
Timothy S. Rich
Political Science Department, Western Kentucky University, USA

ARTICLE INFO
Article history:
Available online 18 April 2014

Keywords:
North Korea
Kim Jong Un
KCNA
Automated content analysis

ABSTRACT
How did North Korea initially frame coverage of the “Great Successor” Kim Jong Un for an international audience? This paper argues that North Korea’s daily English language news reports, while commonly dismissed as purely propaganda, provides potential insights into such framing. Through automated content analysis of daily news reports from 2010 through 2011 coupled with regression analysis, this analysis both suggests an increased focus on Kim Jong Un’s formal positions and less on his pedigree.

1. Introduction

Little is known about Kim Jong Un who took the reins of leadership in North Korea following his father’s death in December of 2011. Prior to 2010, the youngest son of Kim Jong Il was not referenced at all in North Korea’s English language news, this despite appearing earlier in domestic propaganda. While observers continue to debate whether the new leader will usher in political and economic reforms, no rigorous analysis to date attempts to uncover how the regime framed coverage of Kim Jong Un prior to succession. Rather, most coverage prior to Kim Jong Il’s death focused on assumptions of succession instead North Korea’s own accounts. The inherent problem in such an approach is that outside scholars risk missing hints as to the future leadership in the country by dismissing North Korea’s official news as simply propaganda mouthpiece void of any other value.

In this paper I propose using of North Korea’s own news sources as a means to glean insight into how Kim Jong Un was initially positioned as successor prior to his father’s death. In particular I employ automated content analysis—computer software to analyze text—to identify patterns in North Korean rhetoric. Automated content analysis remains rare in political science, largely limited to electoral politics in Western democracies. Using daily reports from the Korean Central News Agency (KCNA) for 2010 and 2011 provides a unique means to empirically measure whether references to Kim Jong Un correlate with common themes within North Korean rhetoric as well as with leaders past and present within the Hermit Kingdom. Through this analysis, the goal is to gain insight into the internal dynamics of North Korea leading up to succession as well generate expectations regarding future actions by the Great Successor.

Predicting North Korea’s collapse has been somewhat a favorite pastime of academic observers. Western predictions of North Korea’s imminent collapse emerged shortly after Kim Il Sung’s death in 1994 and the succession of Kim Jong Il. Scholars certain of the country’s impending regime change identified various rationales behind their position, but surprisingly few appeared to update such predictions as the regime muddled through the end of the Cold War, the famine referred domestically as The Arduous March, and an increasingly interconnected geopolitical environment.1 Despite

1 Scott Snyder commented that “if people keep on predicting that the North is about to collapse, well, one of these days they’ll be right” (Demick, 2004).
economic and political crises, not only has the regime failed to collapse, but Kim Jong Il appears to have successfully
positioned his youngest son Kim Jong Un as his successor without evidence of a backlash or efforts to remove the youngest
Kim short after his father’s death. No evidence to date suggests that Kim Jong Un had any serious competition for this
position, even if de facto leadership remained for awhile in the hands of experienced leadership such as his uncle Jang Sung Taek. Existing evidence suggests a near complete failure of North Korea’s planned economy in the years leading up to
succession, with only limited efforts at economic reform (Haggard and Noland, 2009). Scholars suggest that significant
domestic wrangling was necessary to promote the youngest Kim to the helm, who apparently has no political or military
experience despite propaganda claims to the counter. Yet, most analyses of North Korea rely on disconnected and often
difficult to confirm pieces of evidence rather than attempt to uncover meaning within the breadth of material North Korea
produces for an external audience.

While the amount of information crossing North Korean borders increases, firsthand knowledge by scholars has not made
similar gains. Direct access to the country remains largely confined to humanitarian agencies or government endorsed tours,
with traditional forms of fieldwork limited to interviews and surveys of refugees in China or South Korea. Speculation in the
absence of empirical data unfortunately dominates much of the research, although admittedly this imbalance is slowly
changing. Furthermore, insights from refugees have improved our understanding of everyday politics and economics in the
country, but say little about elite politics under than anecdotal impressions of public opinion. Unlike the heyday of the Cold
War where several well-connected officials in the communist world defected and thus aided American intelligence efforts,
few North Korean elite have defected with the last high level defection—Hwang Jang-Yop—in 1997. Thus the ability to
confirm perceived dynamics of North Korea’s elite politics has advanced only marginally in recent years. A growing literature
increases our knowledge of North Korea in general and especially everyday life in the country (e.g. Cha and Kang, 2003; 
Haggard and Noland, 2009; Hassig and Oh, 2009; Noland, 2000; Park, 2002), yet with each new challenge we are reminded of
our ignorance of elite politics.

Instead of attempting to sift through existing media reports with their various ideological filters, this paper suggests using
one of North Korea’s own sources as a means to track rhetoric and extract meaning. Using daily news reports from the Korean
Central News Agency (KCNA) provides an empirical base for making claims on how Pyongyang frames coverage around Kim
Jong Un and its consistency over time. Created in 1946, the KCNA is the official mouthpiece of the Korea Workers Party (KWP),
with daily English news via the web since 1997. While commonly dismissed as mere propaganda, automated content ana-
lysis—using computer software to code and analyze text—provides a means to uncover patterns in such sources that the
naked eye misses. Through descriptive and inferential statistics, we should gain a greater understanding of how the gov-
ernment intended to present Kim Jong Un to a Western audience. Admittedly using the English language KCNA reports likely
differs from Korean language materials both in content and intended audience (Poneman et al., 2004). However, relying on
North Korea’s own words in English intended for a foreign audience avoids the difficulty of interpreting a Korean translation
through an ideological lens. This source combined with automated analysis also allows for a more precise identification of
shifts in rhetoric over time, rather than attempting to perceive such shifts through interpretive (and possibly idiosyncratic)
reading of the news sources. In sum, automated content analysis of KCNA news reports provides a means to both empirically
verify conventional wisdom as well as uncover patterns potentially missed by outside observers.

