CSCanada
Studies in Literature and Language
Vol. 6, No. 2, 2013, pp. 35-39
DOI:10.3968/j.sll.1923156320130602.4002

ISSN 1923-1555[Print]
ISSN 1923-1563[Online]
www.cscanada.net
www.cscanada.org

Applying Corpus Linguistics in Discourse Analysis

WANG Dongmei[a]*

[a] Lecturer, Department of Foreign Languages, Dalian University of Technology, China.
* Corresponding author.

Supported by the Fundamental Research Funds for the Central Universities (DUT13RW427).

Received 12 February 2013; accepted 2 April 2013

Abstract
The present study employed corpus analytical tools to study and compare BP Company's CSR reports before and after the 2010 oil spill in the Gulf of Mexico. Corpus linguistics served in the research as a methodological tool since it ensured a more qualitative research methodology. And Fairclough's three-dimensional conception was employed to explore how an organization uses justificatory discursive strategies to legitimize its operations after perceived disasters.

Key words: Corpus linguistics; Discourse analysis; CSR reports

WANG Dongmei (2013). Applying Corpus Linguistics in Discourse Analysis. *Studies in Literature and Language, 6*(2), 35-39. Available from:http://www.cscanada.net/index.php/sll/article/view/j.sll.1923156320130602.4002
DOI: http//dx.doi.org/10.3698/j.sll.1923156320130602.4002

INTRODUCTION
Corpus linguistics (CL) focuses on "the study of language based on examples of real life language use" (McEnery & Wilson, 1996). There are two kinds of CL researches. The first kind is called descriptive corpus research, where the researchers work hard to find the linguistic patterns of a language. In the second kind of research, the researchers still work hard to find the linguistic patterns of a language, but in order to explain how the language people use in certain interactional context help constructing the reality they are in. This paper belongs to the second kind of research, applying CL as a methodological tool to understand the construction of discourses. Here we call CL a methodological tool because CL ensures a more qualitative research methodology with the help of large quantity of naturally occurring language data and various corpus software and statistics analytical tools. However, just as Biber (1998) suggests, corpus-based researches (applied corpus researches) have to depend on both the quantitative techniques and qualitative interpretation frameworks. The following are the examples of this kind of research. O'Halloran (2010) analyzed a corpus of British newspaper articles about immigrants, and the researcher employed the theory of Critical Discourse Analysis (CDA) as his interpretative framework. The other studies which are corpus-based and take CDA as the interpretative include Fairclough's (1995) study of media discourse, Hajer's (1997) study of environmental discourse, Chen & Lam's (2012) study on Western perceptions on Hong Kong a decade after the reversion of the sovereignty from Britain to China in 1997, Bhatia's (1997) study on public discourse in Hong Kong, Fang's (2001) study on Chinese print news media discourse, Flowerdew's (2004) study on globalization discourse, etc. The above-mentioned researches conducted quantitative analysis with the help of large general corpora and corpus processes (keyword, frequency, dispersion, concordance). At the same time, the analytical framework of critical discourse analysis was employed to interpret how language, whose linguistic patterns have been found in CL analysis, served as a form of social practice. In the next part, I will take my research on CSR reports as an example to show how to conduct a corpus-based critical discourse analysis.

1. ANALYZING CSR REPORTS
Corporate Social Responsibility Report is a genre that we take to analyze as a discourse because these reports reveal

how an organization or institution discursively construes its public identity and negotiates its relationship with their stakeholders. We chose two CSR reports as samples. The present study chose two CSR texts, one is British BP Company's *Sustainability Review 2009,* and the other is British BP Company's *Sustainability Review 2011*. The reason why these two years' reviews are chosen is that in 2010, BP Company's Deepwater Horizon oil spill caused a massive environmental disaster in the Gulf of Mexico. So we intend to compare the two years' CSR reports before and after the disaster to explore how an organization use justificatory discursive strategies to legitimize its operations after perceived episodes of wrongdoings. The size of the 2009 CSR report corpus is 17200, and the size of the 2011 CSR report corpus is 22620. And for the purpose of analyzing the two corpora, I used WordSmith concordance software.

