A Network Text Analysis of Conrad's *Heart of Darkness*

Starling Hunter[1] & Susan Smith[2]

[1] Tepper School of Business, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

[2] School of Communication, American University of Sharjah, Sharjah, United Arab Emirates

Correspondence: Starling Hunter, Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA, 15289, USA. E-mail: starling@andrew.cmu.edu

Received: September 23, 2014 Accepted: October 6, 2014 Online Published: October 8, 2014

doi:10.5430/elr.v3n2p39 URL: http://dx.doi.org/10.5430/elr.v3n2p39

Abstract

The hallmark of Network Text Analysis (NTA) is the creation of semantic networks or concept maps from linguistic data. Its key insight—one borrowed from studies in Social Network Analysis—is that the *position* of concepts within such networks reveals vital information about the meaning of the text as a whole. A second hallmark of NTA is that the structure and size of a network are independent of the *frequency* of the words of which it is comprised. In this paper we demonstrate the application of NTA to Joseph Conrad's *Heart of Darkness* (HoD). Using morphological and etymological relationships as the basis for the network, we first represent HoD as a network consisting of over 385 nodes. We then compare and contrast the results of our network analysis with those reported in a widely-cited corpus stylistic analysis of HoD. While our results are remarkably consistent with and complementary to results in that study, we also report results not identified in that analysis, results which could only have been identified using NTA.

Keywords: Network analysis, Network text analysis, Corpus stylistics, Content analysis, Social network analysis, Text networks

1. Introduction

Despite intense and vocal criticism (Fish 1973, 1979; MacKay 1996, 1999), corpus-assisted evaluations of literary texts are becoming increasingly objective and systematic (e.g. Stubbs, 2001, 20005; O'Halloran, 2007; Fischer-Starcke 2009; Biber, 2011). Although there are several potential explanations, among them must surely be the increased availability and sophistication of text analysis software and the consequent impact on descriptions and theories of language in use (Smith, Hoffman, Rayson, 2008). That said, there is one widely-used and highly relevant set of methods that has been largely overlooked by corpus stylistics studies—Network Text Analysis (NTA). The hallmark of NTA is the construction of concept networks or maps of linguistic data. Its key insight—one borrowed from studies in Social Network Analysis (SNA)—is that the thematic importance of words and concepts in a text is a function of their *position* within the network, not their frequency of occurrence (e.g. Carley, 1997; Popping, 2003; Atteveldt, 2008). In short, words and concepts that occupy more influential positions in text networks are more likely to convey key themes in a text compared to those occupying less influential positions.

In this paper we demonstrate the application of NTA to Joseph Conrad's (1899/2010) *Heart of Darkness* (HoD). Using morphological and etymological relationships as the basis for the network (Hunter 2014a, 2014b; Hunter & Smith 2013), we first represent HoD as a semantic network consisting of 385 interlinked concepts. We then compare and contrast the results of our network analysis with those reported in *Conrad in the Computer* (Stubbs 2005). In that paper Stubbs used traditional stylistics methods, i.e. the analysis word and phrase frequencies, collocation, distribution, etc. in his analysis of that text. On the whole our results are remarkably consistent with and complementary to his. Specifically, many of his high "keyness" words—the most statistically over-represented words within the text—are centrally positioned within our network representation of HoD. The same applies to the broader themes that they imply. That having been said, we go beyond replicating the findings of these prior research studies and also uncover themes and relationships among them that Stubbs did not identify.

The remainder of this paper is organized as follows. In the next section we describe network text analysis—its theoretical foundations, its distinctive analytical methods, and its many and varied applications. In the third section we describe our application of NTA to Conrad's *Heart of Darkness* along with the results of that analysis. In the fourth section of the paper we compare and contrast our results with those of Stubbs (2005). We conclude the paper

with a discussion of the limitations of our analysis, as well as the opportunities and challenges associated with the application of NTA to literary texts.

2. What is Network Text Analysis?

In short, network text analysis (NTA) involves the encoding of relationships among words in a text and subsequently representing those relationships as a network of linked words and concepts (Popping, 2000). Over the last four decades several methods of NTA have been developed and applied to a wide variety of research questions and types of texts. These include but are not limited to *centering resonance analysis* (Corman et al., 2002), *functional depiction* (Popping & Roberts, 1997), *knowledge graphing* (Bakker, 1987; Popping, 2003), *map comparison/ analysis* (Carley & Palmquist, 1992), *sociocognitive networks* (Carley, 1997; Diesner, 2013), *network evaluation* (Kleinnijenhuis, de Ridder & Rietberg, 1997), *word network analysis* (Danowksi, 1993), *semantic networks* (Sowa, 1992) ; *semantic webs* (van Atteveldt, 2008); *concept maps* (Novak ,1990), *mental models* (Collins and Loftus, 1975); *semantic grammars* (Roberts, 1997), and *morpho-etymological networks* (Hunter, 2014a).

The underlying assumptions motivating all of these approaches are that (1) otherwise invisible structures can be identified when words and concepts are represented and understood as networks (Diesner & Carley, 2005) and (2) that the position of words and concepts in a text network helps to identify prominent themes of the text as a whole. But despite these common foundational assumptions, approaches to NTA differ on a number of important dimensions, namely the level and unit of analysis (e.g. focusing on verbs or nouns or semantic and syntactic categories) and the degree to which the analysis can be automated or computer-supported (Diesner & Carley, 2005). Automated or not, the creation of networks from texts involves the same two fundamental steps: (1) the assignment of words and phrases to conceptual categories and (2) the assignment of links to pairs of those categories. Diesner (2012, pp. 90-1) breaks these two steps into four: *selection*, i.e. choosing which words are included in or excluded from the network analysis; *abstraction*, i.e. assigning the remaining words to higher-level conceptual categories; *connection*, i.e. relating or connecting conceptual categories to one another; and *extraction, i.e.* extracting or inferring meaning and key themes from the completed network. The following example clearly demonstrates each of these four steps. The texts used are the first ten sentences of the first inaugural addresses of the last three Presidents of the United States—Barack Hussein Obama, George W. Bush, and William Jefferson Clinton (Bartelby.com, 2014).

2.1 A Simple Example of Network Text Analysis

As noted previously, that the first step in the network text analysis process is *selection.* For this example the only words of interest are the nouns appearing in each of the first ten sentences of each President's first inaugural speech. Table 1 contains a summary. The second step is *abstraction* and in this example all nouns appearing in the same sentence are assigned to the same category or group. The third step is *connection* and it involves linking the groups or categories defined in step two. In this example we assume a connection between any two of the ten sentences if the same noun occurs in each of them. The last step, *extraction*, involves analyzing the network for evidence of the most influential nodes, as well as the words appearing on the links associated with them. Figure 1 contains the network representations of the linked sentences in the three speeches. In Clinton's speech we see that the fourth sentence (S4) is highly influential because it contains the most links to other sentences—three. In SNA the number of links associated with a node is called degree centrality (Wasserman, 1994). Sentence 4 is also geometrically central because it is the bridge between sentences 8 and 9, on the left side of the network, and sentences 3 and 5, which lie on the right. Also note that the word that links the greatest number sentences is *America* which links sentences 4, 8, and 9. Two other words—*spring* and *world*—each link two sentences.

The text network of George W. Bush's speech is shown in the second panel of Figure 1. In comparison to the network of Clinton's speech, Bush's is notable for two things—a much larger number of interconnected nodes and fragmentation of the network into two components. On the right we see one component wherein four sentences are connected (S1, S2, S3, and S10) while on the left a larger component consists of five interconnected sentences (S5, S6, S7, S8, and S9). In the component on the right, Sentence 2 is the most central because it has links to three sentences (S1, S3, and S10). Notably, the word that co-occurs in the greatest number of sentences is "President." In the larger component, Sentence 7 has a degree centrality of three, the largest of any other. The most frequently co-occurring word is *story*.

