Analysing Private Hospital Websites from a Critical Perspective: Potential Issues of Methodology, Analysis and Interpretation of Findings
Jonathan Rante Carreon
Richard Watson Todd


Abstract
Private hospitals are potentially torn between two conflicting purposes of providing medical care and profit making (e.g. Relman, 2008; Thompson, 1993), and this has implications for the way they communicate their goals. Due to social constraints, they are unlikely to openly express that their main goal is profit making. Conversely, pressure from shareholders may force them to openly and heavily advertise services they offer to attract potential patients. These competing agendas suggest the use of Critical Discourse Analysis (CDA) to uncover opaque and transparent uses of language (Fairclough, 1985; Van Dijk, 2003; Wodak, 2001). However, traditional CDA has come under criticism for allowing analysts to find whatever they want to find (e.g. Widdowson, 1995). Moreover, there is also a tendency to favour the use of certain methods (e.g. metaphor analysis) that conveniently allow the analysts to assign specific critical interpretations within a given text leading to the potential for over-interpretation (O’Halloran, 2007). These criticisms highlight an aspect of CDA that deserves investigation. Thus, this paper presents an approach to conducting a CDA of private hospital websites which addresses these criticisms by reducing the impact of preconceived beliefs on interpretations.

Introduction
Private hospitals are potentially torn between two conflicting purposes of providing medical care and profit making (e.g. Relman, 2008; Thompson, 1993). However, due to social and ethical constraints, these hospitals are unlikely to openly state that their main goal is profit making. Thus, any investigation of hospital priorities need to address the potential opaqueness of relevant information involved. A popular approach that uncovers opaque language uses is Critical Discourse Analysis (CDA) (See for instance the works of Fairclough, 1985, 1988, 1992, 1995, 2003; Wodak, 2004, 2007; Wodak & Meyer, 2001; Van Dijk, 1997, 2001 [1993], 2003; Caldas-Courthard & Coulthard, 1996; Chilton, 1993, 1998, 2002). There are several advantages of CDA. First, as an approach, it is a flexible methodological framework as it allows the analyst to examine from a critical perspective issues and concerns from different fields of study (Chouliaraki & Fairclough, 2010). As a theoretical framework, it permits the analyst to employ whatever methods or observe any linguistic and semiotic phenomena relevant to the purposes of the research study. Second, it is useful in examining how events, practices and texts arise and have been ideologically shaped by power relations and power struggles (Wodak, 2001). Third, it is indispensable in exploring opaque language uses that are assumed by writers as part of the background knowledge (Fairclough, 1985) and how the opacity of relations between discourse and society is itself a factor in securing power and hegemony (Fairclough, 1995; Van Dijk, 2009). Fourth, it is also useful in examining how social practices are bound up with causes and effects that may be not at all clear to people involved (Fairclough, 1995; Van Dijk, 2009). Finally, it may contribute to social emancipation (Fairclough, 2003).

Against these advantages are several occasional (e.g. Hammersley, 1997; Stubbs, 1997) and persistent (Collins & Jones, 2006; Jones, 2004; Jones, 2007; Jones & Collins, 2006; Widdowson, 1995; Widdowson, 1996; Widdowson, 1998) critiques of CDA strongly arguing that while it contributes in the critical analysis of both spoken and written texts and in exposing opaque practices, it is filled with the researcher’s political biases. Out of these criticisms three methodological issues become apparent. First, data selection is criticised for being ideologic. Second, methods in the analysis are those favoured by the analysts and not pilot-tested for their productivity against a sample of the data. Third, analysts are in danger of finding whatever they
want to find in the interpretation of findings, which usually are due to the first two issues. These issues are discussed in detail in the following sections.

CDA is particularly prone to these kinds of criticisms due to the nature of the enterprise. In brief, the CDA poses threats to reliability due to subjective choice of data, methods for the analyses and interpretation of findings. Thus, in this paper we present an approach to conducting CDA that recognizes and addresses the above criticisms with the main goal being to contribute to the methodological development of CDA.

