

A Corpus-Based Critical Analysis of Lexical Choices in Sichuan Earthquake by Chinese and American News Media

WANG Li^{[a]*}; CHEN Shuzhen^[b]

^[a]Associate Professor, College of Foreign languages, Nantong University, Nantong, China.

^[b]Lecturer, Sanming No. 2 Senior High School, Sanming, China.

*Corresponding author.

Supported by the Educational Council of the Chinese Government in the 2013-2016 Academic Year “an investigation into the Acquisition of Chinese Discourse Markers by Learners From Different Language Backgrounds” (13YJC740096).

Received 25 March 2015; accepted 19 June 2015

Published online 26 July 2015

Abstract

By combining the Critical Discourse Analysis (CDA) approach with the corpus linguistic method, this study compared the lexical choices employed by two English newspapers, i.e. China Daily and the New York Times, in their coverage of Sichuan earthquake. This study aims to probe into the different political stances and ideology concealed under the news reports. The comparative analysis of the two corpora was conducted at three levels: word frequencies, part-of-speech (POS) frequencies and semantic category frequencies.

The results show that there are significant differences among the three levels in the lexical choices about Sichuan earthquake between the two media, which reflect different ideologies and political stances of China and America. By highlighting the endeavor of Chinese relief work, the CDC presents a responsible government and a united Chinese nation. On the contrary, the NYTC focuses more on emphasizing the dictatorship and incompetence of the Chinese government and the scientific and technological backwardness in China, hence creates a negative image of China. On the basis of the present findings, this study intends to be helpful to the teaching of English reading in China.

Key words: Sichuan earthquake; Critical discourse analysis; Corpus; Lexical choices; Contrastive study

Wang, L., & Chen, S. Z. (2015). A Corpus-Based Critical Analysis of Lexical Choices in Sichuan Earthquake by Chinese and American News Media. *Canadian Social Science*, 11(7), 111-116. Available from: <http://www.cscanada.net/index.php/css/article/view/7339>
DOI: <http://dx.doi.org/10.3968/7339>

INTRODUCTION

The 2008 Sichuan disastrous earthquake attracts much attention in both Chinese media and American media. Many researchers applied diverse theories and approaches to analyze those news coverage. So far, however, none of the researchers have combined the Critical Discourse Analysis (CDA) approach with the corpus linguistic method to compare the lexical choices employed by Chinese media and American media in the coverage of earthquake. The purpose of CDA is to discover the relationship between language, power and ideology. News discourse is one of the fields in which CDA is interested. The news discourse is composed of the lexical items. Therefore, the analysis of vocabulary can help readers comprehend the purpose of reporters. Although many researchers focused on the study of the news reports, there are few studies which have combined the theory of CDA and corpus linguistic research method to analyze the reports about Chinese events by media both at home and abroad.

Thus, this thesis attempts to find the differences in the lexical choices between Chinese media and American media and to probe into the underlying reasons for those differences. This study serves as an important supplement to existing CDA studies of media discourse. Specifically, this study intends to answer the following three questions.

- a) Are there any significant differences in the lexicon used by the Chinese newspapers and the American newspapers at the level of word frequencies?
- b) Are there any significant differences in the lexicon used by the Chinese newspapers and the

- American newspapers at the level of part-of-speech (POS) frequencies?
- c) Are there any significant differences in the lexicon used by the Chinese newspapers and the American newspapers at the level of semantic category frequencies?

1. LITERATURE REVIEW

Critical Discourse Analysis engages in the analysis of interrelationship between discourse and power by relating discourse to its historical and social environment, with the purpose of exposing the social ideology and the unequal power relations embedded in the discourse and investigating its potential influence in guiding readers or listeners' attitudes or controlling their thoughts (Fairclough, 1989). Its current focus on language and discourse was initiated with the critical linguistics that emerged (mostly in the UK and Australia) at the end of the 1970s (Van Dijk, 1998). CDA can be defined in part by its focus on social issues and social problems (e.g. racism, sexism, unclear disarmament), although these topics are also addressed by discourse analysts who work with other traditions (Wood & Kroger, 2008).

