

# Using Quantitative Methods as a Framework for Qualitative Analyses

Pattamawan Jimarkon

Richard Watson Todd

*Proceedings of the International Conference: Doing Research in Applied Linguistics*

## **Abstract**

*Much of the literature on research in applied linguistics views quantitative and qualitative research as distinct entities embodying contrasting philosophies. In this paper, however, we present an example of how preliminary quantitative analyses of data can inform a qualitative discourse analysis study. Data from an online discussion forum concerning the Thai political crisis of 2010 were initially analysed quantitatively to identify keywords, word clusters, length of postings and user ratings of postings for the contributions from the opposing political factions. Each posting was also rated for level of antagonism and credibility of argumentation. These quantitative data provide an overview of the discussion forum and the patterns of discussion within it which was then used as a framework to guide the main qualitative analysis ensuring that the key revealed meanings and functions were covered in the analysis and reducing potential bias in data analysis and presentation.*

## **Introduction**

Mixed methods research has a long history in disciplines that attempt to explain behavior and social phenomena (Dörnyei, 2007). The practice includes a mix of qualitative and quantitative methods, a mix of quantitative methods or a mix of qualitative methods. The type of mixed methods approach that is most popular and is increasingly employed is the first, the mix of the two, which are often based on different research paradigms.

Single method research is normally criticised by their opposition as inferior and insufficient. In a pure quantitative study, with the focus on theory or hypothesis testing, the researcher may not be sensitive on contextual details. Moreover, it requires a large amount of data to be able to give an effective ground. A qualitative method, on the other hand, is prone to high subjectivity of the researcher and is unlikely to be generalisable. It can only deal with a small size of data, which makes decision making of the overview and conclusive deduction an ordeal.

Four models of mixed methods design are proposed including concurrent design, explanatory sequential design, exploratory sequential design and embedded sequential/concurrent designs (Creswell & Zhang, 2009). The first model, the concurrent design compares and contrasts the results between the two methods to present evidence. Second, the explanatory sequential design utilises the explanation of one set of results to support the other's. In an exploratory sequential design, one method's results are followed by the other's to strengthen the claims made, in the name of generalisability, for instance. In the last design, embedded sequential/concurrent design, a small database is made part of the big database and is used to experiment or enhance the major findings.

Typically, a mixed method research deals with different types of data but it may also mean applying different methods to investigate the same data. Two dimensions of advancements that mixed methods data analyses (MMDA) have to offer can be considered as 1) the design virtue and b) research expertise. The research virtue obtained from mixed methods may refer to the strategies that are used to display trustworthiness of the research such as triangulation, complementarity, development, initiation and expansion (Green et al, 1989). In greater details, four major advantages of MMDA have been put forward (Dörnyei, 2007). First, oversimplification, decontextualisation and reduction of the quantitative analysis can be disputed by in depth meaningful qualitative analysis, while content-specificity and unrepresentativeness can be overcome by the generalisable quantitative analysis. Second, for multi-level analyses, MMDA can add more meaning by converging numbers into words and vice versa. Third, triangulation through multi-methods analyses means increasing validity of the study of the results. Fourth, a study with mixed methods analyses tends to attract more attention from the

audiences from the three paradigms. The research expertise aspect of utilising MMDA includes skill enhancement, trained capacities, growth across discipline boundaries, practical enquiry, dissemination modes and broadened academic opportunities (Brannen, 2005). Chief is the ability that the mixed methods can overcome the challenge and discrepancies of employing either method.

### **The study**

An example of an MMDA of discourse analysis and corpus methodology is presented in this paper. The current study is part of the main study which aimed at investigating the attitudes of the two opposing sides, Reds and Yellows, and how they communicated with each other in an online political discussion forum during the 2010 political unrest, with the focus on the levels of civility and argumentation shown in their postings.

The discourse of political online forums has become an interest for analyses as the online sphere plays an increasingly pivotal role in people's political involvement. Such spaces allow what can be described as deliberative discourse of which patterns are important to understanding the people's contribution to the society. When people feel involved in the topic presented on the political forums, messages posted on posted become robust and heated discussion. Because the anonymity that the spaces have to offer is a platform for ideas, opinions and thoughts are shared freely while feeling safe. Online forums that got attention of audience or are relevant to a wider group of people are composed of long and complex threads of comments.

