Corpus Linguistics in Critical Discourse Analysis: A Case Study on News Reports of the 2011 Libyan Civil War[1]

Sibo Chen
School of Communication, Simon Fraser University

Abstract
This paper reports a comparative analysis of the news coverage of the 2011 Libyan civil war in two national media (*China Daily* and *The New York Times*). The 2011 Libyan civil war attracted wide attention and was extensively covered by various media around the world. However, news discourse regarding the war was constructed differently across various news agencies as a result of their clashing ideologies. Based on corpus linguistics methods, two small corpora with a total of 22,412 tokens were compiled and the comparative analyses of the two corpora revealed the following results. First, although the two corpora shared a lot of commonalities in word frequency, differences still exist in several high ranking lemmas. On the one hand, words such as "Qaddafi" and "war" ranked similarly in the two corpora's lexical frequency lists; on the other hand, the frequencies of the lemma "rebel/rebels" were much higher in *The New York Times* corpus than in the *China Daily* corpus, which indicated that the image of the rebel received more attention in the reports by *The New York Times* than in those by *China Daily*. Second, although the word "Qaddafi" achieved similar frequencies in the two corpora, a follow-up collocation analysis showed that the images of "Qaddafi" contrasted with each other in the two corpora. In *The New York Times* corpus, the words and phrases collocating with "Qaddafi" were mainly negative descriptions and highlighted the pressure on Qaddafi whereas many neutral and even positive descriptions of Qaddafi appeared in the *China Daily* corpus. Based on these findings, the paper further discusses how discursive devices are applied in news coverage of warfare, as well as some methodological implications of the case study.

1. Introduction

The year 2011 witnessed the rise of the "Arab Spring," the revolutionary wave of protests, riots, and even civil wars across the Mideast. Starting on December 18th, 2010, this democratic movement fundamentally has changed the political land-scape of many Arabic countries, and there is no doubt that the Arab Spring should be regarded as one of the most significant news events of 2011. Among all the protests and conflicts during the Arab Spring, the Libyan civil war was the most eye-catching news event as it occurred in a country well known for its authoritarian leader, Muammar Gaddafi, and ended with a dramatic civil war lasting for nearly seven months.

During the Libyan civil war, governments of different countries held differing, often opposed, political stances regarding foreign intervention in Libya's internal conflicts. For instance, Russia and China were strongly against the military intervention proposal by NATO at the initial stage of the Libyan crisis. Nevertheless, as the Libyan crisis eventually turned into a bloody civil war, both countries made concessions and agreed with NATO on military intervention. Nonetheless, it needs to be pointed out that both the Russian and Chinese governments held a neutral stance on the Libyan crisis and maintained that NATO's intervention was inappropriate until the end of this conflict. As indicated by previous studies (e.g. Bhatia, 2006; Hackett & Zhao, 1994; Salama, 2011; van Dijk, 1988, 1993), a government's political stands have significant influence on its national newspapers' reports of news events. For the same news events, there may be different or even

Corresponding Author: Sibo Chen (siboc@sfu.ca)

contradictory press reports by various news agencies owing to their host countries' clashing ideologies. Therefore, for the Libyan civil war, the same news events with different news narratives can be expected due to the conflicting diplomatic struggles behind the civil war, which provides an interesting research topic: what discursive strategies were employed by different news agencies to manipulate their readers' impression of the Libyan civil war?

This paper reports a comparative analysis of the news coverage of the 2011 Libyan civil war by two national media outlets: *China Daily* and *The New York Times*. These media were targeted as they represent two clashing ideologies, namely socialism and liberalism. Based on corpus linguistics methods, the current study investigated how conflicting diplomatic stands regarding the Libyan civil war were realized at the discursive level.

The paper is organized as follows: Section 2 reviews previous studies on war rhetoric and the theoretical debates on applying corpus linguistics methods in critical discourse analysis; Section 3 provides methodological information about the current study; and Section 4 presents the preliminary results and discusses their implications for future studies.

2. Literature Review

2.1 Critical discourse analysis and corpus linguistics

Critical Discourse Analysis (CDA) can be defined as "discourse analysis with critical stances," which concerns "real and often extended instances of social interaction that take a linguistic form or a partially linguistic form" (Fairclough & Wodak, 1997, p. 258). According to Mautner (2009), there are at least three essential principles in CDA: (1) discourse constitutes society and culture; (2) power relations are discursive; and (3) CDA has a commitment to study discourse surrounding urgent social issues. In short, CDA aims at providing interpretations and explanations of social phenomena from a discourse perspective.

