A Network Text Analysis of David Ayer's *Fury*

Starling David Hunter (Corresponding author)

Tepper School of Business, Carnegie Mellon University, 5032 Forbes Ave, Pittsburgh PA, USA

E-mail: starling@andrew.cmu.edu

Susan Smith

College of Mass Communication, American University of Sharjah, Sharjah, United Arab Emirates

E-mail: ssmith@aus.edu

Abstract

Network Text Analysis (NTA) involves the creation of networks of words and/or concepts from linguistic data. Its key insight is that the position of words and concepts in a text network provides vital clues to the central and underlying themes of the text as a whole. Recent research has relied on inductive approaches to identify these themes. In this study we demonstrate a deductive approach that we apply to the screenplay of the 2014 World War II-era film *Fury*. Specifically, we first use genre expectations theory to establish prior expectations as to the key themes associated with war films. We then empirically test whether words and concepts associated with the most influentially-positioned nodes are consistent with themes common to the war-film genre. As predicted, we find that words and concepts associated with the least constrained nodes in the text network were significantly more likely to be associated with the war, action, and biography genres and significantly less likely to be associated with the mystery, science-fiction, fantasy, and film-noir genres.

Keywords: content analysis, text analysis, network text analysis, semantic network analysis, film studies, screenplay, screenwriting, war movies, World War II, tanks

1. Introduction

Network text analysis (NTA) is a term used to describe a wide variety of "computer supported solutions" that model a text as a network "of words and the relations between them" (Diesner & Carley, 2005, p. 83). Constructing these networks is a four-step process and differences in how each is performed account for much of the variety to be found in approaches to NTA (Diesner, 2012). The first step involves the *selection* of which words are to be included or excluded in the analysis. The second step involves the *abstraction* of the included words to higher-order conceptual categories. In the third step, *connections* are established between pairs of related concepts. Subsequent analysis of the resulting network involves a fourth step—the identification or *extraction* of the key themes. Like other forms of content analysis, NTA explicitly assumes that structure encodes meaning (Fischer-Starcke, 2009). Where it differs from traditional content analytic approaches, i.e. those concerned with word-frequency, is that meaning is encoded in the structure of the network. Most specifically, in the extraction phase of NTA, prime importance is placed upon the position or role of concepts within the text network. In short, the more influential the network roles and position occupied by concepts, the greater the assumed thematic or semantic relevance they are assumed to have.

A number of recent studies have focused attention on the last step. In the last five years alone these include network text analyses of *abstracts* of academic journal articles (Beam, et al, 2014), medical school *mission statements* (Grbic, Hafferty, & Hafferty, 2013), *presidential inaugural addresses* (Light, 2014), *violent extremist propaganda* (Morris, 2014), *screenplays* and *novellas* (Hunter & Singh 2015; Hunter & Smith, 2014), *energy policy speeches* (Shim, Park, & Wilding, 2015), as well as newspaper articles about the *global financial crisis* (Nerghes, Lee, Groenewegen, & Hellsten, 2015), *mad cow disease* (Lim, Berry, & Lee, 2015), the *creationism debate* in the US (Shortell, 2011) and two major cities in *Afghanistan* (Martin, Pfeffer, & Carley, 2013). In these studies, measures of concept position within the text networks include *degree centrality* (Shortell, 2011; Grbic, Hafferty, & Hafferty, 2013; Martin, Pfeffer & Carley, 2013; Morris, 2014;) , *betweenness centrality* (Light, 2014; Shim, Park, & Wilding, 2015; Nerges, Groenewegen, & Hellsten, 2015), and network constraint (Hunter & Singh, 2015).

Despite the wide variety of corpora and research questions, the research designs share two important features. The first is the use of exploratory or inductive methods. Put another way, falsifiable hypotheses concerning both the content and positions of the themes are rarely if ever formulated and tested. Rather, the approach has been to select texts on a particular topic, generate semantic networks therefrom, identify the words/concepts occupying the most influential network positions or roles, and for the purposes of the ensuing analysis treat those words/concepts as indicators of the most important themes. For example, Shim, Park & Wilding (2015) undertook to "explore and compare nuclear energy

policy frames" in six countries –Japan, South Korea, USA, UK, France, and Germany—in the two years preceding and two years following the March 2011 Fukushima accident. They created semantic networks from "the speeches and addresses made by top policy makers" in each of the six countries. Their subsequent analysis found many important differences both with and across countries, as well as over time. But the analysis was exploratory and no falsifiable hypotheses or propositions were formulated concerning the differences in the centrality of concepts across the various analytical frames.

Similarly, Beam, Applebaum, Jack, et al (2014) undertook to map the semantic structure of "cognitive neuroscience" a field that links the "biological systems" investigated by neuroscientists to the "processing constructs" investigated by psychologists. Notably, their investigation revealed many significant instances of "negative" structure—what they termed "islands" and "gaps"—as well as "positive" structure—what they termed "hubs" and "branches" (ibid., p. 1958). However, despite the clearly stated role of cognitive neuroscience as a "linking discipline", the authors offered no predictions concerning either which concepts would be central in each of the three domains or which concepts, if any, would serve as linchpins among them.

Two recently studies have undertaken an inductive approach concerning the network position of concepts. One of them is Grbic, Hafferty, & Hafferty (2013, p. 853) who studied the semantic structure of mission statements of 132 US medical schools. They divided the schools into four types—research-focused, social mission-focused, public, and private. Their hypothesis was that "key differences in institutional identity and purpose are projected through (medical) school mission statements, " differences that would "become apparent particularly under the lens of a network approach to text analysis." And as predicted they found that the top 10 most central themes—terms like *leader, biomedical, health, research,* and *community*—across the four hospitals differed. They did not, however, predict what any of the most central themes would be. Nor did they specify a priori *how* the most central concepts would differ across the four types of medical schools—only that they would differ.