Compared to the preceding two, the text network of Obama's speech is most notable for its sparseness. Like Bush's speech, it has two components. The smaller one has two sentences (S2 and S10) linked by the word *nation*. The larger has three sentences (S3, S5, and S8) linked by the words *oath* and *America*. The third sentence (S3) has a degree centrality of two and is, then, the most influential. Each of the three words in the network—*oath*, *nation*, and *America*—co-occur just once in each of the two sentences. Thus, none is more influential than any other. In summary,

the most frequently co-occurring noun in the first ten sentences of Clinton's speech is *America*. For Bush it is both *President* and *story* while for Obama all co-occurring nouns appear with the same frequency. All of those words are more thematically relevant to the degree that co-occurrence indicates an intended emphasis on the part of the authors.

Table 1. Nouns appearing in the first ten sentences of the first inaugural speeches of the three Presidents of the United States

Sentence	Clinton (1993)	Bush (2001)	Obama (2009)
1	Mystery, renewal	President, guest, citizen, transfer, authority, country	Task, trust, sacrifices, ancestors
2	Ceremony, winter	President, service, nation	President, service, nation, generosity, cooperation, transition
3	Words, faces, world, spring	President, contest, spirit, grace	Americans, oath
4	Spring, democracy, vision, courage, America	Leaders	Words, tides, prosperity, water, peace
5	Founders, independence, world, purpose, the Almighty, America	Place, story, end	Oath, clouds, storms
6	Change, sake, ideals, life, liberty, pursuit, happiness	Story, world, friend, liberator, the old, society, servant, freedom, power	Moments, America, skill, vision, office, We the People, ideals, forbears, documents
7	Music, time, mission	Story, people, generation, ideals	
8	Generation, American	Ideals, promise, chance, person	Generation, Americans
9	Nation, predecessor, President, half-century, America	Americans, promise, life, law	Midst, crisis
10	Millions, men, women, steadfastness, sacrifice, Depression, Fascism, Communism	Nation, course	Nation, war, network, violence, hatred

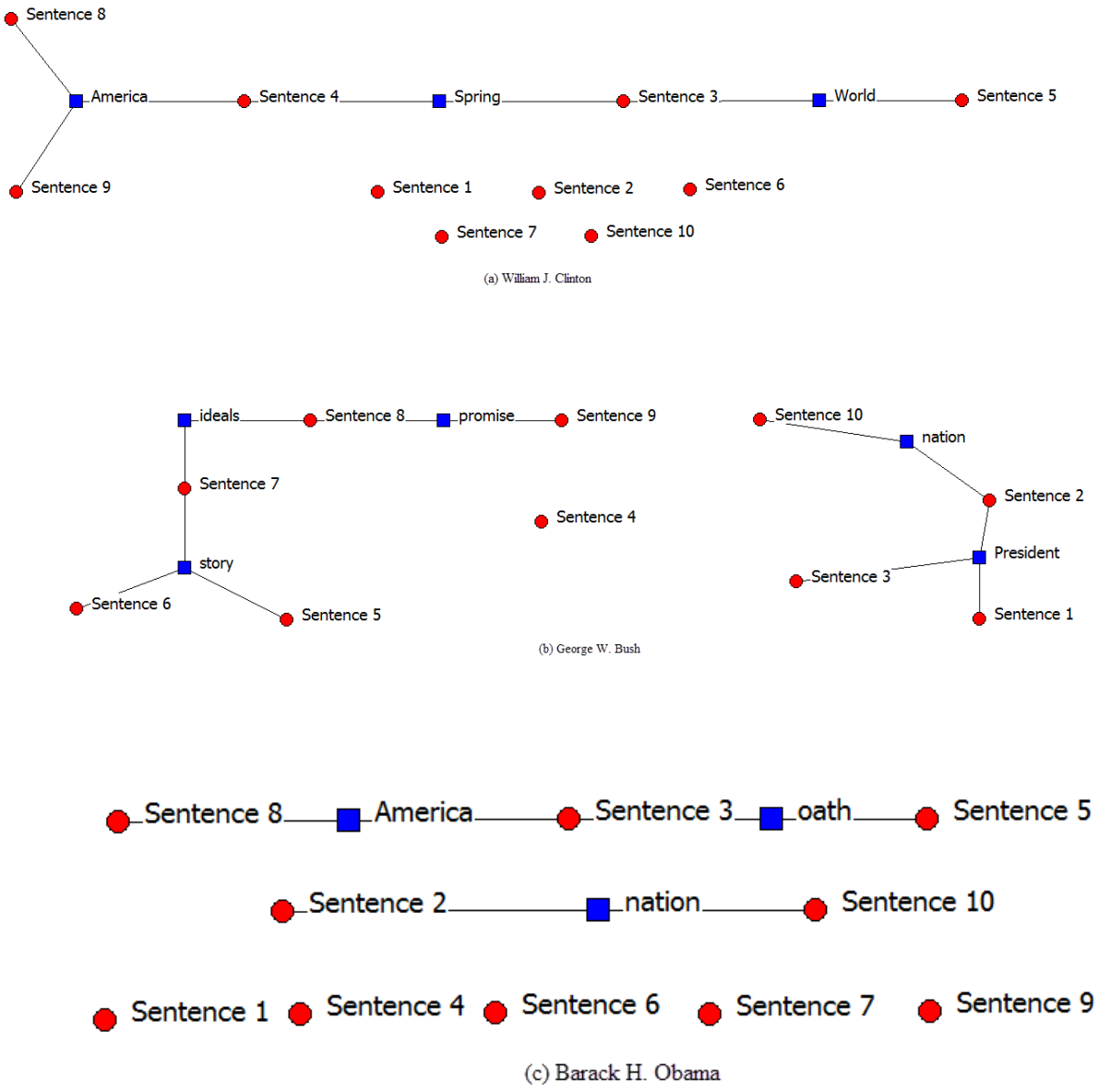

Figure 1. Text networks formed from the first ten sentences of three US Presidents' first inaugural speeches—(a) William J. Clinton (b) George W. Bush, and (c) Barack H. Obama.

2.2 Multi-Morphemic Compounds

As noted in the previous section, there are many different approaches network text analysis. In this study we take the morpho-etymological approach (Hunter & Smith 2013; Hunter 2014a, Hunter 2014b). As the name suggests, the network is constructed from morphological and etymological relationships among words in the text. In terms of the four steps outlined by Diesner (2012), the only multi-morphemic compounds (MMCs) are retained for analysis. These include abbreviations (FBI), acronyms (NATO), blend-words or portmanteau, selected clipped words, several varieties of compound words. Table 2 provides examples of eleven varieties of multi-morphemic compounds.

Table 2. Examples of Eleven Categories of Multi-morphemic Compounds

Type	Examples
Abbreviations and Acronyms:	radar, laser, NATO, AARP, HUMVEE
Blend Words:	moped, guesstimate, dramedy, spork
Clipped Words:	internet, email, pub(lic house), taxi(cab)
Closed Compounds:	afternoon, breakfast, cardboard
Conversion:	photoshop
Copulative compounds:	attorney-client, actor/model
Hyphenated Compounds:	hand-held, off-guard, clean-shaven
Multi-word Compounds:	brother-in-law, sleight-of-hand
Open Compounds:	trade secret, fixed stars
Prefixed:	overstate, underestimate
Suffixed:	software, clockwise

Abstraction is the second stage of a network text analysis each morpheme in a multi-morphemic compound is traced back to its own etymological root. Where possible, that root is Indo-European (Watkins, 2011) and when not, Greek, Latin, Semitic, or other roots are used. For example, the word *breakfast* is comprised of two morphemes—*break* and *fast*. The Indo-European (IE) root for the former is **bhreg-** which means "to break" (p. 12) while the IE root for the latter is **pa-** which means "to protect, feed" (p. 61). These two roots serve as the higher-order conceptual categories which form the nodes in the text network. Thus, if along with *breakfast* the following multi-morphemic compound words—*fastball*, *ballpark*, and *parking brake* were present in the same text sample, then resulting morpho-etymological text network would appear as shown in Figure 2. The resulting network has four node and four links because, according to Watkins (2011, p. 12), the word *ball* descends from the Indo-European root **bhel-2** which means "to blow, swell" while the word *park* descends from the Old French **parc**.