This paper will first highlight how automated content analysis can be applied to coverage of North Korea. A brief intro-
duction to personality cults is followed by an introduction of Kim Jong Un. Next, descriptive and inferential statistics identify
patterns within KCNA reports in regards to when Kim Jong Un is mentioned. Ultimately this analysis highlights that while a
casual view of the KCNA reports may leave one dismissing it as undifferentiated propaganda, patterns uncovered by computer
assisted content analysis suggests an intentional contrast in how Kim Jong Un is presented to a Western audience compared
to domestic propaganda.

2. Applying content analysis to research on North Korea

Social scientists have applied various means as an attempt to explicitly classify parts of text and in the process decipher
meaning (Berelson, 1952; Holsti, 1969). Following similar methods of historians and journalists, early works relied on human
coding of individual texts. The reliance on human coding presents two serious restrictions for large content analysis projects.
Hand coding is a time-intensive process, thus individual scholars either find it necessary to collaborate with others or hire
assistants or reluctantly restrict the scope of the texts under analysis, for example covering a few months of news coverage or
a subset of newspapers rather than all available sources for a time period of interest. Such a restriction as sampling from a
body of texts creates potential selection bias concerns while not taking advantage of the full potential of the data source. In
the absence of the time constraints, maintaining consistency both over time as well as among different coders further restricts
content analysis efforts (Krippendorff, 2004; Neuendorff, 2002; Weber, 1990). Computer assisted content analysis overcomes
both of these concerns as the amount of text that can be analyzed is limited only by computer memory requirements. Meanwhile consistency in coding is ensured through human designed dictionaries or complex algorithms that automatically
classify parts of text.

While content analysis has a long history in political science, including analyses of non-democratic contexts (e.g. Mills,
1985), automated methods surprisingly remain limited largely to areas such as extracting policy positions in Western de-
mocracies (Benoit and Laver, 2008; Laver et al., 2003). A Google Scholar search of “automated content analysis” generates 827
hits, 297 of which also mention politics or political science (Date accessed 12/8/2012). Much of this literature is only from the

While various sources for content analysis are available, relying on North Korea’s English language KCNA news reports prevents fewer challenges in many respects. Framing has long been known to influence perceptions of news coverage and how this translates to public understanding and discourse (Scheufele, 2000; Tuchman, 1978). Without direct data from North Korea it is difficult to separate the government’s intended framing from outsider interpretation. Explicitly written by North Korean officials for a foreign audience, this source does not require translation from Korean and thus intended meanings are less likely to be lost through multiple translations. In other words, KCNA reports should accurately capture North Korea’s desired framing of issues rather than, for example, South Korean interpretation of North Korea’s framing. Similarly, relying on English language sources written about North Korea are subject to similar ideological filtering (Hachten, 1999; Herman, 1993). North Korean editorials theoretically capture more explicitly the government’s position (Kim, 1998: 44), yet a broader set of texts potentially uncover less explicit framing, including patterns in which the writers were unaware.

This is not to suggest that the KCNA reports are factually accurate. Even a cursory analysis confirms the frequency of dubious claims, especially regarding the revolutionary credentials of the Kim family and the miraculous events surrounding the birth of Kim Jong Il (in North Korea, rather than his actual birthplace in Russia). On November 29, 2012 the KCNA reported “Lair of King Tongmyong’s Unicorn Reconfirmed in DPRK”, a report covered widely in Western news as yet another example of outrageous claims emanating from Pyongyang.2 Rather, the assumption here is that these reports are consistent with North Korean intentions to shape perceptions of their government and country, regardless of whether the recipients actually believe such accounts. Furthermore no assumption is made regarding a correlation between domestic Korean language propaganda and KCNA news for a domestic audience, assuming instead that framing likely differs considerably between the two sources. For example, North Korea’s international propaganda commonly portrays the country as vulnerable, whereas the domestic variant emphasizes the political superiority and military strength of the regime (Hecker, 2010). Overall, employing automated content analysis of English language KCNA news reports provides an explicit means to measure North Korean rhetoric, using publicly available resources originating from North Korea.

2.1. Manufacturing images of the Kims

The construction of a master narrative regarding leaders and their role in social harmony remains a common tool for social control in authoritarian regimes. These master narratives, while highlighting the main leader, usually contain references to a supporting cast of similar like-minded revolutionaries. Even in the heyday of the personality cult around Mao Zedong, other Chinese leaders were displayed prominently in Chinese propaganda, although these other leaders were later purged. Even the famed image of Comrade Mao at Tiananmen Square was once paired with another image, that of General Zhu De. North Korea deviates from this pattern with few references to any figures separate from the Kim family and as such paved the way for hereditary transfers of power.