### **Data magnitude**

The data set consisted of 372 messages with 24,023 words (approximately 65 words per message on average). Based on username, there were 173 contributors to the forum, and most contributors to the discussion posted 5 or fewer messages with only 3 contributors posting more than 10 and none more than 20. The postings were a mixture of separate messages and threads of messages.

In analysing the 372 online comments, first qualitatively, saliency of particular words or phrases was used to identify antagonism and agonism. In addition, the data were evaluated on the basis of argumentative discourse, looking at the extent to which opinions, arguments and evidence were presented in the comments. Associated attributes of keywords and other semantic elements and surrounding words, phrases and semiotic occurrences were investigated.

The purpose of the qualitative analysis was to examine the general characteristics of messages in isolation, how messages link to each other and how communication styles change through the discourse. During the process of coding this set of long and complex data, there were decisions to make which data chunks to code and which to pull out, and more importantly which evolves a story to tell. The data needed to be sharpened, sorted, focused, discarded and organised in such a way that the final conclusions can be drawn and verified. The data that was considered irrelevant was eliminated or extracted.

That being mentioned, there are two methodological issues to be addressed, which are the foci of this paper.

Issue 1: How can we trust that the data presented is representative of all of the available data?

Issue 2: How can we conduct a discourse analysis of a very long data set in a way which is comparable to a qualitative analysis of a smaller set?

### **The study's mixed methods data analysis**

Discourse analysis and corpus linguistics were once clearly segregated as the former depends on the text's integrity, the latter the decontextualisation of the text (Breeze, 2011) but recently the marriage of the two methods (e.g. Biber et al., 2007; Upton & Cohen, 2009) is

recognised as a good opportunity to develop new research questions or theories, and to collect additional data for clarification and exploration. A corpus can shed light to the pattern and routine in the text (Stubbs, 2007) by making possible analysis of a large set of textual data.

The underlying rationale for the study's MMDA is best explained as an exploratory sequential design. To illustrate, the goal of the study is to explore the constructs of civility and argumentation in the population of online postings. After the qualitative analysis was conceptualised, a quantitative analysis, in this case a corpus analysis, was conducted as a follow-up to guarantee the generalisability of the sample data. That is to say that the study was driven by the qualitative analysis but the quantitative analysis also received the emphasis.

#### *Solution to issue 1*

The first issue concerns with the fact that while coding the data the researchers carefully read and reread the data, looking for keywords, trends, themes, or ideas in the data that will help outline the analysis. How can we be sure that all the collected data can be made relevant to the detailed analyses of each comment and each thread? The following analyses of threads can show how detailed the process can be.

##### *Sample thread 1*

This is not democracy. The Red Shirts represent nothing but Mob rule. Hell bent on making the entire Kingdom of Thailand surrender to them - no quarter. They should be treated as traitors and shot. (coded as civility = 1; argumentation = 1)

Really? I think the person who led the military coup to over through the original Prime Minister and not hold elections for the new role should but treated as traitors and shot. This could all be solved if he dissolved the parliament and let the people vote for who they want to lead them. (civility = 3; argumentation = 2)

That would probably increase the violence, not diminish it. (civility = 4; argumentation = 2)

really? LizardKing87? [username] This is false information. The election has been hold and two governments have been formed. This is the third government from the same election. Please do some research in wiki- the information is out there man. (civility = 3; argumentation = 3)

Thais are peaceful people, you sound like you have hatred in your heart. (civility = 3; argumentation = 2)

Yes! Nothing better than using force to solve an issue of "force"! This would be like saying that since the previous ruler made so many mistakes that it entitles you to make the same mistakes again....since of course the previous ruler got to do it? Right? Its ironic. And childish. (civility = 2; argumentation = 3)

In this thread, cohesion is created through repetition of phrases, such as treated as traitors and shot, and by directly addressing content of previous messages (such as the third reply addressing the issue of elections introduced in the first reply). In terms of antagonism, the antagonistic starter leads to conciliatory replies which imply that the initial posting is promoting unnecessary violence. In isolation, threads, like this one, may promote agonism, whereas the other threads in the forum promote antagonism.