**Finding whatever one wants to find**

To illustrate this paper’s methodological contribution to CDA, we would like to explore first what makes CDA problematic. This enables us to pinpoint potential causes of problems, so that preventive measures can be introduced to lessen their impact of the interpretations of findings. For instance, in critically investigating the conflicting purposes of providing medical care and profit making among private hospitals by examining their websites, data sets (e.g. Investor Relations webpage) that convey the purpose of profit making can be ideologically chosen to support our claims. Thus, ideologically selecting the Investor Relations webpage of a private hospital which contains mostly financial information of the hospital (e.g. annual profit), we can identify either the multimodal or linguistic information on this page as our “evidence” to support our claim that it prioritises profit making over medical care provision such as that shown in Figure 1 below.

However, this approach to data reflects a heavily criticised way of conducting CDA since hospital priorities were examined using ideologically chosen data that produces pre-determined results and therefore were expected to produce the results the analyst would like to find. This has also implications on how the critical analysis is viewed. Instead of viewing it as a neutral process of uncovering opaque practices and beliefs of institutions and people (where the positive and negative features are weighed against the evidences gathered), it is viewed as a judgmental tool, where concerned institutions and individuals are deemed “evildoers” even before the evidence is presented. Hence to address issues in data selection, in the analyses that follow, we would like to start by explaining how representative data is selected.

**Data and data selection**

As mentioned above, ideologically chosen data potentially introduces bias to the data analyses and interpretations by skewing the findings to the purposes of the research study. To prevent this from happening, data for the analysis may be selected using three criteria: representativeness of the selections, relative importance of the selections and purposes of the study. For instance in the main research of one of the authors looking at diachronic data from a private hospital website, representativeness means that sample data is taken from all the versions of the hospital website.

In this paper, we illustrate the methodological issues raised by critiques of CDA (e.g. Widdowson, 1995) by looking at the priorities of private hospitals through their websites. While examining all pages of the hospital website presents a more vivid picture of the priorities of the hospital, it is impractical to do so because of two reasons. First, the main concern of this paper is to present methodological issues, problematise them and illustrate a potential way to solve or at least mitigate these issues. Second, due to space constraints, the results and findings of linguistically analysing all the webpages are impossible to contain in space allowed for a short paper.

As for data used in illustrating methodological issues, we have selected Bumrungrad International Hospital (BIH) website available from December 1998 to present. BIH is the largest and best-known private hospital in Thailand and the first hospital in Asia to receive the prestigious JCI (Joint Commission International Accreditation Standards for Hospitals).
accreditation. It is also the winner of the 2008 'Best Website for International Travel' award at the Consumer Health World Awards. As we have posited earlier, private hospitals like BIH are potentially torn between two conflicting purposes of providing medical care and profit making. Given the leading role of BIH in Thai private and international healthcare, it would therefore appear to be a Thai private hospital where the conflicting purposes of medical care and profit making are likely to be important and thus useful as source of data for examining methodological issues in CDA. 

When representative data is identified, the second criterion, relative importance of the selections, should also be met. Any page (from the website) can represent the entire website and any page from each version can represent this version. However, of all these pages, which have higher relative importance?

Homepages have higher relative importance than the other pages which are positioned one or more levels lower than the homepages. Generally, the homepage is the first and most commonly visited page (Nielsen, 2000) turning it into a reception area and the most common portal of entry to a website (Knox 2007, Nielsen, 2000). As a portal of entry, it provides an
overview of the website’s contents and where highlights of the entire website are brought to the
front (Knox, 2007). In short, the five homepages were selected as representative samples of the
five website versions. One of the versions (version three) was excluded since only the Splash
page, which “was eye candy for users and served no navigational function” (Cooke, 2003: 171)
is available and the entire version is inaccessible since all links in this version were deadlinks.

Other (webpages) were selected at random. Randomly selected data provides equal
chances to all available data to be chosen and included in the analysis. There are, however,
minor concerns to consider when randomly selecting data such as in case where data is taken
from different versions of a hospital website. Different versions suggest that the contents in
each version vary, as do the number of webpages contained in each of these versions. Doing a
random sampling may result in having data represented only by the version with the highest
number of webpages. Thus, to ensure that all versions are represented, random sampling can be
modified to introduce an internal benchmark against which other portions of the data are
compared.