1.1 Keyword Analysis

There is a function in Wordsmith concordance software called "keyword". By creating keyword list, I obtained all of the words that occur statistically more often in the 2011 CSR reports than in the 2009 CSR reports. Just as Baker points out, "a keyword list gives a measure of saliency rather than frequency" (Baker, 2006). With a default of $p<0.000001$ set up in WordSmith, the two CSR reports data generate a total of 13 keywords.

Table 1
Keywords of Two CSR Texts

N	Key word	Freq.	%	RC Freq.	%	Keyness
1	SPILL	46	0.20	0		51.74
2	GULF	62	0.27	6	0.03	39.28
3	HORIZON	27	0.12	0		30.36
4	DEEPWATER	55	0.24	8	0.05	27.40
5	STRONG	18	0.08	0		20.24
6	HIGH-CONSEQUENCE	14	0.06	0		15.74
7	CHALLENGE	13	0.06	0		14.61
8	COMPLEX	18	0.08	1		14.09
9	ENERGY	11	0.05	27	0.16	-12.23
10	FRONTIER	16	0.07	34	0.20	-12.72
11	ADVANCED	11	0.05	28	0.16	-13.25
12	EFFICIENT	34	0.15	56	0.32	-13.46
13	BEST	3	0.01	17	0.10	-15.17

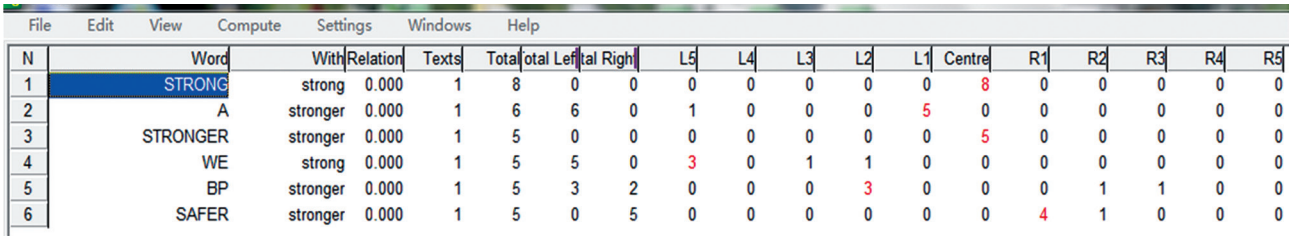
The 13 keywords are presented in order of their keyness. The keyness score starts high at 51.74 for the word SPILL, and gradually decreases to 14.09. After the eighth word COMPLEX, the keyness scores begin to become higher again. With the last word BEST; its keyness has risen to 15.17. This is because half of the table (from number 1 to number 8) is showing the words that occur more frequently in the 2011 CSR report, while the second half of the table (from number 9 to number 13) is showing the words that occur more frequently in the 2009 CSR report. The third column presents the frequencies of each keyword which occurred in the 2011 CSR report, and the fourth column shows the percentages of the keywords in the 2011 corpus. The fifth and sixth column show the frequencies of each keyword which occurred in the 2009 CSR report, and the percentage of each keyword in the 2009 corpus.

1.2 Sorting Concordance Lines of Keywords

A concordance is a list of all the occurrences of a particular search term in a corpus, presented in the context in which it occurs. A concordance is also sometimes called key word in context or KWIC. The 'key word' here refers to the word that we search and examine in a corpus and is different from 'keyword' we discussed in 2.1. With the help of WordSmith Concord tool I carried out a research on some of the keywords in the two CSR reports. In table 1, the first four words whose keyness scores very high in the 2011 CSR text are all about the Deepwater Horizon oil spill accident in 2010. As we all know, in 2010 a BP oil well broke and sent thousands of gallons of crude oil into the Gulf of Mexico, causing one of the most serious environmental catastrophes in the history. No wonder that these four words stay high in the keywod list. Taking a close look at the concordance lines of DEEPWATER HORIZONG, it's not difficult to find out the oil spill in the Gulf of Mexico was labeled by the BP Company as an accident or incident. In other words, BP was the victim who should win the public's sympathy rather than criticism.