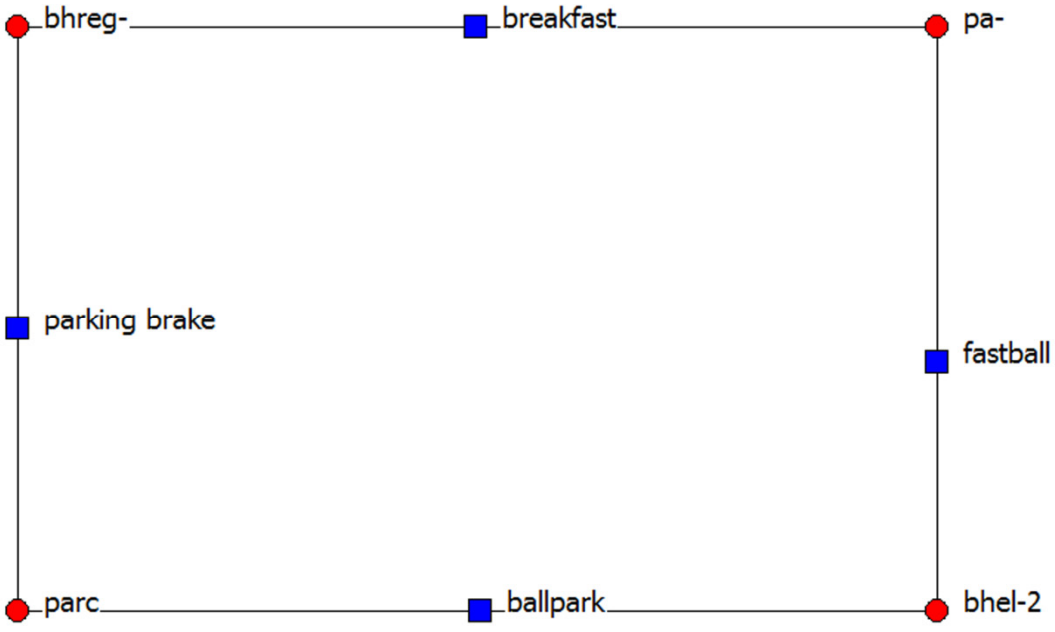

Figure 2. Example of a Small Morpho-Etymological Text network

3. Data & Methods

3.1 Selection

As discussed above, only multi-morphemic compounds (MMCs) were retained for subsequent analysis. Identifying the MMCs was accomplished by first generating a "concept list" with the CASOS Institute's *Automap* software program (Carley, Columbus, and Landwehr, 2013). The "Concept List" output listed each unique word appearing in the text, along with each word's frequency of occurrence. Including differences attributable to inflection, pluralization, and capitalization, there were 5827 words in the list appearing 39430 times. Each author read the entire list individually and identified the MMCs contained therein. Then they met and reconciled their coding. Our combined coding resulted in a list of 382 MMCs, about 6.6% of the total. Several potential MMCs were excluded from the analysis. First among them were numbers that were spelled out, e.g. *eighteen*. However, we retained numbers that were part of compounds, e.g. *two-hundred-mile*, as well as the adjectival forms of numbers when they appeared as part of an MMC, e.g. *second-rate*, *first-hand*. We did not retain neoclassical compounds, e.g. *photograph* but keep pseudo-compounds when the prefix was also a word that descended from the same root as the prefix, e.g. *understand* and *over-whelmed*. By the same rationale, words like *helpless* and *arch-priest* were not retained because the suffix "–less" and the prefix "arch-" descend from a different etymological roots than their corresponding free morphemes—*less* and *arch*.

A second routine in Automap was used to identify open compounds appearing in the text. It is called the "Suggest Ngrams" module. As its name suggest, the routine generated a list of selected multi-word combinations. In this study, that list contained 2022 n-grams. Again, both authors first reviewed the list separately and then reconciled their coding. The result was a list of 34 open compounds such as *back water*, *double doors*, and *first floor*, and another twenty proper nouns. Concerning the latter, we excluded proper names of persons e.g. *Sir Francis Drake* and *Sir John Franklin* but retained those for places and things, e.g. *Pacific Ocean*, *Inner Station*, *Fleet Street,* and the *International Society for the Suppression of Savage Customs*. With the addition of the n-grams the total number of MMCs rose to 436.

3.2 Abstraction

This stage of the analysis involved mapping each individual morpheme back to its Indo-European or other etymological root. As described in the preceding section, the primary source of etymological roots was the 3rd edition of the *American Heritage Dictionary of Indo-European Roots* (Watkins, 2011). Because no software exists for this task, the mapping had to be performed manually. At the conclusion of this process the 436 MMCs were traced back to 385 unique etymological roots, 76% of which were Indo-European. Each of these 385 roots became nodes in the resulting text network.

3.3 Connection

Just as in the preceding examples, a pair of nodes is connected when they co-occur in the same MMC. For example, the roots **ndher-** and **sta-** are connected because they co-occur in the compound *understand*. Among the 385 nodes, there were 476 such connected pairs. Although that may seem like a lot, the number of possible pairs is 73728 and so 476 represents only 0.65% of that number.

In Social Network Analysis (SNA), the term "main component" is given to the largest portion of the network wherein all nodes are mutually reachable. Of the 385 nodes in the text network for HoD, 312 were mutually-reachable. That is to say, there were one or more paths of varying lengths connecting every node (etymological root) to every other node. This network is presented in Figure 3.

3.4 Extraction

The fourth and final stage of a network text analysis involves ascertaining the core meaning or theme of the text from an examination of the properties of the network. This is typically achieved through an analysis of the most central or influential nodes. There are several measures and metrics that are regularly employed in making such an assessment. In this study we make use of two such measures—degree centrality and coreness. As discussed previously, degree centrality is simply a measure of the number of links associated with a particular node or concept which, in this case, is an etymological root. Thus, if a node is connected to ten others, it's degree centrality is 10. As we show in Table 3, the etymological roots **man-1** (man), **uper-** (over), **upo-** (under, up from under, over), **sta-** (to stand), **en-** (in), and **per1** (forward, before, first, chief) have the highest degree centrality scores (Watkins, 2011).

In SNA, coreness is a measure of the degree to which the node in question belongs to groups of nodes which are highly interconnected (the core) or belongs to the remaining nodes which are less interconnected with one another (Borgatti, Everett, and Freeman, 2002). As depicted in the Table 4, only twenty nodes have coreness scores greater than 0.01. The five with the highest coreness are **oi-no-** (one, unique), **sem-1** (one, as one, together with), **kwo-** (stem of interrogative and relative pronouns), **aiw-** (vital force, life, long life, eternity, young), and **lik-** (body, form; like, same).