The function of an internal benchmark may be assigned to version one of the hospital
website. It is assumed that the earliest website version contains the initial and original objectives
of the hospital when it was initially launched. After comparing the contents of the internal
benchmark to other versions of the hospital website, webpages that were similar to the
benchmark were separated from those that are different. Similar webpages were used for the
main analysis. Webpages unique in each version were also examined. The former type of data
may shed light on the priorities of the hospital that were introduced in version one and carried
until the present time. The latter may indicate priorities that are unique in each version. This
kind of data selection allows better representation of all the versions and reflects the
longitudinal component of the research data being studied.

Finally, data can also be chosen based on the purposes of the study. Most if not all
research studies select data that suits their objectives. Thus, when the goal is to investigate
linguistic features of a text, only the linguistic portion is taken. On the other hand, if the
purposes of the research study require only the analysis of the multimodal features, then only
the semiotic components of the text are taken for examination. The same data selection process
was also followed in this paper, so that the analysis and interpretation of findings are done
transparently despite the influence of the purposes of the research in data selection.

Data analysis

There are several potential problems that may affect data analysis in CDA. One is
having predetermined results where instances from the text examined are taken to support any
claim (Widdowson, 1995). Another is in the choice of methods used in examining data where
most methods used are not pilot-tested. Rather, they are chosen based on their popularity (e.g.
metaphor analysis) and based on what the analyst favours.

In order to prevent these potential sources of bias in affecting the data analysis, various
methods commonly used and linguistic phenomena commonly observed in CDA were pilot-
tested against a sample of the data in question. This is to identify productive methods for
investigation and productive linguistic phenomena observable in the data. A combination of 35
different methods and linguistic phenomena commonly used in CDA were pilot-tested against
representative samples taken from the BIH website. However, for purposes of illustration and
due space constraints we are presenting only two of the methods (metaphor analysis and
keyword analysis) considered.

First is metaphor analysis, which a qualitative method heavily exploited in CDA. As a
result of its heavy use in CDA, there is a danger that commonly used metaphors maybe assigned
specific critical interpretation within a given although the choice of metaphor by the writer
simply reflects general language use (O’Halloran, 2007). Hence, metaphor analysis appears to be
a useful example in illustrating how a commonly used method in CDA is potentially
problematic. The second is keyword analysis, which is a quantitative approach to examining data. Despite being a quantitative approach and the most productive among the pilot-tested methods, there are still concerns that needed immediate attention and thorough decision-making. In brief, these two methods are useful in illustrating methodological issues in CDA from both qualitative and quantitative perspectives.

What follows next are sample analyses using two different methods to present a transparent approach to CDA. Each of these methods is discussed briefly; then potential sources of bias while using these methods are illustrated followed by presenting measures that may prevent or at least lessen the occurrence of these biases. It is assumed that attempting to control for these potential sources of biases reduces the impact of preconceived beliefs on interpretations and therefore contributes to CDA methodology.

**Metaphor analysis**

The choices made among different ways of expressing the same concept can provide insights into the conceptual systems, purposes and opaque beliefs of a writer. A key way of examining these choices is to look at metaphorical uses of lexis which can reflect the belief systems and concerns of a writer (Lakoff & Johnson, 1980). Metaphors can be identified by comparing the literal and intended meanings of lexis in a text, and metaphorical uses of lexis can be identified where literal meaning differed from intended meanings. This concept of metaphor is applied in investigating metaphorical uses of lexis on the BIH website. Examples of metaphors identified are shown in Table 1 below.

<table>
<thead>
<tr>
<th>Table 1 Metaphors in BIH</th>
</tr>
</thead>
<tbody>
<tr>
<td>...located in the <strong>heart</strong> of Bangkok</td>
</tr>
<tr>
<td>...that is <strong>home</strong> to many of our centers</td>
</tr>
<tr>
<td>...we may not have discovered the <strong>fountain</strong> of youth just yet</td>
</tr>
<tr>
<td>It's about time to <strong>feel</strong> the difference...</td>
</tr>
<tr>
<td>...<strong>ours</strong> [laboratory tests] <strong>look to see</strong> if you're healthy</td>
</tr>
<tr>
<td><strong>Hospitality</strong> Residence</td>
</tr>
<tr>
<td><strong>Hospitality</strong> Suites</td>
</tr>
</tbody>
</table>