One central debate in critical discourse research over the past decade relates to the application of corpus linguistics in discourse analysis. Following the interpretive tradition of language analysis, CDA tends to be a qualitative research discipline, which is not inherently connected with corpus linguistics. Influential CDA literature (e.g. Fairclough, 1995; Fairclough & Wodak, 1997; van Dijk, 1988) has repeatedly adopted self-solicited research data with an emphasis on the significance of data authenticity and contextual information about collected discourse material. As a result, the decontextualized language data in many early established corpora (e.g. the British National Corpus and American National Corpus) lacks contextual information needed for in-depth CDA. In addition, many discourse and genre scholars have argued that corpus linguistics methods are inappropriate for discourse studies due to the conflict between "bottom-up" and "top-down" processes of analysis. For instance, Swales (2002) points out that interpreting genre and discourse structures essentially requires a "top-down" process for interpreting macrostructures of texts, which makes many corpus linguistics methods (e.g. lexical/phrasal frequency and concordance) incompatible with genre and discourse studies due to their bottom-up processes of analysis and the lack of contextual considerations.

However, recent developments in corpus linguistics have offered a promising perspective for corpus-driven critical discourse studies. The innovation of corpus linguistics software (e.g. Word Smith Tools 6.0) allows for more accurate and quick elicitation of specific lexemes and phrases with their sentential contexts. The rapid expansion of the Internet has also changed the situation of online text availability. In particular, most newspapers today have been digitized, which facilitates data collection and corpus compilation processes. As discussed in Flowerdew (2005), with rigorous research designs, corpus linguistics methods can be an effective tool for discourse and genre research. Recent studies (e.g. Baker et al., 2008; Prentice & Hardie, 2009; Salama, 2011) have suggested several key advantages of applying corpus-driven methods in critical discourse studies,

including the reduction of research subjectivity and the triangulation of research data as well as the improvement of research validity through the incremental effect of discourse[2]. In short, current trends in corpus linguistics seem to suggest that corpus-aided critical discourse analysis will become more prevalent in the near future.

2.2 War rhetoric and media discourse

War rhetoric studies emerged as a subfield of rhetorical and discourse analysis from early studies of language use in Nazi propaganda and the rhetoric of the Cold War (van Leeuwen, 2006). In general, studies on war rhetoric can be divided into the European approach and the American approach, which differ from each other in their theoretical foundations. The European approach to war rhetoric has followed the legacy of critical linguistics since the mid-1970s, focusing not on words and phrases, but on the construction of particular ideas through discourse; by contrast, the American approach is based on traditional rhetorical analysis, with explicit reference to Aristotle and Kenneth Burke (van Leeuwen, 2006). However, recently the two approaches have grown closer together in terms of research methods.

Many recent studies of media discourse of warfare have focused on the justification of war through the construction of "enemy narratives." For instance, Graham, Keenan and Dowd (2004, cited in van Leeuwen, 2006) analysed George W. Bush's declaration of the "war on terror" and identified the following common strategies in call to war speeches: (1) referring to an ultimate moral force to legitimize the war; (2) providing historical and cultural discourse to rally the addressed audience together as "us"; and (3) constructing the enemy as an evil other. Many previous studies have confirmed that various textual features (e.g. narrative stance, metaphors, tone, and argument strategies) can contribute to the construction of different images regarding the same news events. For instance, Hackett and Zhao (1994) analysed newspapers published during the first two weeks of the Gulf War and found that newspapers with different political stances presented contrasting images of the anti-war protests by implementing various textual strategies (e.g. syntactic, lexical, and metaphoric choices). Similarly, Thetela (2001) studied the news reports on the 1998 Southern African Development Community's (SADC) military intervention in Lesotho and concluded that the application of discursive strategies by mass media played a significant role in constructing images of Lesotho's two rival social groups (i.e. the supportive group and the protest group opposing SADC's interventions). In a recent study, Salama (2011) combined corpus linguistics and CDA and explored the lexical collocations in Islamic media regarding complicated Wahhabi-Saudi Islam issues. The results of this study showed that statistically significant collocations can accurately reveal opposing political voices in Arabic media. In sum, rhetorical devices play a crucial role in war propaganda as they create effective storylines to legitimize the use of military force and construct evil enemies to fight.