Table 3. Etymological Roots with the Degree Centrality Greater than 5

Root (Definition)	Degree	Directly Associated Multi-morphemic Compounds
man-1 (man)	21	black men, Chapman lighthouse, Dutchman, Englishman, fireman, fools-helmsman, foreman, gentlemanly, handy men, headman, mankind, man-of-war, poleman, policeman, seaman, white man, wild men, witch-man, An Inquiry into some Points of Seamanship
uper- (over)	17	moreover, overboard, overcast, overcome, overfed, overgrown, overhanging, overheard, overheated, overhung, overland, overpowering, over-seas, overshadowed, overtaking, overwhelmed, summing-up, supernatural,
upo- (up)	16	get-up, got-up, open-mouthed, pent-up, smash-up, summing-up, thereupon, turned-up, up-country, uphill, upkeep, upon, up-river, uproar, upset, upstairs
sta- (stand)	15	assistant manager, camp stool, Central Station, Inner Station, instead, misunderstanding, railway-stations, stand-offish, station-house, stay-at-home, steering wheel, stern-wheel, storeroom, trading post, understand
per1	13	far-off, first-class, fore-end, fore-finger, forehead, foreman, forepart, forerunner, foresaw, foresight, forthwith, straightforward, therefore, unforeseen,
en- (en)	13	by-and-by, inborn, indeed, inland, Inner Station, inshore, inside, insight, instead, into, reined-in, stamped-in, therein, within.
ud- (out)	12	dug-out, look-out, outbreak, outcry, outfit, outlast, outlines, outraged, outside, thereabouts, uttermost, without
s(w)e- (self)	12	herself, himself, yourself, itself, International Society for the Suppression of Savage Customs, yourself, ourselves, themselves, self-defense, self-respect, self-seeking/self-trade, secrets
(s)kel-1 (cut)	12	half-a-crown, half-a-dozen, half-awake, half-caste, half-cooked, half-English, half-French, half-pint, half-shaped, half-speed, half-way
wegh- (way)	10	anyway, archways, doorway, fairway, flatways, half-way, railway, railway-station, railway-truck, waterway,
se-2 (side)	10	river-side, side-spring, water-side, alongside, bedside, besides, fireside, hillside, inside, outside
okw- (eye)	10	red-eyed, shop-window, weak-eyed, wild-eyed, window-holes, eyeballs, eyebrow, eye-glass, eyelids,
aiw- (ever)	10	everlasting, everybody, everyday, everyone, everything, everywhere, however, middle-aged, nevertheless, whatever,
apo- (off)	9	afternoon, because, far-off, man-of-war, matter-of-fact, stand-offish
reg-1 (right)	8	Director of Companies, dressing case, hairdresser, headdresses, railway-stations, railway-truck, South America.
kaput- (head)	8	forehead, handkerchief, headdresses, headlong, headman, headquarters, pin-heads
sem-1 (same, some)	7	quarrelsome, somebody, somehow, someone, something, sometimes, somewhere, unwholesome
rei-1 (scratch, tear, cut)	7	river-bank, river-demon, river-side, river-steamboat, tight-ropes, up-river,
ped-1 (foot)	7	Eldorado Exploring Expedition, footsore, footsteps, foot-warmer, pilot-house, underfoot,
hus (house)	7	madhouse, lighthouse, pilot-house, station-house, trading house, Chapman Lighthouse, custom-house
dhe- (to make)	7	dark-faced, indeed, matter-of-fact, shamefacedly, surface-truth, yellow-faced
leuk- (light)	6	Starlight, sunlight, torchlight, Chapman Lighthouse, lighthouse, moonlight,
kel-2 (cover, conceal)	6	window-holes, bullet-hole, particolored, penholder, shutter-hole
baec (back)	6	back water, backbiting, backbone, back-breaking, backcloth, background
dwo- (two)	6	biscuit-tin, double doors, half-dozen, twenty-mile, two-hundred-mile, two-penny-half-penny
wed-1 (wet, water)	6	back water, fresh-water, water-gauge, water-gourd, water-side, waterway
sae (sea)	6	sea-going, seaman, sea-reach, An Inquiry into some Points of Seamanship, China Seas
tenk-1 (thing)	6	anything, everything, nothing, playthings, something
to-	6	themselves, thereabouts, therefore, therein, thereupon
gene- (give birth)	6	engine-driver, general manager, gentlemanly, International Society for the Suppression of Savage Customs, mankind, supernatural.
nekw-t- (night)	6	fool-nigger, fortnight, good night, midnight, night-air, nightmare
wi- (with)	6	withdrawn, within, without, forthwith, goodbye,

Table 4. Etymological Roots with Coreness Values Greater than 0.10

Etymological Root (Definition)	Coreness	Directly Associated Multi-morphemic Compounds
oi-no- (one)	0.567	another, anybody, anyhow, anyone, anything, anyway, anywhere, eight-inch, nobody, nothing, nowhere, someone,
sem-1 (same, some)	0.394	quarrelsome, somebody, somehow, something, sometimes, somewhere, unwholesome,
kwo-	0.383	somehow, somewhere, whatever, anyhow, anywhere, everywhere, however, nowhere,
aiw- (ever)	0.372	whatever, everlasting, everyday, everyone, everything, everywhere, however, middle-aged, nevertheless
lik- (like)	0.299	barrack-like, death-like, everybody, everyday, everyone, everything, everywhere, fiend-like, likewise,
tenk-1 (thing)	0.189	everything, nothing, playthings, something, anything,
bodig (body)	0.189	nobody, somebody, anybody, everybody,
ne- (no)	0.154	nothing, nowhere, nevertheless
oktu(o)- (eight)	0.112	eight-inch
kailo- (whole)	0.077	holy terror, unwholesome,
kwes- (pant, wheeze)	0.077	quarrelsome
agh-2 (day)	0.073	to-day, everyday
medhyo- (middle)	0.073	middle-aged, midnight, mizzen-mast
leis-2 (small)	0.073	nevertheless
barracas (barrack)	0.058	barrack-like
dheu-3 (to die)	0.058	death-like, death-mask
pe(i)- (to hurt)	0.058	fiend-like
dlegh- (play)	0.037	playthings
de-	0.014	coming-to, into, to-day
so- (this, that)	0.014	nevertheless

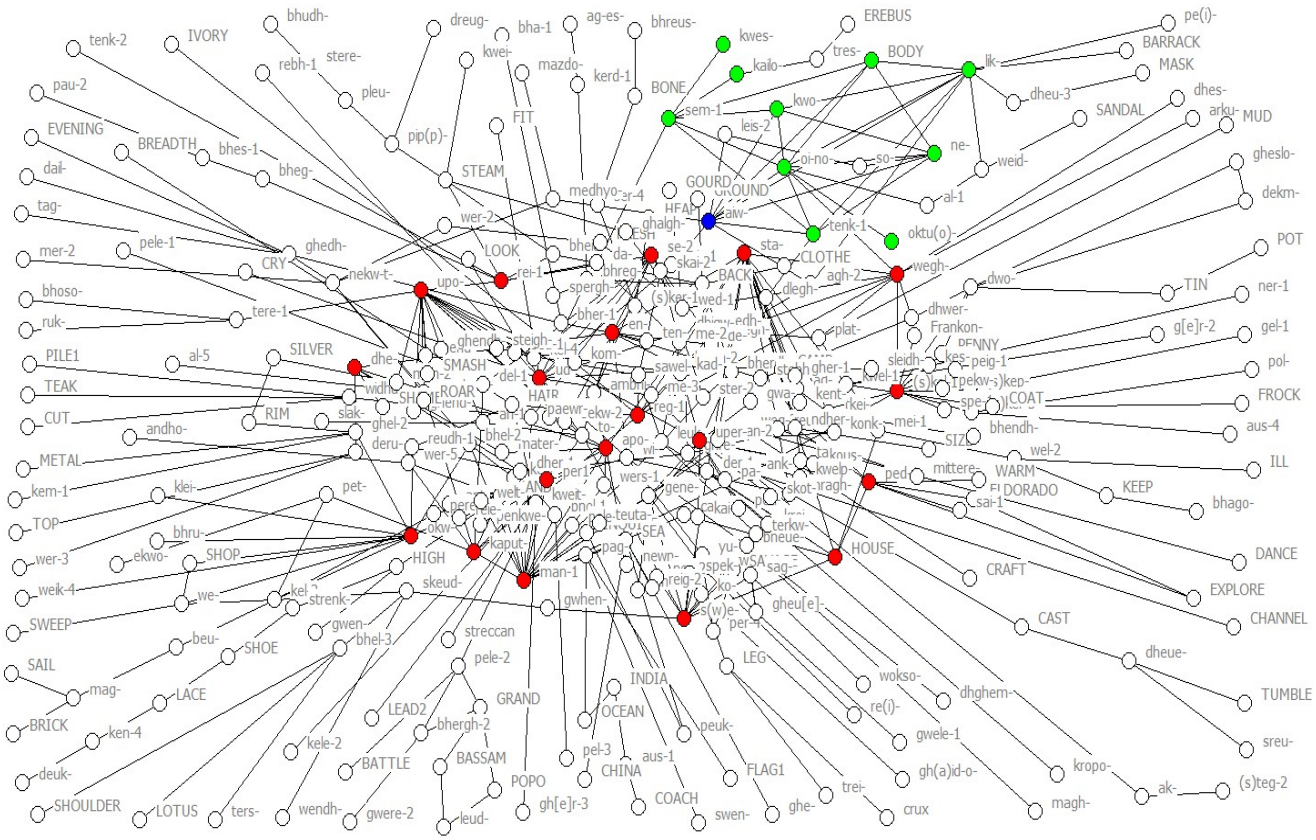

Figure 3. Main Component of a Morpho-Etymological Text Network of *Heart of Darkness*

Legend: Nodes with the ten highest degree centrality scores ($C_d > 7$) are RED. Nodes with the ten highest coreness scores ($C > 0.077$) are GREEN. Nodes that appear in BLUE belong to both groups.