Table 2
Concordance Lines of DEEPWATER HORIZON

N	Concordance		
1	in BP's investigation after the	Deepwater Horizon	accident. The gulf
2	monitors BP's response to the	Deepwater Horizon	accident through
3	management review Following the	Deepwater Horizon	incident. We began
4	the National Commission on the BP	Deepwater Horizon	Oil Spill and
5	to share the lessons learnt from the	Deepwater Horizon	accident and
6	experience. For example, following the	Deepwater Horizon	incident, we have
7	from BP's internal investigation into the	Deepwater Horizon	incident, contractor
8	In the immediate aftermath of the	Deepwater Horizon	explosion, BP
9	report (the Bly Report) into the	Deepwater Horizon	accident, BP
10	on GHG emissions associated with the	Deepwater Horizon	incident or response
11	conclusions into the investigation of the	Deepwater Horizon	oil spill in 2010
12	and incident investigations, including	Deepwater Horizon	The
13	to share the lessons learnt from the	Deepwater Horizon	accident and
14	legal proceedings. Investigations into the	Deepwater Horizon	accident will plan an
15	to natural resources resulting from the	Deepwater Horizon	accident, and from
16	that they would have been in if the	Deepwater Horizon	incident had not
17	events C in February 2010, before the	Deepwater Horizon	oil spill started
18	have been raised even higher since the	Deepwater Horizon	accident and oil spill
19	the Church investors Group since the	Deepwater Horizon	oil spill and
20	from the internal investigation into the	Deepwater Horizon	accident for our



N	Word	With	Relation	Texts	Total	Total Left	Total Right	L5	L4	L3	L2	L1	Centre	R1	R2	R3	R4	R5
1	STRONG	strong	0.000	1	8	0	0	0	0	0	0	0	8	0	0	0	0	0
2	A	stronger	0.000	1	6	6	0	1	0	0	0	5	0	0	0	0	0	0
3	STRONGER	stronger	0.000	1	5	0	0	0	0	0	0	0	5	0	0	0	0	0
4	WE	strong	0.000	1	5	5	0	3	0	1	1	0	0	0	0	0	0	0
5	BP	stronger	0.000	1	5	3	2	0	0	0	3	0	0	0	1	1	0	0
6	SAFER	stronger	0.000	1	5	0	5	0	0	0	0	0	0	4	1	0	0	0

Figure 1
Collocation of Word STRONG

From figure 1, we could see that the word STRONG collocates with the word *we* and its comparative form STRONGER collocate with *a, safer* and *BP*. The main collocates include, *make BP a stronger, safer company*, *we deliver safe and strong operations, we place strong emphasis, we put in place strong foundations, we feel a strong responsibility,* and *we expect strong growth.* By taking a closer look at the concordance line of the words STRONG and STRONGER, we can interpret the message these concordance lines convey in this way: BP (we) has learnt a lot from the disaster, so we will be a stronger one in the future. In this sense, the two words, by serving both as a positive appreciation resource and as a judgment resource, not only help the organization defend itself in face of crisis, but also fulfill a promotional function with self-praising.

Table 3
Concordance Lines of STRONG and STRONGER

N	Concordance		
1	in a responsible and sustainable way, a	stronger	and safer BP
2	Sustainability Review. 2011 Building	stronger	safer company. 2011 was a year
3	with great determination to make BP a	stronger	safer company. We have set three
4	Our employees worked hard to make BP a	stronger	safer company. We recognize
5	the actions being taken to make BP a	strong	safer company. People have high
6	But the overall trend is likely to be one of	strong	growth in energy demand
7	Managed so that we can deliver safe and	strong	Operations. Managing risk from
8	checks and balances. We place	strong	Emphasis on checks and balances
9	practical plans to reduce risk and deliver	strong	sustainable performance
10	Underpinned by technology and relations	Strong	Financial performance is vital
11	We are BP. We are putting in place	strong	Foundations to make BP a safer
12	Growth in developing economies. We feel a	strong	responsibility to help meet
13	We expect the continuation of	strong	growth in the world's most