Another study adopting an inductive approach is Hunter & Singh's (2015) analysis of the screenplay of the 1999 film *Fight Club*. They selected the film not only because of its status as a cult classic, but also because the film and the novel upon which it is based have been the subject of over 100 peer-reviewed journal articles with emphases on literary criticism, film studies, religion, philosophy, media and culture, race and ethnicity, gothic studies, psychotherapy, and the sociology of sport. And it was from the abstracts of such journal articles that they first identified thirteen prominent themes prominent in the academic discourse about the film. The four most frequently occurring in the sample of 52 abstracts were gender, social and individual identity, capitalism, and anarchism. As expected, they found these themes to be clearly associated with the most central and least constrained nodes in the morpho-etymological network that they constructed from the screenplay's text. That said, it is important to note that neither Hunter & Singh (2015) nor Grbic, Hafferty, & Hafferty (2013) offered theoretically-grounded justifications for their predictions. And that is where the present study stands to contribute to the existing literature concerning the extraction of important themes. Our is the first of which we are aware that grounds our predictions in such a manner. As detailed in the next section, in this study we use genre expectations theory (Bignell, 2002; Altman, 1984; Eberwein, 2009) to determine *a priori* what themes should be prevalent in a war film in general and in an Iraq-war film in particular. We then test whether those themes are associated with centrally or influentially positioned nodes in the text network that we construct from the text of the screenplay. We then externally validate those results by having survey respondents attempt to determine the film's genre based only on their examination of words and concepts associated with the most influentially positioned nodes in the network.

2. Literature Review & Hypothesis

The *American Heritage Dictionary of the English Language* defines the word *genre* both as "a type or class" and more specifically—in reference to the arts—as "a category of artistic composition…marked by a distinctive style, form, or content." In film studies the term "genre analysis" refers to the classification of films into recognizable groups and types—e.g. comedy, drama, science-fiction, and horror—as well as the study of the "codes and conventions" that define them (Bignell, 2002, p. 199). Many aspects of a film can convey its genre. These include, but are not limited to the story structure (plots, characters, issues, situations), locations and backdrops, props, the narrative style, dialog, lighting, emphasized camera shots, the musical score and sounds, lighting, etc. (Bignell, 2002; Altman, 1984). "Genres have characteristic features that are known to and recognized by audiences." For example, in a Western we see similar characters, situations, and settings, e.g. native Americans, settlers and homesteaders, horses and men on horseback, stagecoaches and covered wagons, guns and gunfights, corrals and ranches, wilderness and wide open spaces.

Taken together, all of these things—and more—"offer the audience a set of expectations" and allow the film producer a template upon which to base its marketing and promotional discourse. (Bignell, 2002, p. 199). There is currently no universally agreed-upon set of film genres. The *International Movie Database* (imdb.com) classifies films into 21 genres which have remained relatively constant over time (http://www.imdb.com/genre/). *Box Office Mojo* (BOM) , however, currently lists 217 genres and sub-genres (http://www.boxofficemojo.com/genres/). Notably, neither site claim that their genre categories are either mutually exclusive or cumulatively exhaustive, or scientifically precise (Bordwell & Thompson, 2009). In fact, BOM explicitly states that "more genres will be added over time" and it invites readers to indicate what new (sub-) genres should be added. On both sites, a very large number of films are assigned to more than one genre but few appear to have more than four. Although definitions of genres vary, implicit in standard definitions is the notion that films can be classified into groups that have high similarity within the groups and low similarity across

them. In other words, members of a genre share particular certain conventions of "content, e.g. themes or setting" with one another to a much greater degree than they do with those belong to other genres.

In the introduction of his book entitled *The Hollywood War Film*, Eberwein (2009) places the "conventions" that define the genre into two categories—stock "characters" and "basic narrative elements." As summarized in in the first column of Table 1, below, the characters of three types—*males*, *females*, and "*youth/children/pets*." The former are further subdivided into (a) "the older, seasoned leader" (b) "young recruits" (c) "camp/platoon clown (d) "ladies man" [e] ("newly married or recent father" (f) "regional, ethnic, and racial types" and (g) "examples of different social classes." The basic narrative elements are of three kinds—(1) the "basic training" that characterizes the preparation for combat (2) activities characterizing the *specific branches of the armed services* at war (3) activities or *elements common to all branches* of the armed services and where appropriate (4) the *aftermath of war*.

Table 1. Codes and Conventions that Define the War Film Genre—Adapted from Eberwein, (2009)

MALE CHARACTERS	
•	Older, seasoned leader; Young recruits; Camp/platoon clown; Ladies' man; Newly married or recent father
•	Regional, ethnic, & racial types; Different social classes
FEMALE CHARACTERS	
•	Loyal wife, girlfriend, nurse; Prostitute, floozie; Wise, sustaining mother
YOUTH, CHILDREN, & PETS	
•	Eager brothers, boys; Younger sisters; Endangered or killed child; Animals (dogs, cats)
PRE-COMBAT: BASIC TRAINING	
•	Tyrannical squad leader; Demanding exercises, drills; Bonding, pranks; Weekend passes; Sexual initiation
•	Successful graduation, completion of training
COMBAT: ARMY/INFANTRY/MARINES	
•	Water landings, patrols, ambushes, raids, digging in; Combat in jungles, deserts, mountains
•	Tanks, grenades, flamethrowers; Dealing with heat/cold (or elements)
ELEMENTS COMMON TO ALL BRANCHES	
•	Writing letters; receiving mail from home (typically birth announcements and "Dear John" letters).
•	Sharing and observing photographs; Listening to the radio
•	Spontaneous and improvised play to alleviate tension and boredom
•	Singing; prayers/church service; communion; Burials with short, moving eulogies & tributes
•	Leaves and Rest & Recuperation; Reflections on the nature of the enemy
POST-COMBAT: AFTERMATH OF WAR	
•	Recovery/rehab for physical/psychological injuries; Difficulty adjusting to civilian life
•	Reunion with wife, girl, family, friend