4. Conrad in the Network

In his paper *Conrad in the Computer*, Stubbs (2005) reports the results of his corpus stylistic, content-analytic examination of *Heart of Darkness* (HoD). While not all of his results are relevant to this study, many of those discussed in four sections of his paper—*Vague impressions and unreliable knowledge*, *Some (very) simple frequency data*, *Collocations: Words and Words*, and *Phraseology: words and grammar*—can also be placed and understood in the context of a text network analysis.

4.1 Vague impressions and unreliable knowledge

In this first analytical section, Stubbs asserts that a recurring and dominant theme in *HoD* is "Marlow's unreliable and distorted knowledge," his tendency to be found "looking into a fog, uncertain of what he is seeing," and his "confusion by signs and symbols that he cannot decode" (ibid, p. 9). Drawing on Watt's (1988) observation that "mist and haze" represent "persistent imagery", Stubbs identifies several words from that lexical field that appear almost 150 times in the book—an average of over once per page. Those words are *blur** (2), *dark** (52), *dusk* (7), *fog* (9), *gloom** (14), *haze* (2), *mist** (7), *murky* (2), *shadow* (21), *shade* (8), *shape** (13), *smoke** (10), *vague* (5), *indistinct* (4), and *vapor* (1). Also reinforcing this theme is the large number of words beginning with *some* and *seem—something, someone, somewhere, somehow, seem, seems, seemingly*, etc.—which, Stubbs tells us, occur over 250 times in *HoD*.

Before placing these results in a network context, recall that our text network analysis focuses on multi-morphemic compounds (MMCs). These include compound words, acronyms, abbreviations, and blend words, as well as selected clipped words and proper nouns. The words above can only appear in the main component of the network if they or one of their paronyms are part of an MMC that is mutually-reachable by the majority of others. Notably, several of the above-listed words do not meet this criterion. Specifically, the words *vapor, mist, murk, smoke, indistinct, blur, fog, gloom, haze,* and *vague* and their variants are absent from the list of MMCs used to generate the network. Furthermore, no words descending from the same roots, i.e. the paronyms of these words, are found in the list of MMCs. Thus, for

the purpose of this analysis, those words are effectively absent. That said, there are several others which do find a place in the network. These are *dark, dusk, shadow, shade,* and *shape* as well as all of the *some** and *seem** words.

The word *dark* descends from the Indo-European (IE) root **dher-1** which means "to make muddy; darkness" (Watkins, p. 19). *Dark* is the only descendant of that root found among the list of MMCs. They are four in number—*dark-faced, dark-blue, dark-green*, and *dark-red.* However, in the text network, **dher-1** is not among the most central or core nodes listed in Tables 3 and 4.

The word *dusk* descends from the IE root **dheu-1** which Watkins (ibid, p. 19) defines as —"the base of a wide variety of derivatives meaning 'to rise in a cloud,' as dust, vapor, or smoke, and related to semantic notions of breath, various color adjectives, and forms denoting defective perception or wits." This root is implicated in two MMCs—*dumb-founded* and *dust-bin*. Notably, neither **dheu-1** nor the roots of *founded* and *bin* are among those listed as the most central or core.

The words *shadow* and *shade* both descend from the IE root **skot-** which means "dark, shade" (ibid., p. 82). The only relevant MMC is *overshadowed.* While **skot-** is not included among either the most central or core nodes, the other half of the MMC is. Specifically, *over* descends from the IE root **uper-** which means "over." That root is the second most central (C_d = 17).

The word *shape* descends from the IE root **(s)kep-** which is "the base of words with various technical meanings such as 'to cut', 'to scrape', 'to hack'" (ibid, p. 80). The one relevant MMC is *half-shaped.* While the root **(s)kep-** is not among the most core or central nodes, the other root is. Specifically, the word *half* descends from the IE root **(s)kel-1** which means "to cut" (ibid, p. 79). That root is not among the most core nodes but it is the sixth most central (C_d = 12). Among the MMCs that include it are *half-awake, half-caste, half-cooked, half-English, half-French, half-speed,* and *half-way.*

The words *some* and *seem* descend from the same IE root, **sem-1**, which means "one; .. as one, together with" (ibid, p. 77). While *seem* and its variants do not appear in any MMCs, there are several that include *some*. These include *somebody, somehow, someone, something, sometimes,* and *somewhere,* as well as the title of book mentioned by Marlow—*An Inquiry into some Points of Seamanship*. Notably, this root has the second highest coreness score.

In sum, fewer than half of the seventeen words listed by Stubbs are implicated in the main component of the text network. Of those, just four were associated with nodes that were among the most central or core—*dark*, *shadow*, *shape*, and *some*. But even then only the last was directly related to an influential node while the others garnered their influence by being connected to an influential node. Still, with all those caveats, it's worth noting that the MMCs associated with these four words and their roots comprise a large set that emphasizes the same theme. They are *half-shaped, somebody, someone, somehow, something, sometimes,* and *somewhere.* If we add MMCs that include *half* then there is also *half-awake, half-caste, half-cooked, half-English*, *half-French, half-speed,* and *half-way*, all of which suggest, if nothing else, an air of incompleteness or admixture.

4.2 Some (very) simple frequency data

In this section of the paper, the most relevant portion of Stubbs' analysis is his focus on the "content words" among "the top 50 keywords, which both occur 20 times or more in the novel and are also significantly more frequent than in the reference corpus" (p. 11). Excluding the name *Kurtz* and its possessive form, Stubbs identified thirteen such content words: *seemed* (69), *river* (65), *station* (48), *great* (46), *manager* (42), *earth* (39), *ivory* (31), *pilgrims* (31), *darkness* (25), *bank* (25), *forest* (23), *wilderness* (22), and *cried* (20). The former of these was already discussed. And though it appears in no MMCs, its root, **sem-1**, has the second highest coreness value. *Darkness* was also discussed in the last section. In short, its root is associated with four MMCs and none of the implicated roots are highly central or core. Three of these content words are not implicated in any MMCs, nor are their paronyms. Those are *great, pilgrim,* and *earth.* The remaining eight words—and in some cases their paronyms, as well—are all implicated in the text network, some of them very influentially so. Foremost among these is the word *station* which, by Stubbs' account, appears 48 times. It descends from the IE root **sta-** which means "to stand; with derivatives meaning 'place or thing that is standing" (Watkins, p, 86). This root has the fourth highest degree centrality (C_d = 15) and the MMCs including this content word and its variants are *railway-stations*, *station-house, station-yard, Central station*, and *Inner Station*. Notably, paronyms of *station* are implicated in several other MMCs. They are: *standoffish, (mis)understand, stay-at-home, steering wheel, stern-wheel, store-room, trading post, assistant manager, camp-stool,* and *instead.*

The word *river* descends from the IE root **rei-1** which means "to scratch, tear, cut" (Watkins, p. 73). The word itself is implicated in five MMCs—*up-river, river-bank, river-demon, river-side,* and *river-steamboat* while one of its

paronym is implicated in one other—*tight-ropes*. The root has a degree centrality of seven which is in the lower range of the most central nodes.

The word *manager* descends from the root **man-2** which means "hand" (Watkins, p. 52). The word itself was involved in two MMCs—*assistant-manager* and *general manager*. With a coreness of zero and a degree centrality of two, this root is rightly not considered highly influential. However, the word *assistant* descends from **sta-** which noted above as one of the most influential nodes. The word *ivory* descends from a non-IE root and is implicated in only one MMC—*ivory-country*. The root has a zero value of coreness and a very low degree centrality and is, thus, not very influential.