In Kim II Sung’s official biography under the section The Great Manyongdae Family includes the line “there are many families in the world that produced (a) great man. But there is no such a great family as Marshal Kim Jong-II’s family, (all of) whose members were famous as patriots, generation after generation”. Kim II Sung is presented as an idealized symbol of Korea and the Korean family, with an unbounded charisma that justifies the leader’s departure from the limits of official institutional roles. By 1974, the year Kim Jong Il was officially designated as the successor, similar references to his brilliance and revolutionary credentials have emerged. Kim II Sung of course vehemently denied any intentions of extending the personality cult to his son (Martin, 2004).

Even particular terminology appears reserved solely for the Kim’s dynasty, with domestic materials printing their names in boldface and quotes in a separate font from the main text (Beck, 2012: 67). As both the “sun of the nation” and the Eternal President, the eldest Kim’s elevated status is entrenched. Similar phraseology follows Kim Jong II as he is referred to as the brain of the Korean people. The public mourning which followed the deaths of each leader (Kim II Sung on July 8, 1994 and

---

2 North Korea’s English language news reported finding the remains of a unicorn lair from the Koguryo period (277BC-668AD), with this story quickly spreading through Western news outlets. Further investigation suggests a poor English translation to blame, with the story not referencing unicorns at all, but referring to the residence of a mythical Korean creature (kirin) and likely intended not to be interpreted literally but to imply the wisdom of the Kim leadership.
Kim Jong Il on December 17, 2011), while melodramatic by Western standards, thus should not have been surprising considering the image cultivated under both leaders.

The more bizarre claims of the Kims’ propaganda machine receive considerable Western coverage despite such references often absent in existing English language materials from North Korea. Similarly, bizarre stories within the English text, such as references to a unicorn lair, are largely the result of translation errors. Distinctions between domestic and international propaganda suggests a realization of the need to project the North Korean leadership as comparatively less hardline and absent of the most grandiose of claims. Furthermore few appear to analyze the evolution of this personality cult in recent years or its extension to Kim Jong Un.

Despite ample evidence to the contrary, the government continues to both produce materials glorifying the Kims in multiple languages and claim the establishment of pro-Kim organizations abroad. A cursory view of the propaganda for international consumption suggests a connection of the Kim family to responsible actions in the international community, while domestic propaganda likely serves a much different purpose.

Although most assume a distinction between propaganda for domestic versus international consumption, few attempt to uncover deeper meaning within the latter. Considering the cost of developing foreign language materials and the desire for greater international attention, official North Korean resources likely have explicit and implicit messages beyond the standard façade of praise to the Kim family. If so, dismissing the source as a generic template seems unwarranted. Identifying correlations among themes within North Korea’s English language news potentially provides additional leverage. If the propaganda is in fact a generic template, we should see little if any variation on themes. Similarly, if Kim Jong Un is just another name in a list of officials or a face in a crowd, references to the youngest Kim should not correlate with common themes in North Korean materials. A consistent correlation among themes, while not a smoking gun, provides additional evidence suggestive of a level of intentionality within the reports.

3. Introducing Kim Jong Un

Succession issues in dictatorships are always precarious endeavors, threatening the stability of the regime. Similarly, few authoritarian leaders, despite attempts to hand-pick a successor, are successful in passing the torch as envisioned.3 While a full understanding of the dynamics of succession in North Korea is likely to only be known years from now, scholars have suggested the elevation of Kim Jong Un as the Kim Jong Il’s handpicked successor involved considerable wrangling among North Korean elite. McEachern (2009) suggests that diverging policy preferences are emerging within the military and party and this likely extended to preferences in future leadership. Little was known about the youngest son of Dear Leader, with intermitted references in domestic propaganda, and no references in the KCNA before September 27, 2010. So little was known about Kim Jong Un in the outside world that South Korea’s intelligence service acknowledged that they had been misspelling his name (Oh and Hassig, 2010: 96). Conventional wisdom was that Kim Jong Il preferred his youngest son both due to the temperament of the latter and his physical similarities to the country’s founder, Kim Il Sung, but also in response to the eldest son’s (Kim Jong Nam) embarrassing attempt to enter Japan on a fabricated passport in May of 2001. With no known military experience and without the decades of tutelage like his father, the promotion of Kim Jong Un likely required implicit support from both the military and party officials. Similarly, without an apprenticeship or other practical experience, many assumed that North Korea propaganda would highlight the leader’s biological pedigree (Jeon, 2011: 94–95). Yet, just as we have come to expect divergent messages in domestic versus international propaganda, the framing around Kim Jong Un may also differ, especially with the younger Kim remains shrouded in secrecy.

It is assumed that great maneuvering was necessary to promote an inexperienced youth into a position of power, including showing resilience against the South.4 As such several commentators have suggested that the sinking of the South Korean naval vessel the Cheonan on March 26, 2010 may have been related to the succession issue. Similarly, after Kim Jong Un’s accession to the Vice Chairman of the Central Military Commission in 2010, a similar show of force may have been seen as necessary to shore up military support. On November 23, 2010, North Korea shelled military and civilian sites on the South Korean island of Yeonpyeong, just south of the Northern Limit Line (NLL), resulting in two killed and fifteen wounded ROK marines as well as two civilians. Admittedly, inter-Korean military skirmishes are not unusual, with confrontations in 1998, 1999, 2002, and 2009. Yet, the timing of these escalations suggested a possible rationale beyond traditional inter-Korean dynamics.

4. Expectations

This paper suggests that common themes within North Korean news are implicitly connected to Kim Jong Un. Besides giving some indication on the youngest Kim’s formal and informal roles, themes within the KCNA should provide evidence of the general contours of North Korean policy and in part indicate whether we should expect a continuation of rhetoric and policies attached to his father and grandfather or a subtle shift towards, if not reform, a de-emphasis of controversial issues.