As noted previously, prior studies in semantic network analysis operate on the assumption that the most influentially positioned words and concepts in text network embody or illustrate the source texts' most important themes or meaning. Prior research of an exploratory kind on screenplays has already demonstrated that thematically-relevant words are associated with the least constrained and most central nodes in text networks (Hunter, 2014). One of those was *The Hurt Locker*, an Iraq War film about a bomb-disposal team working in and around Baghdad. In the text network of the screenplay, words associated with the least constrained nodes in that network included— *HUMVEE, IED (improvised explosive device), machine gun, shell-shocked, suicide bomber, army-issue, body armor, fireball, gunfire, gunshot, , UN* (United Nations), and *USA*. All of these words, and others, were illustrative of the conventions that Eberwein (2009) and others have identified as defining war films, in general, and Iraq war films in particular. As such, our hypothesis is that

> **H1:** *In a text network constructed from the screenplay of a war film, words associated with the most influentially-positioned nodes will embody or illustrate the codes and conventions of the war genre to a greater extent than they do any other genre to which the film does not belong.*

In other words, the influentially-positioned words should be strongly associated with the WAR genre and not strongly associated with genres to which the film does not belong.

3. Methods & Data

The object of our analysis is *Fury*, a 2014 American war film about US tank crews in Nazi Germany during the final months of World War II. The film stars Brad Pitt and Shia LeBeouf and was written and directed by David Ayer. Among Ayers' other screenwriting credits are one other WWII-era film (U-571, 2000), the first installment of the highly-successful *Fast and Furious* franchise (*The Fast & the Furious,* 2001), a forthcoming comic-book adaptation

(*Suicide Squad, 2016*), and several law-enforcement-themed dramas including *Dark Blue* (2002), *Sabotage* (2014) *SWAT* (2003), *End of Watch* (2012), and *Training Day (2001)*, the last of which starred Denzel Washington in an Academy-Award winning performance. The film had its world premiere in Washington DC on October 15, 2014 followed by a wide release to over 3700 theaters across North America on October 17th. As of July 1, 2015 the film had earned $85.8 million domestically and another $126 Million in foreign revenues for a total of 211.8 million against a budget of $68 million.

Critical response to the film has been largely positive. The film has a "Certified Fresh" by *Rotten Tomatoes*, rating of 77% based on 223 reviews (Rotten Tomatoes, 2015). Thirty-one positive ratings out of forty-five reviews by "top critics" as "fresh" leaves the film with a slightly lower 69% rating. Although the film received no Academy Award nominations, the lead and several supporting actors earned nominations and/or wins for their performances. These included the *Best Actor in an Action Movie* from the Broadcast Film Critics Association, *Best Ensemble* (National Board of Review) and *Best Actor in a Supporting Role* (Phoenix Film Critics Society). The film also was nominated and/or won a number of technical awards including *Outstanding Action Performance by a Stunt Ensemble in a Motion Picture* (Screen Actors Guild Awards), *Best Sound Editing* (Motion Picture Sound Editors), *Best Film Editing* and *Best Art Direction and Production Design* (Satellite Awards).

The copy of the screenplay used in this paper was the "pink" revision date October 30, 2013, a date which is one month after principal photography began. Thus, it is likely the final draft of the screenplay and the one which matches most closely the resulting film. The screenplay was downloaded in PDF format directly from the website of Sony Pictures. (Note 1). The script itself is only 103 pages, approximately 17 less than the 120-page industry standard.

4. Network Text Analysis

Recall that the four steps involved in a semantic network analysis, as detailed by Diesner (2012, pp. 90-1), are (1) *selection*, the determination of which words are to be included and excluded from consideration (2) *abstraction*, i.e. assigning the retained words to higher-level conceptual categories (3) *connection*, establishing a relationship for connecting pairs of conceptual categories and (4) and extraction, i.e. extracting or inferring meaning and key themes from the completed network. Our choices for these four steps were consistent with prior research on the semantic networks of screenplays (Hunter, 2014a, b; Hunter & Singh, 2015). In those studies, the only words that are selected are multi-morphemic compounds (MMCs), i.e. *hyphenated* and *closed compounds*, e.g. heavy-handed or shotgun; *acronyms* and *abbreviations*, e.g. *NATO, radar, laser*; *blend words*, e.g. guesstimate (guess + estimate) and motel (moter + hotel); *clipped words*, e.g. internet(work), e(lectronic)-mail; *multi-word compounds*, e.g. son-in-law, over-the-top; *copulative compounds*, e.g. actor/model, attorney-client; *open compounds*, e.g. trade secret and post office and *words with hyphenated prefixes* of three or more letters, e.g. pro-choice, ultra-sophisticated, anti-establishment. The text of *Fury* contained 4291 unique words repeated a total of 24,872 times. Excluding open compounds, the text contained 227 multi-morphemic compounds, about 5.3% of the number of unique words.