Van Dijk (1998) exploited this CDA theory to analyze the reproduction of ethnic prejudices through ideology in the press. Fairclough (1992) put a three-dimensional model as the theoretical framework for critical news analysis. He suggested a systematic study of news discourse. He analyzed a news report from the British newspaper *The Sun* which mixed its own populist, conversation-like style with the bureaucratic language of the document on which it was reported Brooks (1995) analyzed the headlines of news stories about Africa in the British newspapers to show how the news discourse in Africa rises to a racist representation of Africa.

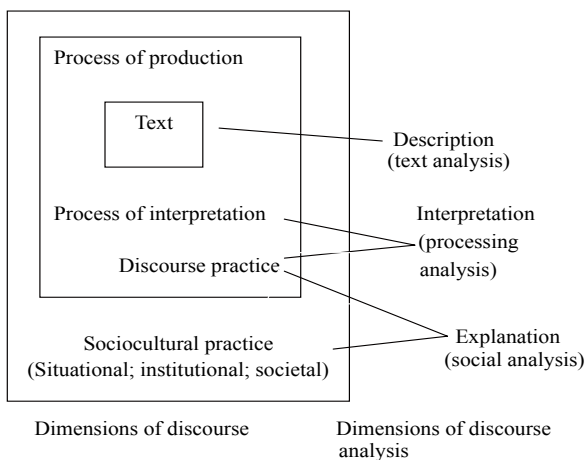


Figure 1
Discourse and Discourse Analysis (Fairclough, 1995, p.98)

As a widely used approach to the analysis of news discourse, Fairclough's three-dimensional model is

adopted in this study. The method of Fairclough's framework includes three aspects, namely, "linguistic description of the language text, interpretation of the relationship between the (productive and interpretative) discursive processes and the text, and explanation of the relationship between the discursive processes and the social processes" (Fairclough, 1995, p.97).

2. RESEARCH METHODOLOGY

2.1 Corpora Used in the Study

In order to make a comparative lexical analysis of the English news reports on Sichuan earthquake between Chinese and American media, two corpora, hereafter briefly called the NYTC and the CDC respectively, were built. The NYTC is composed of 41,064 words on Sichuan earthquake reported by the New York Times. The CDC consists of 41,067 words, which are produced by China Daily. The texts in the two corpora were downloaded from <http://www.nytimes.com> and <http://www.chinadaily.com.cn>, respectively.

2.2 Softwares Used in the Analysis

In order to retrieve the data for analysis, three tools, i.e. AntConc, Log-likelihood calculator¹ and Wmatrix were employed in this study. The most obvious advantage of the AntConc is its ability to compare the frequencies of the same word between the two corpora.

The log-likelihood value (LL value) is calculated for measuring the extent of difference in the relative frequency of a particular item between the two corpora. The larger the absolute value of the LL parameter is, the more different the two corpora are in terms of the relative frequency of the target item. A LL value of 3.8 or higher is significant at the level of $p < .05$ and a LL value of 6.6 or higher is significant at $p < .01$.

The beauty of the Wmatrix is that it can get each word in the corpus parsed and tagged with the appropriate part-of-speech (POS) and semantic category, so that the frequencies of those POSs and semantic categories between the two corpora can be compared.

2.3 Research Procedures

Firstly, two corpora, both of which comprise more than 40,000 words, were established in order to conduct a contrastive study of the lexical choices of the reports on Sichuan earthquake between American media and Chinese media.

Secondly, by running the keyword function of AntConc, a list of keywords which is based on a comparison of the frequencies of all the words in the two corpora was obtained respectively. Then the following analysis was based on the keywords.

¹ Retrieved from <http://ucrel.lancs.ac.uk/llwizard.html>

Thirdly, with the help of Wmatrix, each word in the two corpora was tagged with the appropriate part-of-speech (POS). Then a list of comparing the frequencies of those POSs between the two corpora was retrieved. The next step was data analysis according to the UCREL CLAWS7 Tagset system².

Fourthly, the Wmatrix was used in this step as well. All the tokens in the two corpora were tagged with the semantic category, and then a list was obtained, which compared the frequencies of the semantic categories employed in the NYTC and the CDC respectively. The last step was analyzing the related data on the basis of Semantic Tagset system³.