Out of these metaphors, several interpretations can be made. For instance, the metaphor located in the heart of Bangkok can be interpreted that the writer is making a medical analogy for the role of BIH in Thailand and thus promoting medical issues. Taking into account the purpose of using metaphors which is to present lexical choices that match the purposes and opaque beliefs of a writer, this interpretation might seem to be valid and acceptable. However, while metaphor analysis has been used extensively in CDA, there is a danger of over-interpretation of findings whereby commonly used metaphors are assigned a specific critical interpretation within a given text even though the choice of metaphor reflects general language use rather than specific purposes of the writer (O'Halloran, 2007).

To check whether the metaphors earlier identified reflect the purposes of the writer, all identified metaphors were searched for in the British National Corpus (BNC). The BNC is a 100 million word collection of samples of written and spoken language collected from a wide range of sources (http://www.natcorp.ox.ac.uk/corpus/index.xml). Through a benchmark corpus, like BNC, one can easily estimate how well a corpus fits into it by checking, for instance, clusters or N-grams. N-grams are “multi-word strings of two or more uninterrupted word-forms” (Stubbs and Barth 2003: 62), which are useful in checking words that draws meaning from its context (or co-text as used in Corpus Linguistics and Systemic Functional Linguistics) such as metaphors. Any metaphors appearing in this corpus were excluded from further analysis as they may reflect general language use rather than the belief systems and concerns of the
writer. Only the remaining metaphors were analysed for insights into the purposes and priorities of the writer. For instance, the phrase located in the heart of (a 5 N-gram) occurs 13 times (high frequency as a 5 N-gram) in the BNC and would therefore appear to be a standard phrase whose metaphorical nature does not reflect the specific concerns of the writer, and so was discounted from the interpretation of findings. This means that interpreting the earlier findings as the writer making a medical analogy for the role of BIH in Thailand and thus promoting medical issues is an example of over-interpretation of data. Comparing the metaphors to a benchmark corpus of general English use (the BNC) helped the main researcher decide to exclude metaphor analysis as a method for examining the rest of the data since pilot-testing showed that it is not productive enough. Thus, including it in the investigation may potentially predispose the analyst to over-interpret findings and eventually find whatever s/he wants to find.

**Keyword analysis**

'Keyword' has different meanings depending on whether the term is used as a feature of websites or as a linguistic aspect of a text. In this paper only keywords as linguistic aspect of the text are considered. Keywords can be examined to shed light on the content of the text (e.g. priorities of the author of that text) by virtue of their frequency (Scott, 1997; 2000).

Linguistic keywords can be identified based on frequency. The initial stage is to make decisions about the parameters of the analysis and follow these decisions throughout. An assumption in doing this is to ensure that the same standard of analysis is used. Since keyword analysis involves the examination of lexis, an important parameter that needs to be operationalised is word or lexis, itself. This can be done by vividly identifying the boundaries of a word which can be answered by asking questions such as What is a word?, Are the lemmas of a word considered as separate words or counted as one word?, Are variations in spelling (e.g. American v British English) analysed separately or as one?, Will proper nouns be treated as regular words and analysed as lower case words?, If proper nouns are taken as upper case words, are they analysed as one name or separated into components?, Are words with the same spelling but different meanings separated or analysed as one?, How are phrasal verbs and others that typically go together treated?, Are words in pictures counted as words or included as part of the pictures?, On websites, are words in dropdown menus taken as separate words or only as part of the topics considered?, and many more.

After these questions are raised, two more steps follow. First, these questions are answered. The answer to these questions may vary and could have both positive and negative implications to data analysis and findings. Second, decisions made at this stage should be followed irrespective of the results that come out later. This prevents the analyst from doing any manipulations to the investigation. Table 2 below presents some of the questions that can be considered and how different answers to these questions have implications to the findings of a research paper employing keyword analysis as a method. For instance, defining a word using Answer 1 suggests that words are taken at face value and the analyst may not consider the specific function of words in a clause. One concern that may arise is on the frequency of words. Since a word is only assigned of its part of speech at face value, a word that functions other than its face value function may lead to unreliable results. For instance, the word international is used in the BIH website both as Bumrungrad International and international patients. Identifying the part of speech at face value (e.g. looking at a dictionary) allows international to be analysed only as an adjective (e.g. international patients). There are, however instances where international can be used as a noun (e.g Bumrungrad International). Thus, counting all occurrences under just one part of speech will abnormally increase the frequency of the word.