3 For example, Mao Zedong’s handpicked successor Hua Guofeng fell from power within five years of Mao’s death.

4 Similarly, belligerent actions in 2009 may have been a result of Kim Jong II attempting to win over hardline military officials to back his chosen successor (Klingner, 2010).
Up through the late 1960s North Korea’s economy matched if not surpassed their southern neighbors in terms of per capita GDP, but a far different picture emerges today. While small improvements in the economy and a desire to attract foreign direct investment may suggest a meager attempt at Chinese-style economic reforms, the government’s inability to spawn consistent economic growth has lead the country to fall even further behind. Now, the Hermit Kingdom is left with an economy less than a tenth of the size of South Korea, with per capita GDP only marginally higher than at time of the armistice ending the Korean War. Thus the expectation is for the KCNA not to emphasize the economy, especially in regards to the youngest Kim.

The intention of other common themes in North Korean rhetoric remains unclear. References to North Korea’s homegrown Juche ideology are common in communication. Juche propaganda, with its focus on self-reliance, foreign intervention and independence (even to the point of withdrawing from the world economy), may still resonate at a certain level domestically. However, these Juche ideals contrast sharply with both North Korea’s increased economic dependence on China, but also with the country’s dependence on the outside world for basic humanitarian assistance. While Kim Il Sung was intimately tied to Juche, the term appears to more boilerplate propaganda under Kim Jong Il. As such, we should expect no significant difference in how the term is used regarding Kim Jong Un.

While military rhetoric and the promotion of songun ("military first") politics provided the foundation for Kim Jong Il’s regime, the expectations around Kim Jong Un are less clear. Kim Jong Un’s appointments to military offices as well as the need for at least implicit military support for his succession would suggest a need for increased military-focused rhetoric, yet such references were already ubiquitous long before the youngest Kim’s presence on the national stage. Similarly, while North Korea’s nuclear program traditionally elicited much attention from the Western press towards Kim Jong Il, a cursory analysis of North Korean materials fail to show any strong effort to tie this program to a particular leader. Furthermore, North Korea’s quest for nuclear capabilities far pre-date Kim Jong Il’s rise to power. Attaching nuclear rhetoric to Kim Jong Un also may signal cause for alarm in a Western audience.

While Kim Il Sung transformed North Korea into what propaganda refers to as the “socialist paradise”, Kim Jong Il also shrouded his legitimacy to rule, in part, on maintaining the socialist path. The symbolic importance of a commitment to socialism may partially explain the leader’s hesitancy to enact much needed economic reforms in that such reforms directly conflict with decades of claiming the superiority of a planned economy. Anti-imperialist rhetoric too maintains an image of North Korea resisting capitalist advances, but it remains unclear whether such terminology has any deeper resonance. Finally, references to reunification have been ubiquitous since the Korean War, albeit reunification on P’yongyang’s terms. Thus the expectation is for none of these terms to be disproportionately tied to a particular leader but rather be considered common speech. Automated content analysis provides a means to empirically analyze the prevalence of these common themes over time and in connection to references to the leadership, identifying patterns potentially missed by the naked eye.

5. Research design

Critics argue that automated analysis misses nuances in the original text, although human coding is not immune to similar errors. Furthermore, sophisticated options in today’s automated analysis software allow one to capture meaning and relations within text at multiple levels while providing explicit coding procedures. The English language KCNA website includes daily news reports and these were collected for 2010 and 2011, using the day as the original unit of analysis. While each day included multiple stories, collecting individual files using the day as the unit of analysis decreased collection time for processing. This collection method also avoids overly weighting short news items and multiple stories in a day that are similar in content and length, with the assumption that each day should hold few inconsistencies. Still, computer software allows for descriptive and regression analyses from the day level down to individual paragraphs and sentences. Because this software goes beyond individual word counts and can identify the repetition of phrases and the proximity of words within a document, descriptive and inferential statistics should provide empirically accurate means to evaluate rhetoric.

Based on previous research on North Korea, I constructed a content analysis dictionary coding both common phrases within North Korean rhetoric as well as references to people and countries. I only include explicit references, although additional diagnostic tests ensured that potentially vague terms were not miscoded. For example, like most North Korean propaganda, the KCNA reports routinely shower praise on Kim Jong Il and his father Kim Il Sung, even after nearly two decades since the latter’s death. While over fifty categorizations were initially coded, I chose to focus on terms that appeared frequently or were, from a Western perspective, expected to be important thematically. In addition, measures of proximity within the texts are justifiable in the inclusion of additional terms for further analysis. Unfortunately this software does not provide a means to measure easily the proximity of selected terms per document. Still, through a combination of statistical measures at multiple levels within the KCNA reports, this analysis suggests that important patterns have been overlooked.