3. RESULTS AND DISCUSSION

In this section, the comparison of lexical choice between the NYTC and CDC are carried out in terms of word frequencies, part-of-speech (POS) frequencies and semantic category frequencies.

3.1 Comparison at the Level of Word Frequency

In this part, AntConc was employed to produce two lists of key words which were based on a comparison of the observed frequencies of all the tokens in the two corpora. The first list consists of those words which occur significantly more frequently in the NYTC than in the CDC, as shown in Table 1, and the second list is composed of those words which were used significantly more frequently in the CDC than in the NYTC, as indicated in Table 2. The following two tables list only the top ten words with the highest LL values in each list since these words are most likely to exhibit the differences between the two corpora.

As shown in Table 1, the word *state*, which has a marked contrast in the frequencies in the NYTC and the CDC in that it occurs as many as 51 times in the NYTC but only 5 in the CDC. The different use of this word in the two corpora may reflect the opposite attitudes towards the Chinese government by the two newspapers.

Table 1
The Overused Words in the NYTC

Item	O1	1%	O2	2%	LL
That	481	1.17	190	0.46	+130.51*
A	957	2.33	645	1.57	+61.68*
Officials	88	0.21	19	0.05	+48.25*
State	51	0.12	5	0.01	+43.94*
Government	162	0.39	68	0.17	+39.57*
Media	51	0.12	9	0.02	+32.46*

To be continued

² Retrieved from <http://juiland.comp.lancs.ac.uk/wmatrix2/claws7tags.html>

³ Retrieved from <http://juiland.comp.lancs.ac.uk/wmatrix2/semtags.html>

Continued

Item	O1	1%	O2	2%	LL
Thousands	65	0.16	16	0.04	+31.97*
Dam	33	0.08	4	0.01	+25.95*
Mountains	39	0.09	7	0.02	+24.54*
Miles	29	0.07	3	0.01	+24.45*

Note. O1 is the observed frequency in the NYTC, O2 is the observed frequency in the CDC, 1% and 2% values show relative frequencies in the texts, + indicates overuse in O1 relative to O2, $p^* < 0.0001$; critical value = 15.13

Through a closer observation of the concordances of all the tokens of *state* in the NYTC, it can be found that an overwhelming majority of tokens are used to modify *media*. It is worthwhile to note that the *media* also has a large LL value. *State* also often co-occurs with *run* or *controlled*. Most of these concordances, as illustrated in the examples below, are employed to describe the status of Chinese media.

Table 2
The Overused Words in the CDC

Item	O1	1%	O2	2%	LL
Quake	149	0.36	399	0.97	-118.36*
Sichuan	56	0.14	185	0.45	-72.79*
Country	18	0.04	103	0.25	-65.96*
Yuan	4	0.01	63	0.15	-62.57*
Hit	43	0.10	145	0.35	-58.43*
Monday	48	0.12	145	0.35	-51.04*
Sichuan	149	0.36	286	0.70	-43.88*
Relief	45	0.11	130	0.32	-43.08*
Premier	1	0.00	35	0.09	-40.77*
Chengdu	44	0.11	119	0.29	-35.84*

Note. O1 is the observed frequency in the New York Times

Some of the concordances of *state* in the NYTC are shown below.

1) ...By 9 p.m. local time, the **state** news agency Xinhua quoted provincial disaster relief officials as saying that 3,000 to 5,000....

2) ...were damaged. Four pupils died and more than 100 others were injured, **state** media reported. Another person was...

The concordances show that the media mostly refers to Xinhua News Agency in NYTC. Besides, the word *propaganda* is often used to modify Xinhua. It is best exemplified by the following concordance:

3)...the Xinhua News Agency, the government's leading **propaganda** organ, has offered...

According to the Oxford Advanced Learner's English-Chinese Dictionary (6th ed., 2004), the connotation of *propaganda* is "(usually disapproving) ideas or statements that may be false or exaggerated and that are used in order to gain support for a political leader, party, etc.." Meanwhile, the American media

employs *propaganda* to qualify Chinese officials and department.