A second concern is on the interpretation of findings. Questions to ask here include: Are keywords interpreted as they are or is the co-text considered?, If co-text is considered, how many words to the left or right is called co-text?, Are keywords thematised or analysed as
separate entities?, If thematised, what is the basis of thematisation and how is it done?, What are the bases of interpretations, Are absolute or relative frequencies considered in the interpretations?, If relative frequencies are selected, to what extent are these frequencies relative to a benchmark corpus?, What is the role of socio-political context in the analysis?, Is the socio-political context part of the data to be analysed or are they baselines where findings are compared?

Table 2 Questions and their potential answers and implications

<table>
<thead>
<tr>
<th>Question</th>
<th>Answer 1/implication</th>
<th>Answer 2/implication</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) What is a word?</td>
<td>Answer: Any single unit of language in a clause functions independently in relation to other words in the same clause; part of speech classification of words are not based on their function in a clause. Implication: Almost anything can be a word; words exist independently of others in a text</td>
<td>Answer: A unit of language that exists with its co-text and meaning making is done by referring to co-text. Implication: Words derive their meaning from co-text</td>
</tr>
<tr>
<td>2) How proper nouns are treated?</td>
<td>Answer: Treat individual components of the proper noun as lower case words and consider each component separately. Implication: Proper nouns lose their collective meaning</td>
<td>Answer: Treat all components of proper nouns as upper case words and consider them as one Implication: Proper nouns retain their collective meaning</td>
</tr>
<tr>
<td>3) How phrasal verbs are treated?</td>
<td>Answer: Treat each component separately Implication: Words are independent of each other in a clause</td>
<td>Answer: Treat components as one word except when they are separable in the clause Implication: Words in a clause are dependent on each other and derive their meaning from co-text.</td>
</tr>
</tbody>
</table>

To illustrate the importance of problematising the analysis, one of the questions (Are absolute or relative frequencies considered in the interpretations?) is answered below by considering the top five absolute and relative frequencies generated from the main project of one of the researchers (See Table 3).

Table 3 Absolute and relative frequencies

<table>
<thead>
<tr>
<th>Absolute frequencies</th>
<th>Relative frequencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>and</td>
<td>center</td>
</tr>
<tr>
<td>the</td>
<td>bumrungrad</td>
</tr>
<tr>
<td>center</td>
<td>international</td>
</tr>
<tr>
<td>to</td>
<td>services</td>
</tr>
<tr>
<td>Bumrungrad</td>
<td>surgery</td>
</tr>
</tbody>
</table>
Some initial implications can be drawn from the high frequency of words on the left column: And reflects the frequent use of conjoint phrases (e.g. Diagnostic and Therapeutic Center; understand and treat conditions that threaten the lives and lifestyles); the frequency of center shows how the hospital comprises over 30 specialist centers; the frequency of the may be indicative of the use of definite nouns and Bumrungrad reflects the name of the hospital. However, it is difficult to reach stronger conclusions as absolute frequencies reflect general language use as well as the concerns of the particular texts under investigation.

We therefore need to look at relative frequencies and these were calculated by comparing the absolute frequencies against frequencies in the BNC using log likelihood. (See Rayson and Garside, 2000 for details of log likelihood uses). The five keywords with the highest log likelihoods are given on the right column. It can be seen from Table 3 that the keywords based on relative frequencies are largely different from the absolute frequencies on the left, and the keywords reflect the concerns of the webpages more accurately. They can be categorised as follows: words relating to the hospital itself, words relating to medical priorities and words relating to promotional priorities. The reliability of categorisations can be checked using Cohen’s kappa. The highest-ranked keywords for each theme with examples of use and sources are given in Table 4.