Although previous studies have extensively addressed questions regarding media discourse on warfare, the majority of these studies used quantitative data analysis methods, with very few exceptions such as Salama (2011). As discussed above, there has been a strong trend towards combining corpus linguistics and CDA in discourse scholarship over the past decade. Maunter (2009) proposes that corpus linguistics methods can effectively support quantitative and qualitative research into discourse analysis and thus need to be further explored by future CDA studies. Inspired by Maunter's proposal, the current study implemented corpus linguistics methods to compare the news coverage of the 2011 Libyan civil war in two national media with clashing ideological backgrounds: *China Daily* and *The New York Times.*

3. Methodology

3.1 Research questions

This study focused on news reports on the 2011 Libyan civil war from two national media, *China Daily* and *The New York Times*. The choice of the two media was based on two considerations. First, the 2011 Libyan civil war represents how warfare is covered in a new media era, and the enormous attention paid to this conflict provides sufficient data across different mass media to conduct corpus-driven analyses. Second, English media published in China have rarely been studied. Thus, *China Daily*, as one of the leading official English media outlets operated by the Chinese government, presents a valuable case for investigation. It is interesting to explore whether opinions expressed by *China Daily* dramatically differ from those found in western media such as *The New York Times* since the newspaper's orientations are influenced by the Chinese government as well as its target reader group, westerners living in China.

Three research questions were posed in the current study:
1. Is there any significant difference in terms of lexical frequencies and distributions between *China Daily* and *The New York Times* in their news coverage of the Libyan civil war?
2. What are the collocation patterns of the lexeme "Qaddafi/Gadhafi" in *China Daily* and *The New York Times*? Is there any significant difference in the image of Qaddafi expressed by the two media?
3. If differences of image expressions exist, what factors caused these distinctions?

3.2 Data collection and analysis

The data consisted of 20 news reports on the Libyan civil war randomly selected from *China Daily* and *The New York Times* (10 from each). The selected articles were published between March 1, 2011, and October 23, 2011, covering all the major news events of the conflict. Two small corpora, the *China Daily* (CD) corpus and *The New York Times* (NYT) corpus, were subsequently built, including a total of 22,412 tokens. The articles in *China Daily* have an average length of 1,000 words whereas the average article length at *the New York Times* is 1,200 words.

The data analysis of the two corpora included two stages. First, with the application of WordSmith Tools 6.0, lexical frequency lists of the two corpora were generated and compared. Then, the lexeme "Gadhafi/Qaddafi"[3] was targeted for a collocation analysis as Muammar Qaddafi played a pivotal role during the Libyan civil war. Collocation can be defined as "the phenomenon that certain words often co-occur with each other" (Baker, 2006, p. 96). The collocations of a target word can provide valuable information about semantic preferences attached to it. Table 1 shows some sample collocation lines of "Qaddafi" generated by WordSmith Tools 6.0 from the collected data.

In the current study, collocations of "Gadhafi/Qaddafi" were further coded into four categories: (1) Negative Description, (2) Description of Pressure, (3) Neutral/Positive Description, and (4) Other. These categories are defined in Table 2. The collocations were first coded independently by two coders, and the coding results were then combined for the data analysis. The inter-coder reliability rate of the current analysis was 83.4%. After the completion of the coding process, the number of collocation lines in each category were calculated, and the results were compared between the CD corpus and the NYT corpus.

Table 1
Sample Collocation Lines of "Qaddafi" from the Collected Data

n Tripoli, Musa Ibrahim, a spokesman for the **Qaddafi** government, conducted a bizarre news conference
redulous audience of foreign journalists whom the **Qaddafi** government had invited to Tripoli, Mr. Ibrahim
and Libyans as an alternative to Col. Muammar el-**Qaddafi**'s four decades of freakish rule. But events thi
as it solicits help from abroad to topple Colonel **Qaddafi**. Rebels were dealt military setbacks in Zawiyah
gling local governments. "I am Libya," Colonel **Qaddafi** boasted after the uprising erupted. It was standa
ne rebels here — forging a substitute to Colonel **Qaddafi** in a state that he embodied. "We've found ours
n Libya, largely because few believe that Colonel **Qaddafi** can reconquer a region that long seethed under
hat unfolded beneath cartoons lampooning Colonel **Qaddafi**. Security has begun to deteriorate, with gunfire
bstantial hardware with which to take on the pro-**Qaddafi** stronghold of Tripoli," said a report on Tuesday

Table 2
Coding Scheme for Collocation Lines of "Gadhafi/Qaddafi"

Category	Definition	Example
Negative Descriptions	Derogatory lexemes collocating with "Qaddafi" or direct negative comments on Qaddafi or his military force's behaviour	"Qaddafi's <u>regime</u>" "Qaddafi <u>terrorized and intimidated</u> Libyans"
Description of Pressure	Descriptions of Libyan rebels' movements or NATO's political and military pressure on Qaddafi or his military forces	"NATO has been <u>bombing</u> Gadhafi's forces . . . "
Neutral/Positive Description	Neutral or positive comments on Qaddafi and his governance in Libya	"Under Gadhafi, Libya made <u>considerable economic and social progress</u>."
Other	Other phrases collocating with "Gadhafi/Qaddafi"	"Gadhafi's <u>hometown</u>"