The concordances of *propaganda* in the NYTC are shown below.

4) ...fact, several Chinese journalists confirmed that **propaganda** officials had issued an order prohibiting the.

5) ...effectively ignored orders from the Central **Propaganda** Department not to travel to the earthquake zone a...

On the contrary, Xinhua in the CDC only a simple term to refer to Chinese media, and does not co-occur with *state*, *state-run*, *state-controlled* or *propaganda*.

Therefore, compared with the neutral description of the *media* in the CDC, NYTC focused more on the controlled status of the Chinese media by the Chinese government. The frequent co-occurrence of *state* and *propaganda* demonstrate that the American media believe that Chinese media are controlled and influenced by the government. Thus, Chinese media lack the freedom of speech. There are two reasons that may explain the negative attitude of American media towards the Chinese government. To begin with, social ideological systems of United States are totally different from that of China. For most American people, “Communism” and “Communist Part” are reactionary. They cannot comprehend some politics and regulations set down by the Chinese government. It is well known that most American news agencies are private-owned. In their opinion, “state-owned” means dictatorship. Thus, the Chinese government is a tyrannical government.

The Chinese media accentuate the effort of Chinese rescue, particularly the swift reaction of the Chinese government. As shown in Table 2, the word *relief* has a marked contrast in the CDC and the NYTC since it occurs as many as 130 times in the CDC but only 45 in the NYTC, as are shown below.

Similar to the use of the word *relief*, the word *Premier* often occurred in the CDC to underline the government’s prompt reaction as well. Chinese Premier Wen Jiabao set out his rescue journey from Beijing to Sichuan at 16:40 on May 12, only two hours after the earthquake. As soon as he arrived there, the Premier threw himself into the rescue work. Through Wen Jiabao’s action in the quake-ravaged regions, the CDC presents a positive and responsible Chinese government image.

3.2 Comparison at the Level of Part of Speech (POS)

In this section, the comparison of the CDC and the NYTC is conducted at the level of part-of-speech with the help of Wmatrix. After all the tokens in the two corpora were tagged with the appropriate part-of-speech (POS) according to the UCREL CLAWS7 Tagset system, a list was obtained, which compares the part-of-speech in terms of the extent of difference in their frequencies

between the two corpora. This list shows that among the top forty parts-of-speech with the highest LL values, and most of them are functional words, denoting no semantic meaning at all. The reason why the thirty fifth one JJ, which refers to general adjectives in the UCREL CLAWS7 Tagset system, is chosen to be analyzed is that it not only belongs to the content word, but also occupies the second proportion in both the NYTC and the CDC.

Table 3
The Most Frequently Used General Adjectives in the CDC

Item	POS	Raw frequency	Relative frequency
Chinese	JJ	88	0.23
Local	JJ	69	0.18
Quake-Hit	JJ	57	0.15
Military	JJ	55	0.14
Dead	JJ	50	0.13
Medical	JJ	49	0.13
Provincial	JJ	48	0.12
Other	JJ	37	0.09
Major	JJ	35	0.09
Strong	JJ	27	0.07

Table 4
The Most Frequently Used General Adjectives in the NYTC

Item	POS	Raw frequency	Relative frequency
Chinese	JJ	140	0.36
Dead	JJ	43	0.11
Other	JJ	41	0.10
Local	JJ	40	0.10
Natural	JJ	39	0.10
Official	JJ	37	0.09
Major	JJ	36	0.09
Military	JJ	36	0.08
International	JJ	27	0.07
New	JJ	25	0.06

The result produced by Wmatrix shows that there are 3,133 and 3,039 general adjectives in the CDC and the NYTC respectively. Tables 3 and 4 display the general adjectives with more than 27 occurrences in the CDC and the NYTC.

As indicated in Tables 3 and 4, those words that highlight the aspects of relief work occupy a larger proportion in the CDC than in the NYTC. These words include: *Chinese*, *military*, *medical* and *strong*, with some of the concordances of *military* in the CDC, as is exemplified below.

6) ...from nearby mountains. The **military** continued the disaster-relief work shortly after the recent tremor, and people...