Abstraction involved the assignment of each element of each MMC to a category defined as its etymological root. The source used to determine these roots was the 3rd edition of the Watkins' (2011) *American Heritage Dictionary of Indo-European Roots* (AHDIER). which traces more than 13,000 English words back to over 1,300 Indo-European roots. For example, word *shotgun* is comprised of two morphemes—*shot* and *gun*. According to the AHDIER, the former descends from the IE root *skeud-* which means " to shoot, chase, throw" (p. 81). The latter descends from the IE root *gwhen*- which means "to strike, kill" (p. 36). When Indo-European (IE) roots were not identified, then Greek, Latin, Semitic, or other roots are used, as provided in words' etymology in the *American Heritage Dictionary of the English Language* (AHDEL). Because no software exists that etymological stems words in this fashion, the mapping had to be performed manually. At the conclusion of this process the 227 MMCs were traced back to 255 unique roots, 73% of which were Indo-European.

The relationship between a word and its etymological root is genetic in that it suggests that the former descends from the latter. The choice of relationship used to connect these 403 roots was the co-occurrence of descendants of two or more roots within the same MMC. As shown above, the word *shot* descends from *skeud-* and gun from *gwhen*-. In semantic network of the screenplay of *Fury*, these two roots are linked or connected because their descendants co-occur within the same MMC—shotgun. And because roots have many descendants that may co-occur with many other descendants of other roots, the result is a semantic network where the nodes are etymological roots and the linkages representing the MMCs in which the roots co-occur. The 227 nodes in the semantic network for the screenplay of *Fury* were connected by 250 links or ties. As shown in **Figure 1**, the main component of the network—i.e. the largest group of interconnected or mutually-reachable nodes—contained 130 nodes connected by 165 linkages.
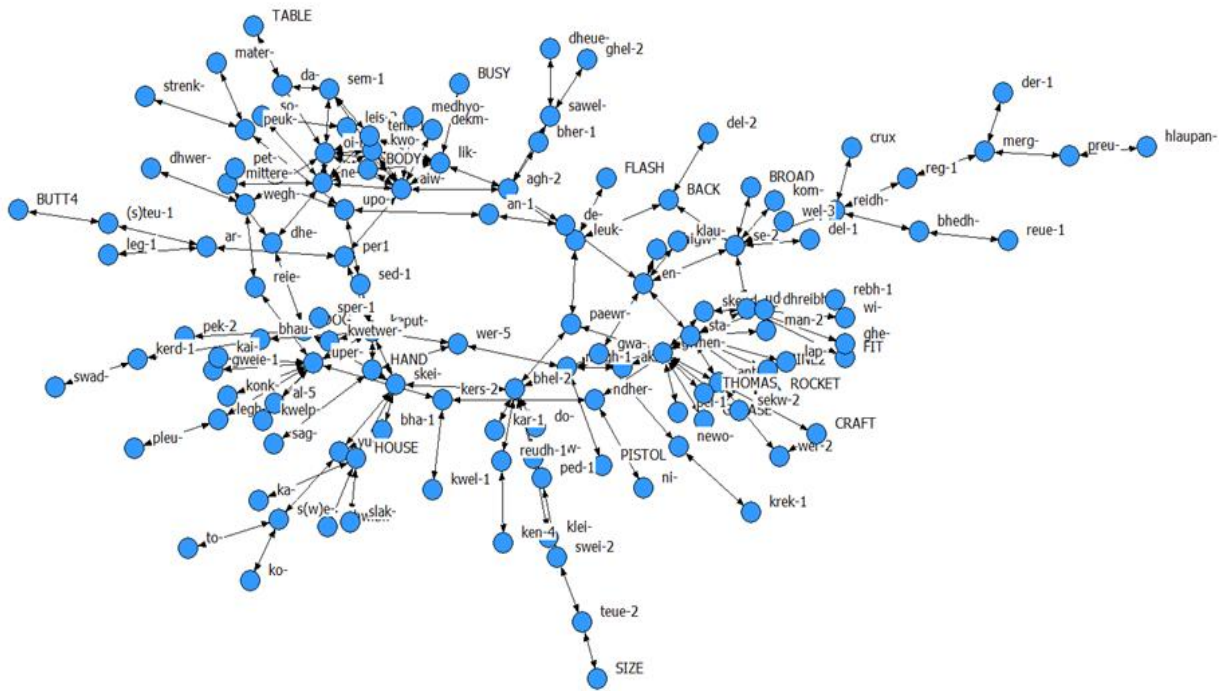
Figure 1. Main Component of the Morpho-Etymological Network of the Screenplay of David Ayers' *Fury*

4.1 Identifying Key Themes

As discussed above, prior research in semantic network analysis has relied upon a variety of node-level measures to identify the most influentially-positioned concepts in a text network, degree and/or betweenness centrality being the most common (Nerges, Groenewegen, & Hellsten, 2015). In this study, we defined "influentially-positioned" nodes as those with betweenness centrality scores in the top 10% of the 130 nodes (etymological roots) in the main component. Those nodes and their betweenness scores are shown in Table 2, below:

Table 2. The Most Influentially-positioned Nodes and Betweenness Centralities and Multi-Morphemic Compounds Associated with Them

Node (definition)	Betweenness Centrality (Normalized)	Associated Multi-morphemic Compounds
kaput- (head)	28.9	forehead, handkerchief, headbutt, headquarters, HQ, headset, overhead, shithead, spearhead
bhel-2 (ball)	24.9	baseball, bulldozer, bullshit, eyeball, fireball, football, hardball, redball
skei- (to cut, split)	24.5	bullshit, dogshit, shithead, shithouse, shit-yer-pants,
en- (in)	21.9	close-in, dug-in, incoming, instead
se-2 (side)	20.8	backside, broadside, countryside, roadside, sidewalk
agh-2 (a day)	20.3	daylight, everyday, Sunday
gwhen (to strike, skill)	18.0	anti-tank gun, COAX (coaxial machinegun), greasegun, gunfire, gunpowder, gunshot, gunsight, MG (machine gun), outgunned, shotgun, sub-machinegun, tommy-gun
aiw- (vital force; life; eternity)	17.9	everyday, forever, middle-aged, nevertheless, teenager
sta- (to stand)	16.3	anti-tank gun, assistant-driver, TC (tank commander), anti-tank mine, anti-tank rocket, lamppost, outpost, instead, understand
uper- (over)	16.0	hangover, overalls, overhead, overheated, overlay, overrun, overwhelming, supercharge, superquick.
per1 (forward, through)	13.2	forearm, forehead, forever.
paewr- (fire)	12.7	gunfire, fireball, firelight,
de- (to)	12.1	today