The word *bank* descends from the IE root **bheg-** which means "to break' (Watkins, p. 9). The only MMCs with which it is associated are *river-bank* and *sandbank*. As shown above, the word *river* is associated with **rei-1** which is among the nodes with greater degree centrality. Thus, the word *riverbank* is important by way of its other root's connection to **rei-1**.The word *forest* descends from the root **dhwer-** which means "door, doorway" and which is neither highly central ($C_d = 3$) or core (Coreness = 0). The word *forest* itself is not implicated in any MMC. However, one of its paronyms is *door* and that word appears in three MMCs—*doorstep*, *doorway*, and *double doors*. The word *wilderness* descends from the IE root **welt-** which means "woods" (Watkins, p. 101). Because the root is implicated with only two roots—*wild-eyed* and *wild men*—its degree centrality is very low (Cd= 2). Its coreness is zero and thus the root is not influential. That said, the two roots with which it is associated are **okw-**, which means "to see," and **man-1** which means "man" (Watkins, p. 62, 52). As noted in Table 3, these two roots have degree centrality scores of 10 and 21, respectively. The latter is the highest among all roots. As such, *wilderness* is thematically important due to its connection to two other highly central and thus influential roots.

Finally, the content-word *cried* has a non-IE root and is implicated in only one MMC—*outcry.* The other root there is **ud-** which means "up, out" (Watkins, p. 97). That root has a degree centrality of 12 which is the among the highest. Other MMCs with which **ud-** is associated are *look-out, outbreak, outfit, outlast, outlines, outraged, outside, thereabouts, uttermost, without,* and *dug-out.* Thus again we have a content word identified by Stubbs which is influential because of the connection of its root to a more central one.

In this section Stubbs also made note of the ten most common verb lemmas—*say, see, look, know, come, make, seem, hear, take* and *think.* Among these, three and their paronyms are associated with no MMCs—*say, think*, and *know*. A fourth verb, *seem* has already been discussed. As shown previously, its paronyms include *some*, the root of which has the second highest coreness. None of the roots associated with the remaining six verbs are either highly central or core. However, at least half are linked to roots that are. For example, the verb *see* descends from **sekw-2** which means "to perceive, see" (Watkins, p. 77). The two MMCs with which the root is associated are *unforeseen, foresight, foresaw,* and *insight.* Notably, these stand in direct opposition to the over-arching themes of uncertainty, vagueness, and unreliable knowledge. Although the root **sekw-2** is neither highly central or core in the text network, three of the words are associated with the root **per1**, the one from which *fore* descends. According to Watkins (p. 67), this root has over 100 derivatives and meanings including "forward, through, in front of, before, early, first, and chief", among others.

The verb *look* has no IE root and is associated with only one MMC—*outlook.* As noted above, the root of *out* is **ud-**, one of the most central. The verb *come* descends from the IE root **gwa-** which means "to come, go" (ibid, p. 34). The two MMCs with which it is associated are *overcome* and *coming-to.* The two other roots with which **gwa-** is associated are **uper-** and **de-**. The former means "over" and is one of the most central. The latter is a demonstrative stem which is found in the lower tier of the most core nodes.

The verb *make* descends from the IE root **mag-** which means "to kneed, fashion, fit" (p. 51) and it or its paronyms are associated with three MMCs—*sailmaker, boiler-maker*, and *brickmaker*. Neither **mag-** itself nor the three roots to which it is linked are either highly central or core.

The verb *hear* descends from the IE root **kous-** which means "to hear" (ibid., p. 45) and it is associated with one MMC—*overheard*. While **kous-** is not central or core, the other one is **uper-** and it is highly central.

Finally, the verb *take* descends from the IE root **tak-2** which means "take" (ibid., p. 92). The two MMCs with which it is associated are *overtaking* and *undertaken*. The two other roots with which it is associated are **uper-** and **ndher-**. As noted before, the former of these two is second most central ($C_d = 17$).

4.3 Collocations: Word and Words

In this section Stubbs draws several important conclusions about the text's key themes from an examination of selected collocations. First among these is the word *green*. He notes that this word is "usually associated with death, decay, and

desolation." He adds that "It sprouts through the stones in the city of the dead…and through the bones of a dead man; old machinery is abandoned in it."

The word *green* descends from the IE root **ghre-** which means "to grow, become green" (Watkins, p. 32). *Green* and its paronyms are associated with five MMCs—*dark-green, grass-roofs, green-lined, overgrown,* and *undergrowth*. As it pertains to network collocation, i.e. roots that are adjacent or directly linked to one another, neither *green* nor its paronyms are associated with either death, decay, or desolation. Possibly only the *dark* in *dark-green* suggests this. Otherwise we see spatial and geometrical associations, i.e. *under*, *over*, *line*, and *roof* which perhaps implies *over*. Stubbs also takes note of the collocates of a group of words—*glitter, gleam, glisten*, and *glint*:

> The words GLITTER <14>, GLEAM <8>, GLISTEN <3> and GLINT <2> connote things which are ominous and dangerous: GLITTER collocates withdark, sombre, gloom, and the infernal stream; GLEAM collocates with blood and fire; people's eyes glitter, glisten and gleam; arrows glint when they are being shot at Marlow. Some such associations are signalled explicitly in the text. In Brussels, Marlow looks at the coloured patches on a map of colonial countries. Later, he meets the Russian harlequin figure whose clothes are covered with coloured patches… (p. 15).

One thing that Stubbs does not mention here is that all four of these words are paronyms: they are derivatives are the same IE root **ghel-2**, which means "to shine; with derivatives referring to colors, bright materials, gold (probably 'yellow metal')…" (Watkins, p. 29). Interestingly, none of the four words mentioned by Stubbs are associated with any MMCs. However, several other paronyms are: *eye-glass, Golden Hind, gold-rimmed*, and *yellow metal*. Note that one of these is broadly consonant with Stubbs observation. He remarked that "people's eyes glitter." Among the MMCs we find the hyphenated-compound *eye-glasses* which, as we see, is composed of *eye* and a paronym of *glitter*. That said, none of the four MMCs shows this root associated with anything "ominous" or "dangerous."

4.4 Phraseology: Words and Grammar

In this section Stubbs makes especial note of some of Conrad's widely-recognized "recurrent lexico-grammatical patterns" (Stubbs, 2005, p. 15). One of these patterns is particularly relevant to this study: Conrad's use of "a large number of words with negative prefixes" (ibid). Most of these are adjectives like *impossible, uneasy, unexpected, unearthly*, and *unsound* among many others. Stubbs also notes that there are also fifty words ending with the suffix *–less*, as well as 500+ occurrences of *no, not, never, nothing, nobody,* and *nowhere*. The "total frequency of these negatives," he says, is over 800, approximately 2% of the total.

From a text network perspective we note that all of the negative prefixes that Stubbs mentions and all of the *no*-words descend from the IE root **ne** which means "not" (Watkins, p. 59). The MMCs with which it is associated are only four in number—*nevertheless, nobody, nothing,* and *nowhere*. This root is the seventh most core of the 312 nodes in the network's main component and among the most central, too, with a degree centrality of six.

4.5 Summary of Results

On the whole there is a great deal of overlap among the findings reported by Stubbs in four sections of his study and those reported herein.

4.5.1 Vague impressions and unreliable knowledge

Words involving *some* and *seem* were said by Stubbs to be very frequent and indicative of the theme of vagueness and unreliable knowledge. Our text network analysis revealed these words to be associated with the second most core node in the network. Further, the roots associated with three other such vague words—*shape, shade, and shadow*—were connected to the second and sixth most influential roots as measured by degree centrality.

4.5.2 Frequency data

Stubbs identified several frequently occurring "content words." None of these was linked by him directly to the major theme of vagueness and unreliability. Rather they were notable for occurring much more frequently in the text than in fiction specifically or in the British National Corpus (BNC) more generally. The text network analysis found the etymological roots of several of these words either have high degree centrality (e.g. *station*) or coreness (e.g. seem), or are linked directly to those that are (e.g. *wild man*, *assistant manager*, and *outcry*). The same was found to be the case for the ten most common verb lemmas.