4. Results and Discussion

4.1 Results of the lexical frequency analysis

Table 3 shows the top seven content lexemes of the CD corpus and the NYT corpus. It can be observed that the two compiled corpora share many commonalities in top-ranking content lexemes. For instance, in both corpora the lexeme "Gadhafi/Qaddafi" is the top content lexeme while the lexeme "government" ranks low (seventh). On the other hand, the results in Table 3 also indicate several differences between the CD corpus and the NYT corpus. In particular, the lexeme "rebel" and its plural form "rebels" have much higher frequencies in the NYT corpus than in the CD corpus. Although the NYT corpus has nearly 80% more tokens than the CD corpus (14,395 vs. 8,017 tokens), the NYT corpus has a total of 158 tokens of "rebel/rebels," which is more than two times that of the CD corpus result (40 tokens of "rebels" and 27 tokens of "rebel"; the frequency of "rebel" is not shown in Table 3). Another difference between the two corpora is the rankings of the lemma "Libya/Libyan." The lexeme "Libya" ranks second and the lexeme "Libyan" ranks fourth in the CD corpus. By contrast, the NYT corpus has Libya ranking only fifth in its list. It is surprising to find that the total token quantity of "Libya/Libyan" is higher in the CD corpus than in the NYT corpus despite the fact that the NYT corpus has many more tokens than the CD corpus.

Table 3
Top Seven Content Lexemes in the CD Corpus and the NYT Corpus

The CD corpus (8,017 Tokens)			The NYT corpus (14,395 Tokens)		
Rank	Quantity	Lexeme	Rank	Quantity	Lexeme
1	108	Gadhafi	1	170	Qaddafi
2	81	Libya	2	144	said
3	78	Said	3	98	colonel
4	41	Libyan	4	81	rebels
5	40	Rebels	5	78	Libya
6	33	Forces	6	77	rebel
7	33	Government	7	71	government

4.2 Results of the collocation analysis

Table 4 shows the collocation analysis results of the CD corpus and the NYT corpus. For the lexeme "Gadhafi/Qaddafi," the analysis of the CD corpus generated a total of 108 collocation lines while the result of the NYT corpus was 170. As shown in Table 4, there are many differences in the distribution of collocation lines between the two corpora. In particular, the NYT corpus has a much higher percentage of collocation lines in the "Negative Description" category than the CD corpus (31.31% vs. 13.84%). By comparison, the CD corpus includes a higher percentage of collocation lines in the "Neutral/Positive Description" category than the NYT corpus (39.23% vs. 31.31%).

Table 4
The Collocation Analysis Results of the Lexeme "Qaddafi/Gadhafi"

	CD	NYT
Negative Description	18 (13.84%)	57 (31.31%)
Description of Pressure	43 (33.07%)	51 (28.02%)
Neutral/Positive Description	51 (39.23%)	57 (31.31%)
Other	18 (13.84%)	17 (9.34%)
Total	130	182

Note. Some collocation lines in the CD and the NYT corpora were coded as both "Negative Description" and "Description of Pressure" (e.g. "the Gadhafi regime must be overthrown").

4.3. Discussion and Conclusion

The results in both Tables 3 and 4 provide evidence supporting the research hypothesis that the same news event (the 2011 Libyan civil war) was presented differently in *China Daily* and *The New York Times* via the implementation of various discursive strategies. The results of the CD corpus suggest a neutral political stand whereas results of the NYT corpus indicate clear support for Libyan rebels and for NATO's military actions against Muammar Qaddafi. To be specific, the high rankings of the lexeme "Qaddafi/Gadhafi" in Table 1 suggest Qaddafi was regarded as playing a central role of the Libyan conflict by both *China Daily* and *The New York Times*. Meanwhile, the low rankings of "government" in Table 3 imply that both media recognized the Libyan government's ineffective role in the direction of the war.

On the other hand, the ranking differences between the two corpora in the lemma "Libya/Libyan" and the lemma "rebel/rebels" indicate the political and narrative stances of *China Daily* and *The New York Times*. The results of Table 3 demonstrate that *The New York Times* paid much attention to individual rebel figures during the war and provided much exposure of the rebels' military and political actions whereas *China Daily* described the war from the Libyan government's perspective, and its major concern was the future geopolitics of the Middle East. In short, the rebel groups of the Libyan civil war were somewhat neglected by *China Daily*. The above finding is further supported by the results in Table 4, in which we can see that many more negative descriptions of Qaddafi can be found in the NYT corpus than in the CD corpus. The portrait of Qaddafi in *The New York Times* was of a ruthless dictator who conducted a brutal genocide against Libyan civilians. However, such highly negative adjective phrases cannot be found in the CD corpus, which described Qaddafi as a neutral or even somewhat positive political figure who, despite some minor political errors in his home country, still received significant support from the Libyan people and bravely fought against NATO's military intervention.