Recall that in all other semantic network analyses reviewed above—and all of which we are more generally aware—words are nodes in the networks whereas in the morpho-etymological approach the nodes are etymological roots and the words are associated with the edges. As such, while the determination of the most influentially-positioned nodes is a relatively straight-forward process, the only contingency or element of uncertainty being which of many measures of influence to use. In the table above, the third column contains all of the MMCs associated with each of the 13 most influentially-positioned nodes. After excluding prepositions and pronouns, these nodes were associated with the following 57 MMCs : anti-tank (mine), anti-tank (gun), anti-tank (rocket), assistant-driver, backside, baseball, birthday, broadside, bulldozer, bullshit, close-in, COAX (coaxial machine gun), countryside, daylight, dogshit, dug-in, everyday, eyeball, fireball, firelight, football, forearm, forehead, forever, grease-gun, gunfire, gunpowder, gunshot, gunsight, handkerchief, hardball, headbutts, headquarters, headsets, incoming, instead, lamppost, MG (machine gun), middle-aged, nevertheless, outgunned, outpost, overhead, Redball, roadside, shithead, shithouse, shit-yer-pants, shotgun, sidewalk, spearhead, Sunday, TC (Tank Commander), teenager, today, Tommy-gun, and understand.

4.2 Results

Several of the above MMCs typify several of the codes and conventions of war films defined by Eberwein (2009). Recall that his "COMBAT" category included participation in and participation for combat ("ambushes, raids, digging in"), the locations where combat takes, and the weaponry itself (tanks, grenades, flamethrowers, etc.). Clearly, several MMCs are relevant to this category including *anti-tank mine, anti-tank gun, anti-tank rocket*, *COAX* (coaxial machine gun), *dug-in*, *fireball, grease-gun, gunfire, gunpowder, gunshot, gunsight, incoming, machine gun, outgunned, shotgun, TC (tank commander),* and *Tommy-gun*. Additionally, the term *spearhead* is used in this context figuratively and takes the meaning of the "the leading forces in a military thrust." The MMC *countryside* is used the phrase "German countryside" in reference to the location of all combat which takes place in the film. The term *outpost* appears in the phrase "outpost tank" and refers to a defensive combat role performed by a specific tank in the column. The term *overhead* appears in reference to American P47 warplanes that provided air-cover for the tank column. Associated with the "MALE CHARACTERS" is the troupe of "regional, ethnic, and racial types" representing "different social classes." Among the MMCs used to convey this convention were the names of several characters—*WarDaddy, Coon-Ass, good-old-boy*, *redneck, peachfuzzy teenager*, and *Redball*. Only the latter term was associated with the most influentially nodes and it refers to the supply trucks driven by black American soldiers.

4.3 Survey Results

Taken together, the above MMCs are adequate to support the first hypothesis, i.e. that the influentially positioned concepts (nodes) in the text network of *Fury's* screenplay embody the conventions of the genre. As noted previously, it is at approximately this point that most network text analyses conclude. We, however, opted to further validate our coding with an approach not previously undertaken in any study of which we are aware. Specifically, developed a survey that would allow us to directly compare how well the MMCs convey not the codes and conventions of the war genre, by comparing it to genres to which it does not belong. Specifically, we created a survey instrument that asked respondents to first review the above list of MMCs and then required them to indicate the film genre to which they best correspond. (See Appendix 1 for details of the questions which appeared in the survey instrument). Respondents were from Amazon.com's *mTurk* e-worker service (mTurk.com). All survey respondents were located in the USA, had previously completed at least 5000 human intelligence tasks (HITs) and had a 98% or better approval rates from other employers. Respondents were told in the introduction to the survey that they would be matching keywords extracted from the screenplay of a film to types or genres of films. After viewing just one of the three groups of keywords, respondents were asked to answer a series of 20 questions, each of which provided a definition of a genre and which required the respondent to rate on a 1-10 scale the likelihood that the resulting film belonged to that genre. Higher scores represented higher likelihoods. The question specific to the war genre read as follows: "How likely is it that this list of words was taken from a WAR film, i.e. one that contains numerous scenes and/or a narrative that pertains to a real war (i.e., past or current)." The question regard westerns was worded similarly: "How likely is it that this list of words was taken from a WESTERN film, i.e. a film that contains numerous scenes and/or a narrative that portrays frontier life in the American West during 1600s to contemporary times." The eighteen other genres about which the respondents provided opinions were Action, Adventure, Animation, Biography, Comedy, Crime, Drama, Family, Fantasy, Film-Noir, History, Horror, Music/Musical, Mystery, Romance, Science-Fiction, Sport, and Thriller. Their definitions embedded in the questions were taken directly from the *International Movie Database* (IMDB). (Note 2). Consistent with our qualitative analysis, we found the predicted relationship between the network position of concepts and the film's genre. Table 3 contains the descriptive statistics the scores for each of the 158 usable responses.