---

5 Some suggest that the North Korean government intentionally overlays the need for assistance in order to encourage future aid shipments, rather than carry the burden of feeding their population themselves on dwindling state resources.
6 This paper uses WordStat software from Provalis (http://www.provalisresearch.com/wordstat/WordstatFeatures.html).
7 For example, several references to America and American were in relation to Latin America. These were recoded so that only references to the US are included for analysis.
6. Analysis

Even a cursory view of the KCNA reveals the rarity of references to Kim Jong Un. Whereas Western analysts presumed an explicit public grooming of the younger Kim, at least in the English language media, the KCNA reports paid little attention to him with no references before mid-2010. Daily references to Kim Jong Un ranged from 0 to 76 in 2010–2011, with an average slightly above one (1.12). When restricted to just 2010, Kim Jong Un is barely mentioned at all, maxing at four references in a day with an average close to zero (0.102). This pales in comparison to references to his father and grandfather in the same time frame. References to Kim Jong Il peaked at 244 (in the days directly after his death) and averaging 9.28 references, while Kim Il Sung peaked at 67 references but averaged 9.28 references over the time period. Not surprisingly, Kim Jong II was mentioned more frequently in 2011, both because of his death but also his trips abroad, averaging 20.84 references per day compared to 12.08 in 2010. References to Kim Il Sung also increase in 2011, from an average of 7.38—11.19. Individual graphs of references to each official over the two years again suggest that Kim Jong Un was not immediately exalted as his pedigree may suggest. While Kim Il Sung and Kim Jong Il references fluctuate considerably over time, the youngest Kim is rarely mentioned at all, only seeing a spike in the period after his father’s death (Fig. 1).

Table 1 presents word or phrase frequencies in the KCNA reports for several common phrases, separated by year. Selected words and phrases with a frequency of a thousand or more were included along with that of Kim Jong Un. None of the themes are particularly surprising, although the change in usage from 2010 to 2011 is surprising in some cases. The frequency of references to Dear Leader nearly doubles in 2011, with a slightly smaller increase for the Eternal President Kim Il Sung. References to the military, socialism, and economics too see a large increase (and Juche a smaller increase) in references in 2011, while the other themes largely stay flat. Meanwhile, despite apparent coverage in domestic propaganda, Kim Jong Un was mentioned only thirty seven times in 2010, increasingly more than twentyfold in 2011, but still failing to hit a thousand references.

A more precise means to measure the importance of these terms or phrases in a set of documents is the term frequency-inverse document frequency (TF-IDF) weight. The TF-IDF measures the importance of a word within a document, while weighing terms by their inverse frequency, thus decreasing the weight of terms that occur frequently:

\[
TF-IDF = \left( \frac{N_{ij}}{N_j} \right)^a \log(D/D_i)
\]

where \(N_{ij}\) is the number of times \(i\) in document \(j\), \(N_j\) as the number of words in document \(j\), \(D\) as the number of documents, and \(D_i\) the number of documents where the term \(i\) appears.

The formula thus assigns to the term \(i\) a weight in document \(j\) that increases when \(i\) occurs frequently in only a few documents, lower when \(i\) occurs fewer times in a document or occurs in many documents, and lowest when it occurs in all documents. This in part controls for differences in the window for coverage (for example, Kim Jong Un was not mentioned at all in the first half of 2010) as well as common repetitions that may not intend to convey new information (for example, the ubiquity of references to Kim Jong Il). While most of the terms see little difference between the two years, the TDF-IDF score drops considerably for socialism and economics as well as for Kim Jong Il, suggesting that while all three appear frequently, the terms fail to have greater salience. While one would expect a higher coefficient for Kim Jong Un in 2011, the tenfold increase in the TF-IDF measure suggests an importance that contrasts with its frequency.

I next calculated the Jaccard coefficient (a measurement of the proximity of terms within text) at three levels: the newsday, the paragraph, and the sentence.\(^8\) This statistic measures the frequency in which two terms coexist versus used separately and essentially can be viewed as a percentage of similarity. While over fifty terms and phrases were originally coded, I focused on

---

\(^8\) The Jaccard coefficient \((J)\) is a measure of the intersection of two terms divided by the union of these two terms or \(J = C/(A + B - C)\) where \(A\) is the frequency of term \(A\), \(B\) the frequency of term \(B\) and \(C\) the occurrence of both \(A\) and \(B\). Because of the length of the daily news, coefficients at the sentence and paragraph are expected to be small. Nevertheless, the Jaccard coefficient provides a means for comparison in the co-occurrence of words.
terms that, based on previous research, would be expected to occur frequently and at least indirectly to coverage of Kim Jong Un. Terms in which failed to co-occur with Kim Jong Un at least ten times at the newsday level were omitted. Furthermore, I generated separate statistics for 2010 and 2011 due to the general increased coverage of Kim Jong Un in the latter.

Several clear patterns emerge. The highest Jaccard coefficients consistently were terms related to Kim Jong Un’s official titles: the Central Military Commission (CMC), the National Defense Commission (NDC), the Korean People’s Army (KPA) and the Korean Worker’s Party (KWP). This is not particularly surprising considering Kim’s appointment to these bodies, especially if the intent of the reports is to not draw a more explicit connection between the younger Kim and his pedigree. In addition, references to Kim’s uncle Jang Sung Taek score high (see Table 2), but the strength of this relationship weakens in 2011 whereas the others all strengthen if not stay relatively flat. Meanwhile references to Kim Il Sung and Kim Jong Il do not occur at similar rates. In 2010, Jaccard coefficients for both were 0.075 and 0.74 respectively at the newsday level and barely breaking zero at lower levels. These rates increased considerably in 2011 (0.26 and 0.261 at the newsday level), but remain far below other references at each level. Admittedly this is largely due to the limited number of references to the younger Kim compared to the nearly daily references to the elder Kims.