In sum, although the current study's data is limited in scope and thus prevents any strong conclusion, the data analysis procedure clearly shows the potential of implementing corpus linguistics methods in CDA to reveal how salient textual devices are adopted in media texts, representing different or even conflicting ideological stances. The clashing ideological backgrounds between *China Daily* and *The New York Times* may be the major cause for the differences observed in the data analysis, and the findings of the current study suggest methodological implications for future studies on war rhetoric that take a quantitative approach.

Notes

1. Reprinted by Permission of Canadian Association for the Studies of Discourse and Writing. Sibo Chen, *2013 CASDW conference proceedings*, 17-28. © 2013 Canadian Association for the Studies of Discourse and Writing.
2. A specific genre or particular type of discourse is a collection of texts sharing common lexico-grammatical and discoursal features. Corpus linguistics methods, with the application of corpus analysis programs, are particularly good at revealing language patterns by processing large text collections.
3. "Gadhafi" was used by *China Daily* whereas "Qaddafi" was used by *The New York Times.*

References

Baker, P. (2006). *Using corpora in discourse analysis*. New York, NY: Continuum.

Baker, P., Gabrielatos, C., Khosravinik, M., Krzyzanowski, M., McEnery, A. M. & Wodak, R. (2008). A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society, 19*(3), 273-306.

Bhatia, A. (2006). Critical discourse analysis of political press conferences. *Discourse & Society, 17*(2), 173-203.

Fairclough, N. (1995). *Critical discourse analysis: Papers in the critical study of language*. New York, NY: Longman.

Fairclough, N., & Wodak, R. (1997). Critical discourse analysis. In T. A. van Dijk (Ed.), *Discourse as social interaction* (pp. 258-284). London, England: Thousand Oaks.

Flowerdew, L. (2005). An integration of corpus-based and genre-based approaches to text analysis in EAP/ESP: Countering criticisms against corpus-based methodologies. *English for Specific Purposes, 24*(3), 321-332.

Graham, P., Keenan, T., & Dowd, A-M. (2004). A call to arms at the end of history: A discourse-historical analysis of George W. Bush's declaration of war on terror. *Discourse & Society, 15*(2/3), 199-221.

Hackett, R., & Zhao, Y. (1994). Challenging a master narrative: Peace protest and opinion/editorial discourse in the US press during the Gulf War. *Discourse & Society, 5*(4), 509-541.

Mautner, G. (2009). Corpora and critical discourse analysis. In P. Baker (Ed.), *Contemporary corpus linguistics* (pp. 32-46)*.* New York, NY: Continuum.

Prentice, S., & Hardie, A. (2009). Empowerment and disempowerment in the Glencairn Uprising: A corpus-based critical analysis of early modern English news discourse. *Journal of Historical Pragmatics, 10*(1), 23-23.

Salama, A. H. Y. (2011). Ideological collocation and the recontexualization of Wahhabi-Saudi Islam post-9/11: A synergy of corpus linguistics and critical discourse analysis. *Discourse & Society, 22*(3), 315-342.

Swales, J. M. (2002). Integrated and fragmented worlds: EAP materials and corpus linguistics. In J. Flowerdew (Ed.), *Academic discourse* (pp. 150–164). London, UK: Longman, Pearson Education.

Thetela, P. (2001). Critique discourses and ideology in newspaper reports: A discourse analysis of the South African press reports on the 1998 SADC's military intervention in Lesotho. *Discourse & Society, 12*(3), 347-370.

van Dijk, T. A. (1988). *News as discourse*. Hillsdale, NJ: Lawrence Erlbaum Associates.

van Dijk, T. A. (1993). Principles of critical discourse analysis. *Discourse & Society, 4*(2), 249-283.

van Dijk, T. A. (2005). War rhetoric of a little ally: Political implicatures and Aznar's legitimatization of the war in Iraq. *Journal of Language and Politics, 4*(1), 65-65.

van Leeuwen, T. (2006). War rhetoric. In K. Brown (Ed.) *Encyclopaedia of language and linguistics* (2nd ed.) (pp. 516-520). Oxford, UK: Elsevier.