Table 3. Descriptive Statistics for Survey Responses

Genre	Min	Max	Median	Average	Std. Dev.	Skewness
Action	1	10	8.5	8.29	1.60	-1.15
War	1	10	8	8.02	2.08	-1.21
Crime	1	10	6	6.08	2.34	-0.43
Adventure	1	10	6	6.08	2.23	-0.62

Thriller	1	10	6	5.53	2.23	-0.24
History	1	10	6	5.33	2.54	-0.09
Drama	1	10	5	5.18	2.44	0.05
Bio	1	10	4	4.24	2.36	0.46
Film-noir	1	10	4	4.16	2.47	0.49
Comedy	1	10	3	3.97	2.48	0.65
Horror	1	10	3	3.90	2.34	0.56
Mystery	1	10	4	3.89	2.03	0.52
Scifi	1	9	3	3.83	2.29	0.55
Sports	1	10	3	3.73	2.46	0.67
Western	1	10	3	3.46	2.54	0.82
Fantasy	1	10	2	3.21	2.28	1.07
Animated	1	10	2	2.72	2.13	1.48
Romance	1	10	1.5	2.24	1.87	2.03
Music	1	10	1	2.16	1.88	2.31
Family	1	10	1	1.98	1.77	2.51

In the IMDB, *Fury* belongs to three of the 20 aforementioned genres—*Action, War,* and *Drama.* As shown in the table above, the mean scores of the corresponding survey responses were 8.29, 8.02, and 5.18, respectively. Notably, these are the first, second, and seventh highest scores among the 20 genre-specific averages. The skewness of the *Action* and *War* responses were also the most highly negative with values of -1.15 and -1.21, respectively. This indicates that the left tail of the distribution of these scores is longer and, accordingly, that the mass of the distribution is concentrated to the right. This skewness can be interpreted to indicate that there was relatively less uncertainty in the scores for these two genres. By comparison, the skewness of the scores for the *Drama* genre was only 0.46, thereby indicating that there was a slight right-hand shift in the distribution. The standard deviation of the distribution of the *Action* and *War* genres were 1.60 and 2.08, the lowest and fourth lowest scores respectively.

While these results strongly suggest that the most influentially-positioned concepts—those with the high betweenness scores—were easily recognizable as belonging to the *War* genre, they only partially support the hypothesis. Recall that our specific prediction was not just that the most influentially-positioned words would illustrate or embody the *War* genre but not do the same for genres to which it did not belong. As we can see from the results in Table 1, the lowest average scores are those associated with the following five genres—*Fantasy* (3.21), *Animated* (2.72), *Romance* (2.24), *Music & Musicals* (2.16) and *Family* (1.98). Notably, these five genres also have the five highest levels of skewness— values of 1.07, 1.48, 2.03, 2.31, and 2.51, respectively. This pattern of results suggests that the survey respondents had very little uncertainty associated with their determinations concerning these five genres in particular. Seventeen t-test of means were also conducted in which *War* genre scores were compared to scores for the remaining seventeen genres to which the film did not belong. In every instance the scores for the *War* genre were very significantly larger ($p < 1E-10$). Taken together these results make clear that the most influentially-positioned concepts—those with the highest betweenness centrality scores—were strongly associated with the War genre in the absolute sense and the relative. As such, the study's hypothesis is strongly supported.

5. Conclusion

As noted in the introductory section of this paper, the extraction of meaning from text or semantic networks involves an examination of the most influentially positioned nodes in the network. That said, we are aware only one other study wherein the themes associated with a text network's most influential nodes were specified *a priori.* Thus, the present study is distinguished from most of the prior literature in this regard. Where the present study is most distinctive concerns the theory we used to generate our falsifiable two hypotheses. Specifically, it was genre theory within the broader film studies literature, and research on the war-film genre that we applied to the screenplay of *Fury.* As discussed in the preceding section, our results supported both hypotheses, i.e. that the words associated with the most influentially-positioned nodes would embody the codes and conventions of the war film genre and that they would so more accurately than words associated with the least influentially-positioned nodes. Recall that we used network constraint (Burt, 2000) as our measure of positional influence and found that several words associated with the subset of least constrained—and thus most influentially-positioned—nodes were clearly associated with the codes and conventions of war films, words like *anti-tank mine, anti-tank gun, anti-tank rocket, COAX* (coaxial machine gun), *dug-in, fireball, grease-gun, gunfire, gunpowder, gunshot, gunsight, incoming, machine gun, outgunned, shotgun, TC (tank commander),* and *Tommy-gun.* In marked contrast, our reading of the words associated with the most constrained—and thus least influentially-positioned—nodes revealed that many fewer evoked the codes and conventions of the war genre. But unlike prior studies of this kind, we also externally validated our impressions of these two sets of words. Specifically, we administered surveys to over 100 participants recruited through Amazon's mTurk service and, as discussed above, we found further and stronger support for our hypothesis. To the best of our knowledge, ours is the only network text analysis that has externally validated its findings.