Based on these preliminary findings, I next employ a negative binomial regression model with daily references to Kim Jong Un as the dependent variable. Regression analysis should allow for a more structured analysis, allowing us to control for the individual impact of terms and potentially uncover the importance of individual factors. Frequencies for each term in Tables 1 and 2 are included as independent variables. In addition, a dummy variable for 2010 (as no reference to the youngest Kim is made before September 27th) and for the period after Kim Jong Il’s death as well as the total word count for the daily news are included as controls.

Table 3 displays the results. In Model 1, references to Kim Il Sung and Kim Jong Il have coefficients with opposite signs, yet neither reaches statistical significance. This initial findings are consistent with claims that rhetoric around the youngest Kim is not a boilerplate propaganda that relies on a continuity through relying on the pedigree of the successor. Meanwhile references to Jang Sung Taek correlate with an increase in references to Kim Jong Un, statistically significant at the 0.001 level. References to the Korean Workers Party, the Korean People’s Army, and to a lesser extent the National Defense Commission all positive correlate with references to Kim Jong Un at the 0.05 level or better. Perhaps surprisingly, references to reunification negatively correlate with Kim Jong Un, but only significant at the 0.10 level. Nuclear references also negative correlate but fail to reach significance. Furthermore, dummy variables for the year 2010 (where Kim Jong Un was not mentioned in the first half) and the period after Kim Jong Il’s death both reach significance at the 0.001 level in opposite directions. Nearly identical results are found after included a lagged dependent variable, with the assumption that news in time period \((t)\) is partially conditioned by yesterday’s news \((t – 1)\). Additional specifications with added variables produced largely consistent results as well (omitted for brevity).

---

**Table 1**

Term frequency of selected topics in the KCNA news.

<table>
<thead>
<tr>
<th>Term</th>
<th>2010 Frequency</th>
<th>TF*IDF</th>
<th>2011 Frequency</th>
<th>TF*IDF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kim Il Sung</td>
<td>2685</td>
<td>42.4</td>
<td>4050</td>
<td>34.3</td>
</tr>
<tr>
<td>Kim Jong Il</td>
<td>4859</td>
<td>23.3</td>
<td>8229</td>
<td>9.9</td>
</tr>
<tr>
<td>Military</td>
<td>4342</td>
<td>26.1</td>
<td>6257</td>
<td>37.8</td>
</tr>
<tr>
<td>Reunification</td>
<td>3772</td>
<td>54.9</td>
<td>4077</td>
<td>64.8</td>
</tr>
<tr>
<td>Juche</td>
<td>2591</td>
<td>53.8</td>
<td>3204</td>
<td>35</td>
</tr>
<tr>
<td>KWP</td>
<td>2412</td>
<td>332.5</td>
<td>2040</td>
<td>320.9</td>
</tr>
<tr>
<td>Songun Politics</td>
<td>2092</td>
<td>86.3</td>
<td>2111</td>
<td>65.6</td>
</tr>
<tr>
<td>Anti-Imperialism</td>
<td>2032</td>
<td>91.9</td>
<td>1798</td>
<td>79.4</td>
</tr>
<tr>
<td>Socialism</td>
<td>1940</td>
<td>62.7</td>
<td>2909</td>
<td>35.4</td>
</tr>
<tr>
<td>Nuclear</td>
<td>1569</td>
<td>267</td>
<td>1778</td>
<td>198.3</td>
</tr>
<tr>
<td>Economics</td>
<td>1124</td>
<td>94.4</td>
<td>2334</td>
<td>42.9</td>
</tr>
<tr>
<td>Kim Jong Un</td>
<td>37</td>
<td>41.8</td>
<td>786</td>
<td>460.3</td>
</tr>
</tbody>
</table>

Source: Korean Central News Agency of the DPRK

---

**Table 2**

Jaccard coefficients regarding Kim Jong Un and selected terms.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>CMC</td>
<td>0.426</td>
<td>0.276</td>
<td>0.292</td>
<td>0.585</td>
<td>0.303</td>
<td>0.327</td>
</tr>
<tr>
<td>NDC</td>
<td>0.144</td>
<td>0.049</td>
<td>0.049</td>
<td>0.049</td>
<td>0.055</td>
<td>0.062</td>
</tr>
<tr>
<td>KPA</td>
<td>0.131</td>
<td>0.034</td>
<td>0.034</td>
<td>0.034</td>
<td>0.031</td>
<td>0.034</td>
</tr>
<tr>
<td>KWP</td>
<td>0.104</td>
<td>0.02</td>
<td>0.021</td>
<td>0.036</td>
<td>0.061</td>
<td>0.062</td>
</tr>
<tr>
<td>Jang Sung Taek</td>
<td>0.223</td>
<td>0.185</td>
<td>0.179</td>
<td>0.565</td>
<td>0.101</td>
<td>0.063</td>
</tr>
</tbody>
</table>

Source: Korean Central News Agency of the DPRK
To test the robustness of these findings, the original basic models were retested with a binary dependent variable, simply whether Kim Jong Un was mentioned that day or not (Table 4). Probit regression analysis finds roughly similar results. Here references to Kim Il Sung and Kim Jong Il both have positive coefficients, but neither reaches statistical significance, while references to Jang Sung Taek remains significant at the 0.001 level. References to the KPA and KWP largely match that of the previous models, while nuclear references and general references to the military reach statistical significance only in the probit model. Adding a lagged dependent variable produces largely consistent results. Post-estimation diagnostics find the models correctly classify the dependent variable in 90.91 and 91.32 percent of cases respectively.