4.5.3 Collocations

Stubbs reported here that the word *grass* was collocated with others that suggested "death, decay, and desolation" while the words *gleam, glint, glisten,* and *glint*—all of which descend from the same IE root—were associated with

things which were "ominous and dangerous." This was one instance where the results of the network analysis did not confirm Stubbs'. We found that while the roots associated with these words were modestly central, their paronyms occurred in MMCs more frequently and to greater effect than did the ones identified by Stubbs. In one case there was mild support for his hypothesis while in the other, a theme emerged which seems unrelated to his.

4.5.4 Phraseology

Stubbs reported on the very high frequency of negative words in *HoD*. Our results placed all of these words squarely within the core of the text network. Thus again, there was confirmation of Stubbs' findings.

In the final analytical section of this paper we report on findings that are unique to the network analytical approach, results that highlight findings not reported by Stubbs. We accomplish this by identifying the MMCs associated with the four most central nodes, a little more than the top 1% in terms of degree centrality.

5. Extraction of Meaning via Connected Ego Networks

The fourth and final step of a network text analysis to be *extraction* of meaning. This is accomplished by a variety of means and varies according to the analyst's methods and goals. In this study, the extraction process begins with the identification of the most influential nodes in the text network. And here, for simplicity's sake, we define influence by degree centrality of the etymological roots, which, as we know are the nodes in the network. We then examine the MMCs associated with each of these nodes. Of particular importance are those that link or bridge one or more roots or nodes. Determining these linkages requires first creating "ego networks." In SNA an ego network consists of a focal node (the "ego"), the nodes that are directly connected to it (aka, the "alters) plus any connections or ties among the alters (Ego Networks, 1998). The network graph depicted in Figure 4 contains the ego networks for the four nodes (etymological roots) with the highest degree centrality scores, i.e. greater than or equal to 15. Omitted from the graph are those nodes with only one link, also known as pendants. On the links between nodes appear the MMCs associated with each pair or group of nodes. The etymological roots in the network—in order of degree centrality scores—are **man-1** (man), **uper-** (over), **upo-** (up, over, above), and **sta-** (stand). Notably they are all Indo-European roots. Thus, at the broadest level, we could say that the three major themes in HoD concern (1) men (2) that which is over, up above, or superior (3) and that which stands or is standing or stationary. A very simple combination of these elements could lead us to conclude, or to hypothesize, that the text deals with a *man or men* who *stand(s) up* (*up* to?) or *stands over* another *man* or *men*. And if we take note of the fact that both *superior and supremacy* descend from **uper-**, then three other possibilities—among many—are a *superior man stands up*, a *superior man stands over* (other) *men* and the *supremacy* of one *man over* another *man* or *men.*

The next step is to examine the specific MMCs associated with each node. MMCs associated with the former root, and depicted in the diagram, are *seaman(ship), fool-helmsman, foreman, gentleman, mankind, headman, handy-men,* and *witch-man*. Other MMCs not shown in the diagram include *Dutchman, Englishman, fireman, man-of-war, poleman, policeman, white man, wild men,* and *black men*. Differentiation among "men" stands out as a clear theme here. This list of men differ not just by skin color (*white man, black men*) but also by nationality (*Dutchman, Englishman*), by occupation or function (*seaman, helmsman, handy-men, fireman, poleman, policeman*), by rank or position (*foreman, headman*), and by disposition (*gentleman, wild men*). Again, though it is not mandatory to do so, it is possible to place value judgments on these essentially categorical distinctions, i.e. that *white* is superior to *black*, that the *gentleman* is superior to the *wild men*, or that the Western and "civilized" (*Englishman, Dutchman*) is superior to superstitious (*witch-man*) and uncivilized (*wild men*). All of these need not be offensive distinctions, however. For example, in a social organization, team, or workplace setting, the *headman* or *foreman* is very much the *superior* of those assigned to tasks narrower in scope or importance.

The MMCs associated with **uper-** and appearing in the diagram include *over-sea, supernatural, overfed, overgrown, overtaking, overland, and summing-up*. Others not shown in diagram include *overboard, overcast, overhanging, overhung, overheard, overheated, overpowering, overshadowed,* and *overwhelmed/ing.* Broadly, what is emphasized here are things or states of being that stand over and above others. More specifically, we might note that a three-fold division is apparent among these words. In the first would be those concerned with the physical and material world, with time and or the three dimensions of space, e.g. *over-sea, overfed, overgrown, overtaking, overland, overboard, overcast, overhanging/hung*, and *overheated*. A second and much smaller category deals with mostly emotional and cognitive acts and reactions, e.g. *overwhelmed/ing*, *overheard, overcome* ("…as if overcome with great weariness"), *overpowering* ("what made this emotion so overpowering was…") and *summing-up* ("I like to think my summing-up would not have been a word of careless contempt"). A final grouping—if such a word may be used for a set of one—is world that lies outside of and perhaps above the realm of the five senses, emotion and cognition—the *supernatural.* There is clearly a hierarchy here in terms of physicality/action vis-à-vis affect and cognition which are generally seen

to be actions' motivation. Another hierarchical or value-laden distinction could be that between the *natural* and *supernatural*.

The third most central root in the main component is **upo-** which means "up, over." MMCs associated with it in this graph are *up-country* and *summing-up*. Other MMCs not depicted include *open-mouthed, pent-up, smash-up, turned-up, uphill, upkeep, upon, up-river, uproar, upset, up-stairs, get-up,* and *got-up.* Again, in the broadest sense we see emphasis on that which is *up* or above relative to that which is below, be that in the physical, affective, or cognitive realms of action

Finally, the fourth most central root is **sta-** which means "stand." MMCs associated with it are *storeroom, station-house, trading post, railway-stations, assistant manager, Inner Station, understand*, and *misunderstanding*. Other MMCs not depicted in Figure 4 include *stand-offish, station-yard, stay-at-home, steering-wheel, stern-wheel, camp stool*, and *Central Station*. Clearly emphasized here are places and positions, as well as things that stand or that are stable. And there is implied verticality here, too. When those things are working and function, they could be said to stand up. When they are neglected, broken, or useless, they could be said to have fallen (down) into disrepair.

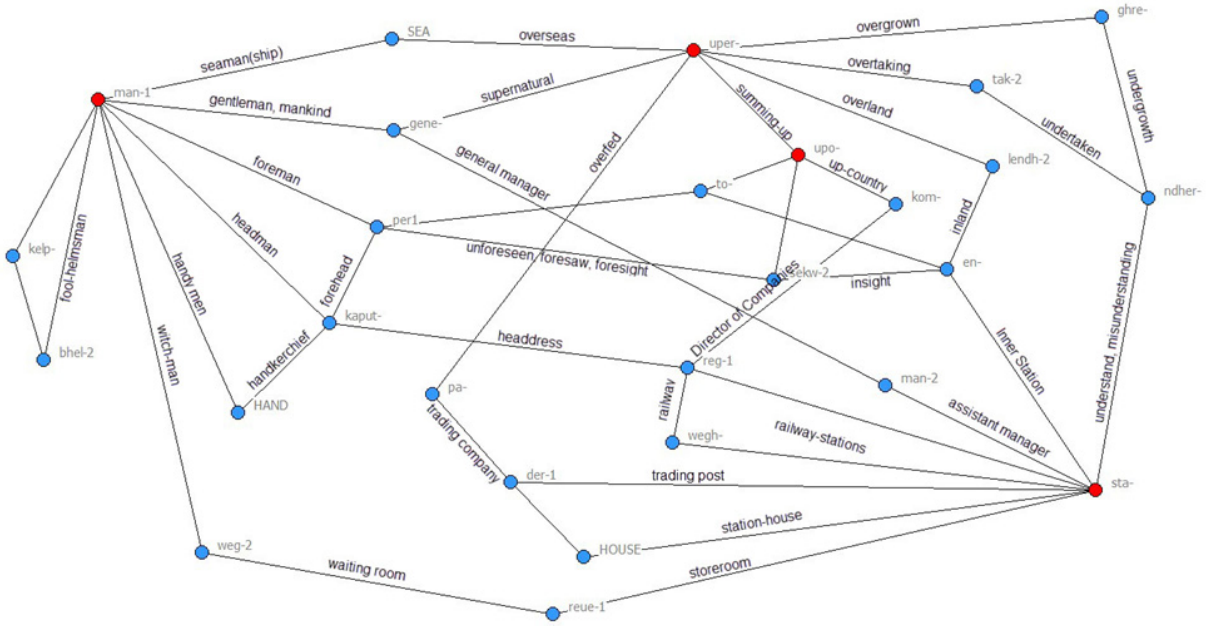

Figure 4. Ego Network Formed by the Four Etymological Roots with the Largest Degree Centrality