There are at least two important implications of our findings that should be explicitly noted. First, as noted above, this study is the first of which we are aware that developed specific, falsifiable hypothesis concerning the role and positions of relevant concepts and externally validated the hypothesis. It is our belief that this approach should become a standard practice followed by future network text analyses. Secondly, the approach outlined here may have implications for the practice of screenplay "coverage"—the term used to describe the qualitative analysis of screenplays by readers, directors, agents, and producers. As noted by Eliashberg, Hui, and Zhang (2007), "deciding which scripts to produce is dauntingly difficult task, as the number of submissions always greatly exceeds the number of movies that can be made" (p. 881). The approach used by all major studios to select screenplays from the thousands submitted each year has been described both "age-old" and "labor-intensive" (ibid, p. 881). In short, they hire 3-4 readers who read the script and provide a written "synopsis of the story line" along with a "recommendation on whether the screenplay should be produced into a movie and the changes, if any, that are needed before actual production" (ibid, p. 881-2) Because the evaluations are entirely subjective and qualitative, it means that the success of a movie relies in large part on the "quality of the available readers and their acumen in picking out promising scripts" (ibid., p. 882). In addition, because readers can and do frequently disagree—as do producers—the process is characterized by high levels of uncertainty and strikes many as arbitrary (Hague, 2011). Although this study did not attempt to link the position of key concepts to the screenplay's quality, it's worth noting that it's possible that screenplays will vary systematically in this regard. More specifically, it is possible that in higher quality screenplays will be characterized by a greater number of genre-relevant concepts in the more influential positions within the text network. If so, then it means that an objective and much-needed measure of screenplay quality could be developed, one based on the process and approach outlined here.

We should note that there are several limitations to this study that should be explicitly recognized, limitations that may place bounds around on the generalizability of the result. The first concerns the nature of the text analyzed. Whereas all other network text analyses of which we are aware use non-fiction, we used a screenplay, one adapted from an autobiography. This is important because contemporary screenplays adhere to very well-defined set of story-telling conventions, plot devices, and narrative structures (Field, 2005; Snyder, 2005) and are characterized by a level or thematic and lexical repetition not common to texts in other domains (Hunter& Smith, 2013). Secondly, we analyzed a screenplay from a film genre that is long-standing and well-defined in the minds of American movie-goers and media consumers. It's possible that widespread familiarity with war films enhanced survey respondents' ability to correctly identify the genre and it's an open question as to whether a romantic comedy or a family drama would be so easily identified through a similar approach. Third, we should note that the method upon which we based our findings relies on a single source for tracing multi-morphemic compounds back to their etymological roots—the *American Heritage Dictionary of Indo-European Roots* (Watkins, 2011). Recall that approximately 75% of the individual morphemes in the network model were traced back to that source. It is possible that some of the remaining 25% could have been traced to common roots—Indo-European or otherwise—described and defined in other well-known sources, e.g. the *Barnhart Concise Dictionary of Etymology* (Barnhart, 1995) or *The Concise Oxford Dictionary of English Etymology* (Hoad, 1993). Finally, we should note that it is unclear whether and to what degree the approach described here can be applied to languages other than English. While it is widely accepted in comparative linguistics that compounding is a common to all languages, the same compound words are not always used across languages to describe the same thing. For example, the English compound *laptop* doesn't translate directly into many other languages. Furthermore, research on the Indo-European roots of those languages may not have progressed as far as it has in English. As such, there may be serious limitations associated with the application of the methods described herein to other Indo-European or other language families.

References

Altman, R. (1984). A semantic/syntactic approach to film genre. *Cinema Journal*, 6-18.

Barnhart, R. K. (Ed.). (1995). *The Barnhart Concise Dictionary of Etymology*. New York, NY: Harper-Collins.

Beam, E., Appelbaum, L. G., Jack, J., Moody, J., & Huettel, S. A. (2014). Mapping the semantic structure of cognitive neuroscience. *Journal of Cognitive Neuroscience*, *26*(9), 1949-1965.

Bignell, J. (2002). *Media semiotics: An Introduction*. Manchester, UK: Manchester University Press.

Box Office Mojo (2015), Retrieved April 23, 2015, from http://www.boxofficemojo.com/movies/?id=savingprivateryan.htm

Burt, R. S. (2000). The network structure of social capital. *Research in Organizational Behavior*, *22*, 345-423.

Diesner, J., & Carley, K. M. (2005). Revealing social structure from texts: meta-matrix text analysis as a novel method for network text analysis. *Causal mapping for information systems and technology research: Approaches, advances, and illustrations*, 81-108.

Diesner, J. (2012). *Uncovering and managing the impact of methodological choices for the computational construction of socio-technical networks from texts*. Carnegie-Mellon University, Pittsburgh PA. Institute of Software Research International.

Eberwein, R. (2009). *The Hollywood War Film*. Malden, MA: John Wiley & Sons.

Eliashberg, J., Hui, S. K., & Zhang, Z. J. (2007). From story line to box office: A new approach for green-lighting movie scripts. *Management Science*, *53*(6), 881-893.

Field, S. (2005). *Screenplay: The Foundations of Screenwriting* (2nd ed.). New York: Delta.

Fischer-Starcke, B. (2009). Keywords and frequent phrases of Jane Austen's *Pride and Prejudice*: A corpus-stylistic analysis. *International Journal of Corpus Linguistics*, *14*(4), 492-523.

Grbic, D., Hafferty, F. W., & Hafferty, P. K. (2013). Medical school mission statements as reflections of institutional identity and educational purpose: A network text analysis. *Academic Medicine*, *88*(6), 852-860.

Hauge, M. (1991). *Writing screenplays that sell*. New York, NY: Harper-Collins.

Hoad, T. F. (Ed.). (1993). *The Concise Oxford Dictionary of English Etymology*. Oxford: Oxford University Press.

Hunter, S. (2014). A Novel Method of Network Text Analysis. *Open Journal of Modern Linguistics*, *4*(2), 350-66.

Hunter, S., & Singh, S. (2015). A Network Text Analysis of Fight Club. *Theory and Practice in Language Studies*, *5*(4), 737-749.

Hunter, S., & Smith, S. (2013). Thematic and Lexical Repetition in a Contemporary Screenplay. *Open Journal of Modern Linguistics*, *3*(1), 9-19.

Hunter, S., & Smith, S. (2014). A Network Text Analysis of Conrad's *Heart of Darkness*. *English Linguistics Research*, *3*(2), p39.