7) ...in order to get urgently needed to help into the area. He had urged **military** personnel to do their best to open the roads. Road...

The word *military* is one of the most frequently used words to present the Chinese relief work. Two minutes after the earthquake, the troop which consists of the local garrison and armed police got to rescue people. It breaks the record of Chinese military's action of fighting the natural disaster. No wonder that some media refer to this action as "the Long March of rescue". The military plays an indispensable role, which has highly been praised by the Chinese media. After the disaster happens, the most important thing is to save people's lives. The military is needed, so does the medical aid.

The second finding of the present research is that the same words used by the CDC and the NYTC serve different purposes. The word *Chinese* is a case in point. This word has two functions in the CDC. On the one hand, the CDC attempts to put much emphasis on the great relief work done by the government and the army. On the other hand, the CDC employs this word to express that this earthquake is the whole nation's disaster, and the Chinese people are united together to help and support their compatriots in Sichuan. The following concordances can demonstrate these points.

The concordances of *Chinese* in the CDC are shown below.

8) ...in the disaster areas, under the strong leadership of **Chinese** government, can overcome losses caused by the disaster, help...

9) ...Present Hu Jintao called for **Chinese** army and armed police stationed near the quake-hit areas to make effort to assist...

Instead of pinpointing the Chinese relief work, the NYTC makes mention of *Chinese* merely because those events happen in China and it wants to help its American readers to understand the location and the people involved in the news.

The concordances of *Chinese* in the NYTC are shown below.

10) ...spreaders. The United States agreed to provide **Chinese** authorities with satellite images of the earthquake zone and two...

11) ...destroying 80 percent of structures in some of the towns and small cities near its epicenter, **Chinese** officials said. Its...

3.3 Comparison at the Level of Semantic Category

Table 5 shows ten semantic categories with the greatest significant differences between the two corpora, and which are ranked in terms of the absolute value of the LL parameter. As is evident from Table 5, the Y1 category (the LL value between the two corpora being 82.36) and the Q4 category (the LL value between the two corpora being 51.93) occur much more frequently in the NYTC than in the CDC. And thus these two categories are chosen for further discussion.

Table 5
Comparison of the Frequencies of Semantic Categories Between the NYTC and the CDC

Item	O1	1%	O2	2%	LL	Category
Z99	905	2.30	1886	4.83	-361.15*	Unmatched
Y1 114	0.29	16	0.04	+82.36*	Science and technology in general	
E5-	185	0.47	361	0.93	-59.35*	Fear/shock
I1	50	0.13	154	0.39	-56.53*	Money generally
Q4	88	0.22	17	0.04	+51.93*	The Media
B3	82	0.21	195	0.50	-48.49*	Medicines and medical treatment
W3	656	1.67	426	1.09	+47.22*	Geographical terms
T1.2	30	0.08	102	0.26	-42.15*	Time: Momentary
S8+	276	0.70	441	1.13	-39.81*	Helping
M7	757	1.92	1011	2.56	-38.93*	Places

Notes. O1 is the observed frequency in the New York Times

The Y1 category denotes science and technology in general, as is exemplified in Table 6.

According to Table 6, the word *scientists* occupy the highest proportion in Y1 category. The concordances of *scientists* reflect the backwardness of Chinese science and technology in NYTC. The NYTC believes that Chinese scientists are incapable of predicting the earthquake and they are mainly responsible for this catastrophe. From their point of view, this disaster could have been avoided if Chinese science and technology had been advanced enough. The American media tries to convey such kind of message in two ways. The first way is by describing the common people's comment about Chinese scientists. For example, "... another 4,500 are missing. 'Those seismological **scientists** are wasting taxpayer money,' said ...". The other way is by quoting the Chinese scientists' words. For instance, "...a group of environmentalists and scholars said the fact that government **scientists** had underestimated the risk of the May earthquake raised questions about...". Unless inconsistent with their personal beliefs and experiences, recipients tend to accept beliefs (knowledge and opinions) through discourse from what they see as authoritative, trustworthy or credible sources, such as scholars, experts, professionals or reliable media (Van Dijk, 1998). Thus, quoting scientists' words would definitely help its readers accept the opinion of NYTC. The reason why the American media put much emphasis on the backwardness of the Chinese science and technology is that the American media may be influenced by Orientalism. Orientalism is primarily a term used for the imitation or depiction of aspects of Eastern cultures in the West by writers, designers and artists. American scholar Edward Said (1978) published his influential