In appraisal theory, evaluative expressions can be classified into judgment, appreciation, and affect. In 2011 CSR report, we found other three words belong to the category of "appreciation", and all these three words occurred significantly more frequently here than in the 2009 CSR text. The three words are HIGH-CONSEQUENCE, CHALLENGE AND CONPLEX. By taking a closer look at the three words with the help of the tool in WordSmith called KWIC searches, we found that all the three words are used to evaluate the problems the whole oil industry has to face to: oil industry is a dangerous business, so it is challenging for us to provide energy to meet the ever increasing need in a secure and efficient way.

Table 4
Concordance Lines of CHALLENGE

N	Concordance		
1	The scale of the	challenges	is such that it can only be
2	to choose wisely. Global economic	challenges	have reduced the focus of
3	logistical, infrastructure and cost	challenges	Concerns about nuclear
4	Create many engineering and technical	challenges	The oil and gas reservoir
5	Climate change. Addressing the global	challenges	of climate change will require
6	environments create some unique	challenges	for oil and gas operations
7	Helped us to recognize and address	challenges	and long-term
8	personnel can assist	challenges	and escalate or intervene as
9	many other companies signed	challenges	sponsored by
10	them. Following a procedural	challenges	However, the matter was
11	to keep rising, the global energy	challenges	is becoming
12	Energy security represents a	challenge	In its own right. More than
13	environmental issues. The energy	challenges	with energy demand
14	technical, environmental and social	challenges	ahead in every area
15	The energy future Today's	challenges	is meeting the growing
16	Meeting the global energy	challenges	requires a diverse mix of fuels
17	and the South China Sea	challenges	of deepwater exploration
18	will be essential in addressing the	challenges	of energy security and
19	We believe the global energy	challenges	can only be met through a
20	Logistical, infrastructure and cost	challenges	presented by the 450 ppm

For the word COMPLEX, only the sentence in line 5 is talking about the oil spill accident. Putting the sentence in a bigger context, we found that it is about the investigation result from some external investigations: This accident is complicated because it had multiple causes and many parties got involved. So the seemingly negative appreciation "complex" can be interpreted as a justificatory wording to the interest of BP company.

Table 5
Concordance Lines of COMPLEX

N	Concordance		
1	The situation was	complex	As BP's personnel were
2	for the accident, rather than a	complex	inter-linked series of
3	and faces diverse and sometimes	complex	environmental regulations at
4	These proceedings include	complex	civil litigation which has largely
5	Namely that this was a	complex	accident that involved multiple
6	Legal processes around the incident are	complex	The exact shape form and
7	We understand that operating in	politically-complex	regions and

The other word that attracts our attention is "HIGH-CONSEQUENCE". Just like the word COMPLEX, HIGH-CONSEQUENCE was not used as a negative appreciation to depict any consequences brought about by the accident. It collocates with the word "potential", indicating that oil industry is a hazardous industry, mistakes cannot be avoided. So, in other words, through the Deepwater Horizon accident, we have learned a lot about how to respond and manage the dangers in the future. But sometimes dangers are inevitable.

Table 6
Concordance Lines of HIGH-CONSEQUENCE

N	Concordance		
1	major accident risk and other potentially	high-consequence	risks.
2	environmental audits for certain potentially	high-consequence	activity. Within our
3	supply chain that involve potentially	high-consequence	activities. Our
4	that use contractors in potentially	high-consequence	activities. The
5	supply chain that involve potentially	high-consequence	activities. Individual
6	Including for low-probability, potentially	high-consequence	scenarios. Our

The other five keywords in the second half of the keyword list are more frequently occurred in the 2009 CSR report. They are ENERGY, FRONTIER, ADVANCED, EFFICIENT, and BEST. By using the automatic semantic annotation system to tag the 2009 CSR report, the four words among the five keywords, FRONTIER, ADVANCED, EFFICIENT, and BEST are labeled as PJ, which means the four words are semantically positive self-judgment evaluations. Looking at the concordance lines of the words that are labeled as positive self-judgments, we can see that the BP company tries to depict itself as a moral, capable, and responsible pioneer company in the oil industry. The following table shows the concordance lines of the positive self-judgment evaluations in the 2009 CSR report, with some of the words that are not keywords but belong to the same semantic category. Baker (2006) believes that finding out the key categories of a corpus can help the researcher to discover the unique style of a text and its semantic prosodies.