Table 3
Negative binomial regressions on daily references to Kim Jong Un.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coeff</th>
<th>SE</th>
<th>Coeff</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kim II Sung</td>
<td>0.0027</td>
<td>0.0130</td>
<td>0.0027</td>
<td>0.0130</td>
</tr>
<tr>
<td>Kim Jong II</td>
<td>-0.0067</td>
<td>0.0059</td>
<td>-0.0065</td>
<td>0.0067</td>
</tr>
<tr>
<td>Jang Sung Taek</td>
<td>0.7048***</td>
<td>0.1198</td>
<td>0.7042***</td>
<td>0.1207</td>
</tr>
<tr>
<td>CMC</td>
<td>0.0325</td>
<td>0.0234</td>
<td>0.0329</td>
<td>0.0249</td>
</tr>
<tr>
<td>NDC</td>
<td>0.0733*</td>
<td>0.0336</td>
<td>0.0731*</td>
<td>0.0339</td>
</tr>
<tr>
<td>KPA</td>
<td>0.0544**</td>
<td>0.0173</td>
<td>0.0545**</td>
<td>0.0174</td>
</tr>
<tr>
<td>KWP</td>
<td>0.0229***</td>
<td>0.0057</td>
<td>0.0229***</td>
<td>0.0057</td>
</tr>
<tr>
<td>Anti-Imperialism</td>
<td>-0.0212</td>
<td>0.0244</td>
<td>-0.0211</td>
<td>0.0245</td>
</tr>
<tr>
<td>Economics</td>
<td>0.0030</td>
<td>0.0160</td>
<td>0.0031</td>
<td>0.0164</td>
</tr>
<tr>
<td>Juche</td>
<td>-0.0083</td>
<td>0.0120</td>
<td>-0.0083</td>
<td>0.0120</td>
</tr>
<tr>
<td>Military</td>
<td>0.0058</td>
<td>0.0098</td>
<td>0.0058</td>
<td>0.0098</td>
</tr>
<tr>
<td>Nuclear</td>
<td>-0.0251</td>
<td>0.0157</td>
<td>-0.0251</td>
<td>0.0157</td>
</tr>
<tr>
<td>Reunification</td>
<td>-0.0215t</td>
<td>0.0113</td>
<td>-0.0216t</td>
<td>0.0114</td>
</tr>
<tr>
<td>Socialism</td>
<td>0.0114</td>
<td>0.0152</td>
<td>0.0112</td>
<td>0.0159</td>
</tr>
<tr>
<td>Songun Politics</td>
<td>-0.0172</td>
<td>0.0140</td>
<td>-0.0173</td>
<td>0.0144</td>
</tr>
<tr>
<td>Year 2010</td>
<td>-1.3547***</td>
<td>0.2534</td>
<td>-1.3543***</td>
<td>0.2537</td>
</tr>
<tr>
<td>KJI Death</td>
<td>4.5448***</td>
<td>0.5859</td>
<td>4.5525***</td>
<td>0.6129</td>
</tr>
<tr>
<td>Word Count</td>
<td>0.0000</td>
<td>0.0001</td>
<td>0.0000</td>
<td>0.0001</td>
</tr>
<tr>
<td>Lagged DV</td>
<td>-0.0007</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.9278***</td>
<td>0.2883</td>
<td>-1.9285***</td>
<td>0.2887</td>
</tr>
<tr>
<td>N</td>
<td>726</td>
<td>726</td>
<td>726</td>
<td>726</td>
</tr>
<tr>
<td>Pseudo $R^2$</td>
<td>0.4050</td>
<td></td>
<td>0.4054</td>
<td></td>
</tr>
</tbody>
</table>

***p < .001, **p < .01, *p < .05, t p < .10.
Source: Korean Central News Agency of the DPRK

To test the robustness of these findings, the original basic models were retested with a binary dependent variable, simply whether Kim Jong Un was mentioned that day or not (Table 4). Probit regression analysis finds roughly similar results. Here references to Kim Il Sung and Kim Jong Il both have positive coefficients, but neither reaches statistical significance, while references to Jang Sung Taek remains significant at the 0.001 level. References to the KPA and KWP largely match that of the previous models, while nuclear references and general references to the military reach statistical significance only in the probit model. Adding a lagged dependent variable produces largely consistent results. Post-estimation diagnostics find the models correctly classify the dependent variable in 90.91 and 91.32 percent of cases respectively.