### Table 4 Keywords categorised by theme

<table>
<thead>
<tr>
<th>Keyword</th>
<th>Freq</th>
<th>LL</th>
<th>Source</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Words relating to the hospital itself</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bumrungrad</td>
<td>580</td>
<td>3741.9351</td>
<td>All versions</td>
<td>Bumrungrad Hospital; Bumrungrad International</td>
</tr>
<tr>
<td>International</td>
<td>350</td>
<td>2258.0643</td>
<td>All versions</td>
<td>Bumrungrad International is Thailand’s premier hospital</td>
</tr>
<tr>
<td><strong>Center</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Center</td>
<td>788</td>
<td>4014.6498</td>
<td>All Versions</td>
<td>Clinical Research Center; Extensive Diagnostic Center; Horizon Regional Cancer Center</td>
</tr>
<tr>
<td>Surgery</td>
<td>226</td>
<td>1288.4151</td>
<td>All Versions</td>
<td>Coronary artery bypass graft (CABG) open heart surgery; … surgery in a sterile operating theater environment; ...life-saving surgery</td>
</tr>
<tr>
<td><strong>Words relating to promotional priorities</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Services</td>
<td>331</td>
<td>1363.251</td>
<td>All Versions</td>
<td>Food &amp; Beverage Services; …quality of products and services</td>
</tr>
</tbody>
</table>

The findings in Table 4 can now be interpreted. The first theme reflects the topic of the website. Thus, we might expect similar keywords in other communications concerning BIH (e.g. hospital brochures and pamphlets). The second and the third theme show the purposes of the hospital with the medical care provision prioritized with some promotional activities probably to allow this private hospital to acquire enough profits for its continued operation.

As mentioned earlier in this paper, while CDA have advantages in examining texts and discursive practices from a critical perspective, it has also shortcomings as an approach. That is, the analysis and interpretation of findings are vulnerable to biases which could have been mitigated. This explains why guidelines were set at the beginning of this paper and strictly followed throughout.
Early decision-making matters

First, let us state that the discounted methods such as metaphor analysis do not imply that they are no longer useful in examining other texts from a critical perspective. Similarly, productive methods (e.g. keyword analysis) may not be necessarily productive or useful when they are used to investigate texts other than those mentioned in this paper. As shown earlier, it is indispensable to have a set of guidelines in the selection of data and methods for investigation. Methods for the analysis should be chosen based on their productivity when pilot-tested against a sample of the entire research data and not merely because they are favoured by the analyst. Guidelines should also be followed in the interpretation of findings and in the choice of statistical measures for reliability checks. All of these guidelines are followed throughout the research study irrespective of what findings they produce. The key point is that decisions are made as much as possible at the beginning of the research study (e.g. from pilot testing), which leads to the preparation of guidelines for analysis that governs the entire investigation. This restricts the impact of researcher's biases in the study being conducted.

Process: a methodological contribution to CDA

The ultimate goal of this research study is to present an approach to CDA that recognizes and addresses the problematic issues raised by its critiques. Due to choices made by researchers, most investigations include certain levels of subjectivities that may progress to biases. Like most discourse analyses, investigations employing CDA are by nature a subjective enterprise, which exposes the analysts to potential sources of bias throughout the course of the study. While it is impossible to eradicate all of these potential sources of bias, what matters most is whether certain moves are initiated by the analysts to prevent, mitigate or address them. In this paper, we have illustrated that one way of dealing with the potential sources of biases is to problematise the methodological process prior to conducting any major analysis. This allows the analysts to outline a robust methodological pathway that addresses potential loopholes in the investigation. We hope that our illustration contributes to the methodological development of CDA.

References


**The authors**

Jonathan Rante Carreon is a PhD candidate and a recipient of a co-tutelle agreement between King Mongkut’s University of Technology Thonburi and Macquarie University. His interests include corpus-based and multimodal approaches to (critical) discourse analysis that involves data from professional medical communications and medical classroom settings.

Richard Watson Todd is Associate Professor and Head of the Centre for Research and Services at King Mongkut’s University of Technology Thonburi and holds a PhD from the University of Liverpool. His interests include text linguistics, computer applications in language use, and innovative research methodologies.