6. Conclusion

Notably, none of these broad themes described in the last section are evident anywhere in Stubbs' corpus stylistic analysis of *Heart of Darkness*. That should not, however, be taken as any kind of criticism of Stubbs' approach or results. Rather, it suggests that the extraction of meaning displayed here is more of broad-brush strokes than fine lines. But as already shown, one can drill down into parts of the network and find results or draw inferences about very specific matters in the text. Our method also permits analysts to place words, concepts, and themes in a unified context. That is to say, the text network approach described here displays specific chains or links among words and groups of words in a way that traditional frequency-based content analysis and stylistics can not. Further, it allows not just the representation of those relationships, but also their quantification along such dimensions as strength, connectedness, and distance. As such it would seem that network text analysis can serve as a complement to corpus stylistic analysis rather than as a replacement. It may also have uses of its own, uses for which in-depth analysis is neither practical nor warranted, e.g. the comparative and quantitative analysis of large samples of texts like Hunter's (2014) network text analysis of 150 contemporary screenplays. That study reported the results of a network text analysis of 75 award-winning screenplays and 75 unproduced (amateur) screenplays. Text network variables like network size (as measured by the number of etymological roots) and coreness were calculated for all 150 screenplays. He reported that award-winners were found to have significantly larger networks with smaller cores compared to unproduced screenplays. A corpus stylistic analysis of this data set would have been both impractical—because of the size of the

sample—and difficult to quantify in terms generally applicable to the entire sample. However, in the case of in-depth analyses of single texts, the methods are, we think, more nearly comparable and very complementary. Future research should undertake to determine when and where each approach is more applicable or warranted.

References

Bakker, R. R. (1987). *Knowledge Graphs: representation and structuring of scientific knowledge*.

Bartleby.com. (2013). *Inaugural Addresses of the Presidents of the United States*. Retrieved from http://www.bartleby.com/124/

Biber, D. (2011). Corpus linguistics and the study of literature: Back to the future?. *Scientific Study of Literature*, *1*(1), 15-23. http://dx.doi.org/10.1075/ssol.1.1.02bib

Borgatti, S.P., Everett, M.G. and Freeman, L.C. 2002. Ucinet for Windows: Software for Social Network Analysis. Harvard, MA: Analytic Technologies.

Carley, K. M. (1997). Network text analysis: The network position of concepts. *Text analysis for the social sciences: Methods for drawing statistical inferences from texts and transcripts*, 79-100.

Carley, K., & Palmquist, M. (1992). Extracting, representing, and analyzing mental models. *Social forces*, *70*(3), 601-636. doi: 10.1093/sf/70.3.601

Carley, Kathleen M & Columbus, Dave & Landwehr, Peter. (2013). AutoMap User's Guide 2013. *Carnegie Mellon University, School of Computer Science, Institute for Software Research, Technical Report, CMU-ISR-13-105*

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407. http://psycnet.apa.org/doi/10.1037/0033-295X.82.6.407

Conrad, J. (1899/2010). *Heart of Darkness*. Macmillan.

Corman, S. R., Kuhn, T., McPhee, R. D., & Dooley, K. J. (2002). Studying Complex Discursive Systems. *Human communication research*, *28*(2), 157-206. http:// 10.1111/j.1468-2958.2002.tb00802.x

Danowski, J. A. (1993). Network analysis of message content. *Progress in communication sciences*, *12*, 198-221.

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.

Diesner, J. (2013). From Texts to Networks: Detecting and Managing the Impact of Methodological Choices for Extracting Network Data from Text Data. *KI-Künstliche Intelligenz*, *27*(1), 75-78. http:// 10.1007/s13218-012-0225-0

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.

Ego Networks (5 August, 1998). *Social Network Analysis Instructional Website*. Retrieved from

http://www.analytictech.com/networks/egonet.htm

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. http://dx.doi.org/10.1075/ijcl.14.4.03fis

Fish, S. (1973). What is stylistics and why are they saying such terrible things about it? *Approaches to poetics*, 109-52.

Fish, S. (1979). What is Stylistics and Why Are They Saying Such Terrible Things about It?-Part II. *Boundary 2*, 129-146. http://dx.doi.org/10.2307/303144

Hunter, S. (2014a). A Novel Method of Network Text Analysis. *Open Journal of Modern Linguistics*, *4*(02), 350-66. 10.4236/ojml.2014.42028

Hunter, S. D. (2014b). A Semi-Automated Method of Network Text Analysis Applied to 150 Original Screenplays. In *Proceedings of the Joint Workshop on Social Dynamics and Personal Attributes in Social Media* (pp. 68-76).

Hunter, S., & Smith, S. (2013). Thematic and Lexical Repetition in a Contemporary Screenplay. *Open Journal of Modern Linguistics*, *3*(01), 9. 10.4236/ojml.2013.31002

Kleinnijenhuis, J., De Ridder, J. A., & Rietberg, E. M. (1997). Reasoning in economic discourse: An application of the network approach to the Dutch press. *Text analysis for the social sciences: Methods for drawing statistical inferences from texts and transcripts*, 191-207.

Mackay, R. (1996). Mything the point: A critique of objective stylistics. *Language & Communication*, *16*(1), 81-93. http://10.1016/0271-5309(95)00008-9

Mackay, R. (1999). There goes the other foot-a reply to Short et al. *Language and Literature*, *8*(1), 59-66. http://10.1177/096394709900800104

Novak, J. D. (1990). Concept mapping: A useful tool for science education. *Journal of research in science teaching*, *27*(10), 937-949. http:// 10.1002/tea.3660271003

O'Halloran, K. (2007). The subconscious in James Joyce's *Eveline*: a corpus stylistic analysis that chews on the Fish hook'. *Language and Literature*, *16*(3), 227-244. http:// 10.1177/0963947007072847

Popping, R. (2000). *Computer-assisted text analysis*. Sage.

Popping, R. (2003). Knowledge graphs and network text analysis. *Social Science Information*, *42*(1), 91-106. http:// 10.1177/0539018403042001798

Popping, R., & Roberts, C. W. (1997). *Network approaches in text analysis* (pp. 381-389). Springer Berlin Heidelberg.

Roberts, C. W. (Ed.). (1997). *Text analysis for the social sciences: Methods for drawing statistical inferences from texts and transcripts*. Lawrence Erlbaum Associates.

Smith, N., Hoffmann, S., & Rayson, P. (2008). Corpus tools and methods, today and tomorrow: Incorporating linguists' manual annotations. *Literary and Linguistic Computing*, *23*(2), 163-180. http:// 10.1093/llc/fqn004

Sowa, J. F. (1992). Conceptual graphs as a universal knowledge representation. *Computers & Mathematics with Applications*, *23*(2), 75-93. http:// 10.1016/0898-1221(92)90137-7

Stubbs, M. (2001). *Words and phrases: Corpus studies of lexical semantics*. Blackwell Publishers.

Stubbs, M. (2005). Conrad in the computer: examples of quantitative stylistic methods. *Language and Literature*, *14*(1), 5-24. http:// 10.1177/0963947005048873

van Atteveldt, W., Kleinnijenhuis, J., & Ruigrok, N. (2008). Parsing, semantic networks, and political authority using syntactic analysis to extract semantic relations from Dutch newspaper articles. *Political Analysis*, *16*(4), 428-446. http:// 10.1093/pan/mpn006

Wasserman, S. (1994). *Social network analysis: Methods and applications* (Vol. 8). Cambridge university press. http://dx.doi.org/10.1017/CBO9780511815478

Watkins, C. (Ed.). (2011). *The American heritage dictionary of Indo-European roots*. Houghton Mifflin Harcourt, 3rd Edition.

Widdowson, H. G. (2008). The novel features of text. Corpus analysis and stylistics. *Language and Computers*, *64*(1), 293-304.