Table 4
Probit regressions on daily references to Kim Jong Un.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coeff</th>
<th>SE</th>
<th>Coeff</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kim II Sung</td>
<td>0.0083</td>
<td>0.0140</td>
<td>0.0087</td>
<td>0.0142</td>
</tr>
<tr>
<td>Kim Jong II</td>
<td>0.0071</td>
<td>0.0095</td>
<td>0.0072</td>
<td>0.0096</td>
</tr>
<tr>
<td>Jang Sung Taek</td>
<td>0.8727***</td>
<td>0.1325</td>
<td>0.8954***</td>
<td>0.1343</td>
</tr>
<tr>
<td>CMC</td>
<td>0.5292***</td>
<td>0.0753</td>
<td>0.5215***</td>
<td>0.0759</td>
</tr>
<tr>
<td>NDC</td>
<td>-0.0414</td>
<td>0.0490</td>
<td>-0.0444</td>
<td>0.0492</td>
</tr>
<tr>
<td>KPA</td>
<td>0.0577***</td>
<td>0.0221</td>
<td>0.0594***</td>
<td>0.0224</td>
</tr>
<tr>
<td>KWP</td>
<td>0.0283***</td>
<td>0.0075</td>
<td>0.0283***</td>
<td>0.0074</td>
</tr>
<tr>
<td>Anti-Imperialism</td>
<td>-0.0438</td>
<td>0.0251</td>
<td>-0.0402</td>
<td>0.0253</td>
</tr>
<tr>
<td>Economics</td>
<td>-0.0039</td>
<td>0.0180</td>
<td>-0.0040</td>
<td>0.0183</td>
</tr>
<tr>
<td>Juche</td>
<td>-0.0026</td>
<td>0.0186</td>
<td>-0.0062</td>
<td>0.0189</td>
</tr>
<tr>
<td>Military</td>
<td>0.0214*</td>
<td>0.0103</td>
<td>0.0207</td>
<td>0.0103</td>
</tr>
<tr>
<td>Nuclear</td>
<td>-0.0448**</td>
<td>0.0153</td>
<td>-0.0426**</td>
<td>0.0155</td>
</tr>
<tr>
<td>Reunification</td>
<td>-0.0074</td>
<td>0.0109</td>
<td>-0.0056</td>
<td>0.0110</td>
</tr>
<tr>
<td>Socialism</td>
<td>0.0060</td>
<td>0.0190</td>
<td>0.0078</td>
<td>0.0188</td>
</tr>
<tr>
<td>Songun Politics</td>
<td>-0.0168</td>
<td>0.0183</td>
<td>-0.0146</td>
<td>0.0186</td>
</tr>
<tr>
<td>Year 2010</td>
<td>-1.3113***</td>
<td>0.2362</td>
<td>-1.2815***</td>
<td>0.2409</td>
</tr>
<tr>
<td>KJI Death</td>
<td>1.4298</td>
<td>1.1692</td>
<td>1.3327</td>
<td>1.2068</td>
</tr>
<tr>
<td>Word Count</td>
<td>-0.0001</td>
<td>0.0001</td>
<td>-0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>Lagged DV</td>
<td>0.4300**</td>
<td>0.1911</td>
<td>0.4300**</td>
<td>0.1911</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.2304***</td>
<td>0.2797</td>
<td>-1.3329***</td>
<td>0.2870</td>
</tr>
<tr>
<td>N</td>
<td>726</td>
<td>726</td>
<td>726</td>
<td>726</td>
</tr>
<tr>
<td>Pseudo $R^2$</td>
<td>0.5413</td>
<td></td>
<td>0.5489</td>
<td></td>
</tr>
</tbody>
</table>

Correctly Classified 90.91% 91.32%

***p < .001, **p < .01, *p < .05, t p < .10.
Source: Korean Central News Agency of the DPRK
As a whole, regression analysis confirms rhetorical patterns within the text and suggests subtle distinctions in how Kim Jong Un is presented to a Western audience. While some of the findings affirm conventional wisdom, others bring up additional questions. Empirical analysis finds the references to Kim Jong Un are more closely tied to his office position and his uncle than his father or grandfather. That references to the Korean Workers Party positively correlated to Kim Jong Un requires additional unpacking. Considering that the KCNA reports are the official mouthpiece of the KWP, one would expect frequent references, perhaps to the point that such terms were simply boilerplate propaganda terms. Instead, the positive correlation may suggest an intention to re-emphasize the party over other sources of power in North Korea, especially since the party largely played a secondary role under Kim Jong Il. Without his own network to justify his rise to power, Kim Jong Un likely requires broad party support. Whether this bump is temporary unfortunately requires more longitudinal data. Lastly, of particular interest perhaps to Western analysts is that nuclear references negatively correlate with the Great Successor. While this of course could be unintentional or a statistical artifact, this preliminary evidence suggests an intention to downplay the issue until after support for Kim Jong Un as the heir apparent was solidified. It may also suggest an intention to move beyond nuclear rhetoric and the use of the nuclear card to entice concessions from regional powers or the return to the Six Party Talks framework with the US, China, Japan, Russia, and South Korea.

7. Concluding remarks

North Korea’s internal dynamics are not the black box that much of the Western media seems to suggest. Nevertheless, a full understanding of elite politics in the Hermit Kingdom is unlikely until the regime collapses. Automated content analysis allows scholars to fully exhaust what information is available, rather than simply reiterating claims of the impenetrability of the regime or dismissing the possibility of greater understanding of contemporary North Korean politics. Building upon this research, future analyses can compare the framing of Kim Jong Un before and after coming to power, identifying whether a consistent frame or shifts are evident over time. For example, an analysis of 2012 and 2013 KCNA news reports potentially could shed light on what led to the execution of Kim Jong Un’s uncle Jang Song Taek. If not predicting such shocking behavior, such an analysis may help outsiders identify how the regime fully intended to frame this action. Similarly, whether rhetorical shifts are seasonal, implicitly tied to recurrent domestic or international issues, requires more longitudinal data. While admittedly a preliminary analysis of news coverage around the Great Successor, this study highlights the possibility of extracting information from existing sources on North Korea and provides a means, however imperfect, of testing assumptions about North Korean intentions.

Acknowledgement

This research was funded by a Korean Studies Grant from the Academy of Korean Studies (d) (AKS-2011-R-11).

References

Herman, E.S., 1993. The media as a role in U.S. foreign policy. J. Int. Aff. 47, 23–43.
Kim, S., 2010. Leadership Change in North Korea
Klingner, B., 2010. Leadership Change in North Korea