and controversial book, *Orientalism*; he used the term to describe a pervasive Western tradition, both academic and artistic, of prejudiced outsider interpretations of the East, shaped by the attitudes of European imperialism in the 18th and 19th centuries. A central idea of Said is that Western knowledge about the East is not generated from facts or reality, but from preconceived archetypes that envision all “Eastern” societies as fundamentally similar to one another, and fundamentally dissimilar to “Western” societies. Therefore many Americans believe that China, a country of the Orient and the Third World, is backward.

Table 6
The Selected Words of Y1 in the NYTC

Item	Category	Raw frequency	Relative frequency
Scientists	Y1	34	0.09
Technology	Y1	10	0.03
Geology	Y1	9	0.02
Science	Y1	8	0.02
Engineers	Y1	8	0.02

Table 7
The Selected Words of Q4 in NYTC

Item	Category	Raw frequency	Relative frequency
Media	Q4	48	0.12
Published	Q4	13	0.03
Coverage	Q4	9	0.02
Publish	Q4	5	0.01
Editors	Q4	3	0.01

As shown in Table 7, the Q4 category stands for the media-related words. Table 8 shows the most frequently used five words in the I1 category, which represents rescue work generally. As is obvious in Table 8 that yuan occupies the highest proportion in I1 category. After the earthquake, the Chinese people try their best to give their hands so as to help the victims. These concordances present a united and strong image of Chinese nation.

Table 8
The Selected Words of I1 in the CDC

Item	Category	Raw frequency	Relative frequency
Yuan	I1	65	0.17
Insurance	I1	18	0.05
Cash	I1	15	0.04
Money	I1	8	0.02
Financial	I1	7	0.02

Table 9 lists the most frequently used five words in the B3 category, which represents those words whose

meanings are related to medicines and medical treatment, as exemplified below.

Table 9
The Selected Words of B3 in the CDC

Item	Category	Raw frequency	Relative frequency
Medical	B3	49	0.13
Hospital	B3	38	0.1
Doctors	B3	15	0.04
Medicine	B3	9	0.02
Treatment	B3	8	0.02

The concordances of *medical* in the CDC are shown below.

12) ...for disaster relief comprising eight departments responsible for helping with rescue work , **medical** care, security, information...

13) ...normal. The Sichuan Provincial health department has sent 24 **medical** teams to Sichuan, Mianzhu and Shifang...

The earthquake results in so great casualties that the medical treatment is in great need. Just as one of the disaster relief officials said, “Saving lives in the aftermath of the Sichuan earthquake is the top priority and rescue operations will continue as long as hope exists that people can be saved.”

CONCLUSION

To sum up, American media and Chinese media approached the Sichuan earthquake from different perspectives. American media pays attention to the negative aspects, such as the lack of freedom of Chinese media, the backwardness of Chinese military and technology. On the contrary, Chinese media focuses on the positive aspects. The present findings have some implications for the English teaching in China. Newspaper reading is an essential part in EFL learning and media discourse is often biased. Therefore, teachers should foster their students’ critical awareness in newspaper reading.

REFERENCES

- Fairclough, N. (1989). *Language and power*. New York: Longman Inc.
- Fairclough, N. (1992). *Discourse and social change*. Cambridge: Polity Press.
- Fairclough, N. (1995). *Critical discourse analysis*. Cambridge: Polity Press.
- Fowler, R. (1991). *Language in the News: Discourse and ideology in the press*. London: Routledge.
- Hornby, A. S., et al. (2009). *Oxford Advanced learner’s English-Chinese dictionary* (6th ed.). Beijing: The Commercial Press.
- Van Dijk, T. A. (1998). *The handbook of discourse analysis*. Blackwell Publishers.
- Said, E. (1978). *Orientalism*. New York: Pantheon Books