Table 6
Concordance Lines of the Positive Self-Judgments (samples)

NO.			
1	in producing their resources. Our	frontier	skills are being applied in new
2	We have three of the world's most	advanced	CCS projects in our portfolio
3	a revitalized BP is driving innovative,	efficient	and responsible operations
4	focus on four key areas represents the	best	way forward in terms of
5	and international levels. We are making	effective	use of resources and networks
6	taking practical steps to ensure BP is a	diverse	and inclusive company where
7	values BP is progressive, responsible,	innovative	and performance driven

CONCLUSION

The present study is about how to conduct a discourse analysis by using corpus analytical techniques, among which, two techniques, keyword lists and concordance lines are employed. Keyword list derived from the comparison of the two little texts help us generalize significant lexical differences before and after the environmental disaster. By sorting the concordance lines of those keywords, we are exposed to the contexts where the keywords occur. But the patterns we found from the contexts are still subject to the researchers' interpretation. The quantitative results derived from the corpus analytical software help counter the bias of researchers' subjective interpretation. All the keywords that occur more frequently in the 2011 CSR report are employed to help the BP Company to legitimate its managerial and operational behaviors. Adopting Fairclough's (1992) three-dimensional conception, discourse is a text, a discursive practice and a social practice. In this sense, a CSR report is itself a text that can be analyzed in a linguistic method. Researcher can talk about its linguistic patterns, semantic prosodies, and rhetorical styles. A CSR report is also a discursive practice whereby a company discloses the social, environmental, legal impacts of their operational practices. At the same time, a CSR, especially when released after an environmental accident, is a social practice whereby a company addresses the concerns of potential readers and defuses the criticisms. In face of a major environmental disaster, BP deploys interpersonal resources to portray itself as an indispensable provider of energy for the whole world, a survivor from the disaster, and a student who is good at learning from a lesson.

REFERENCES

Baker, B. (2006). *Using corpora in discourse analysis*. Continuum: London.

Bhatia, V. K. (1997). Democratizing government decision-making: A study of public discourse in Hong Kong. *Journal of Pragmatics, 28*, 515–532.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: investigating language structure and use*. Cambridge: Cambridge University Press.

Chen, W., & Lam, P. (2012). Western perceptions of Hong Kong Ten years on: a corpus-driven critical discourse study. *Applied Linguistics, 33*(5), 1-19.

Fang, Y. J. (2001). Reporting the same events? A critical analysis of Chinese print news media texts. *Discourse and Society, 12*(5), 585–613.

Fairclough, N. (1992). *Discourse and social change*. Cambridge: Polity Press.

Fairclough, N. (1995). *Media discourse*. London: Hodder Arnold.

Flowerdew, J. (2004). Identity politics and Hong Kong's return to Chinese sovereignty: Analysing the discourse of Hong Kong's first chief executive. *Journal of Pragmatics, 36*(9), 1551–1578.

Hajer, M. (1997). *The politics of environmental discourse: ecological modernization and the policy process*. Oxford: Oxford University Press.

McEnery, A., & Wilson, A. (1996). *Corpus linguistics*. Edinburgh: Edinburgh University Press.

O'Halloran, K. (2008). Fleeing, sneaking, flooding: a corpus analysis of discursive constructions of refugees and asylum seekers in the uk press. *Journal of English Linguistics, 36*(1), 5-38.

Scott, M. (2008). *Oxford wordsmith tools 5.0 Manual*. Oxford: Oxford University Press.