Light, R. (2014). From Words to Networks and Back: Digital Text, Computational Social Science, and the Case of Presidential Inaugural Addresses. *Social Currents*, *1*(2), 111-129.

Lim, S., Berry, F. S., & Lee, K. H. (2015). Stakeholders in the Same Bed with Different Dreams: Semantic Network Analysis of Issue Interpretation in Risk Policy Related to Mad Cow Disease. *Journal of Public Administration Research and Theory*, in press.

Martin, M. K., Pfeffer, J., & Carley, K. M. (2013). Network text analysis of conceptual overlap in interviews, newspaper articles and keywords. *Social Network Analysis and Mining*, *3*(4), 1165-1177.

Morris, T. (2014). Networking vehement frames: neo-Nazi and violent jihadi demagoguery. *Behavioral Sciences of Terrorism and Political Aggression*, *6*(3), 163-182.

Nerghes, A., Lee, J-S., Groenewegen, P. & Hellsten, I. (2015). Mapping Discursive Dynamics of the financial crisis: a structuralist perspective of concept roles in semantic networks. *Computational Social Networks, 2*(16), 1-29.

Rotten Tomatoes (2015). Retrieved August 4, 2015, from http://www.rottentomatoes.com /m/fury_2015

Shim, J., Park, C., & Wilding, M. (2015). Identifying policy frames through semantic network analysis: an examination of nuclear energy policy across six countries. *Policy Sciences*, *48*(1), 51-83.

Shortell, T. (2011). The conflict over origins: A discourse analysis of the creationism controversy in American newspapers. *Mass Communication and Society*, *14*(4), 431-453.

Snyder, B. (2005). Save the cat! The last book on screenwriting you'll ever need. Studio City, CA: M. Wiese Productions.

Watkins, C. (Ed.). (2011). *The American Heritage Dictionary of Indo-European Roots*. Boston: Houghton Mifflin Harcourt.

Notes

Note 1: http://www.sonypictures.com/movies/fury/awards/scripts/fury_script.pdf
Note 2: http://www.imdb.com/help/search?domain=helpdesk_faq&index=2&file=genres

Appendix

1. How likely is it that this list of words was taken from an **ACTION** film, i.e. a film that contains numerous scenes where action is spectacular and usually destructive?

2. How likely is it that this list of words was taken from an **ADVENTURE** film, i.e. one that contains numerous consecutive and inter-related scenes of characters participating in hazardous or exciting experiences for a specific goal.

3. How likely is it that this list of words was taken from an **ANIMATED** film, i.e. a film where the majority of its scenes are wholly or partly animated, hand-drawn, computer-generated, stop-motion, etc?

4. How likely is it that this list of words was taken from a **BIOGRAPHICAL** film, i.e. a film that focused on the depiction of activities and personality of a real person or persons, for some or all of their lifetime?

5. How likely is it that this list of words was taken from a **COMEDY**, i.e. a film that mostly contains characters participating in humorous or comedic experiences?

6. How likely is it that this list of words was taken from a **CRIME** film, i.e. a film that contains numerous consecutive and inter-related scenes of characters participating, aiding, abetting, and/or planning criminal behavior or experiences usually for an illicit goal?

7. How likely is it that this list of words was taken from a **DRAMA**, i.e. a film that contains numerous consecutive scenes of characters portrayed to effect a serious narrative throughout?

8. How likely is it that this list of words was taken from a **FAMILY** film, i.e. the film is aimed specifically for the education and/or entertainment of children or the entire family?

9. How likely is it that this list of words was taken from a **FANTASY** film, i.e. it contains numerous consecutive scenes of characters portrayed to effect a magical and/or mystical narrative?

10. How likely is it that this list of words was taken from a **FILM NOIR**, i.e. a film that features dark, brooding characters, corruption, detectives, and the seedy side of the big city?

11. How likely is it that this list of words was taken from a **HISTORY** film, i.e. a film whose primary focus is on real-life events of historical significance featuring real-life characters?

12. How likely is it that this list of words was taken from a **HORROR** film, i.e. a film that contains numerous consecutive scenes of characters effecting a terrifying and/or repugnant narrative throughout?

13. How likely is it that this list of words was taken from an **MUSICAL/MUSIC** film, i.e. a film that contains several scenes of characters bursting into song aimed at the viewer and/or contains significant music-related elements, e.g. portrays a concert or a story about a band?

14. How likely is it that this list of words was taken from a **MYSTERY** film, i.e. a film that contains numerous inter-related scenes of one or more characters endeavoring to widen their knowledge of anything pertaining to themselves or others?

15. How likely is it that this list of words was taken from a **ROMANCE** film, i.e. a film that contains numerous inter-related scenes of a character and their personal life with emphasis on emotional attachment or involvement with other characters, especially those characterized by a high level of purity and devotion?

16. How likely is it that this list of words was taken from a **SCIENCE FICTION** film, i.e. a film that contains numerous scenes, and/or the entire background for the setting of the narrative, based on speculative scientific discoveries or developments, environmental changes, space travel, or life on other planets?

17. How likely is it that this list of words was taken from a **SPORTS** film, i.e. a film whose focus is on sports or a sporting event, either fictional or actual?

18. How likely is it that this list of words was taken from a **THRILLER**, i.e. a film that contains numerous sensational scenes or a narrative that is sensational or suspenseful?

19. How likely is it that this list of words was taken from a **WAR** film, i.e. a film that contains numerous scenes and/or a narrative that pertains to a real war (i.e., past or current)?

20. How likely is it that this list of words was taken from a **WESTERN**, i.e. a film that contains numerous scenes and/or a narrative that portrays frontier life in the American West during 1600's to contemporary times?