Such threads which may promote agonism are, however, outnumbered by threads which promote antagonism, such as sample threads 2 and 3.

*Sample thread 2*

My guess is that this could turn into another Tiananmen Square. Let's hope not. (civility = 4; argumentation = 2)

It is the red protesters that behave more like fascists by intimidating any group that dares to oppose their agenda. A government crackdown would be an improvement over the anarchy on the streets now. Everyday citizens are the victims here and they are only oppressed by the red shirts. It is a completely different scenario than Tiananmen Square. (civility = 2; argumentation = 3)

*Sample thread 3*

Red shirt are not the majority of the whole population, however if there is an election the current government will lose by big margin because the red are the majority of the people who can register to vote. This problem in Thailand is an accumulation of social injustice that exist in the country for many many years. It is not just Thaksin and his money. The problem is real and it needs long time to solve and heal. The poor are fed up with inequality in the country. (civility = 4; argumentation = 3)

what!?! red shirt are the majority of people who can vote? such a liar. Even until this stage they still use abused information to stir the mop up for more violence!!! this's not DEMOCRACY!! you are to destroy our country!!! I still remember what you did last summer!!! you burnt our CITY!!! NO MORE RED SHIRT!!! (civility = 1; argumentation = 2)

No doubt, Thaksin is behind this mob. The evidence is his appearance on the red shirts stage's monitor and display. He is the symbol this mob. His true democracy is to dismiss all his criminal cases, to return all of his corrupted money and to bring him back to the power. I believe that all Thai people have the same opportunity. For example of Mr. Thaksin, he was a normal people but now he is a billionaire. Is there no opportunity? (civility = 2; argumentation = 3)

In these two threads, a reasonable civil starter draws responses which appear likely to exacerbate feelings. In sample thread 2, the civil starter message hopes for a peaceful resolution, but leads to a reaction which redefines the actors in such a way that a violent ending to the protests appears to be preferred. In sample thread 3, the reasoned starter concerns injustice and inequality. However, the responses focus on side issues in the original message - the voting preferences of the majority, and the role of Thaksin - and use these as opportunities to attack the original contributor. The fact that such threads outnumber those which move towards agonism again suggests that the discussion forum promotes antagonism.

From the three sample threads, analyses are easily influenced by selection of threads and comments that can fit nicely into the researcher's personal biases and idiosyncrasies. However, this can be counteracted by providing a contextual summary of the whole data set. To make the qualitative data more meaningful, average numbers of words, Likes and, particularly, levels of civility and argumentation are presented to show a numeric overview of each side.

Table 1 suggests that Red messages were generally longer, were more civil and contained more or clearer argumentation, but that Yellow messages attracted a greater number of positive ratings (perhaps due to the greater number of Yellows in the forum community).

**Table 1** Average numbers of words, likes and levels of civility and argumentation

	All messages	Red	Yellow	Neutral	Unclear
Average no. of words	56.94	83.93	57.35	68.77	23.50
Average no. of Likes	5.45	4.78	7.06	3.27	3.44
Average level of civility	2.74	2.84	2.13	3.53	3.37
Average level of argumentation	2.25	2.78	1.99	2.87	1.92

When the data is composed of more than one group of data, decisions can be made to chunk or subgroup the data for certain characteristics to emerge. In the present study, the main corpus was divided into two sub-corpora, Red Shirt and Yellow Shirt. Keywords for Reds and Yellows against one another were identified.

**Table 2** Keywords for Reds and Yellows

Ranking	Red Shirts		Yellow Shirts	
	Keyword	Log likelihood	Keyword	Log likelihood
1	Democrat	27.037	Red	39.025
2	Middle	19.240	Shirt	24.180
3	Military	18.771	We	21.375
4	The	17.913	Our	20.487
5	PAD	16.835	Very	14.292
6	Supported	13.676	Please	13.577
7	Bought	12.149	Democracy	12.463
8	Party	12.072	Leaders	12.148
9	Because	11.222	Peace	12.148
10	Or	10.883	Bring	10.719

Only a corpus-analysis can reveal the salient pattern from the keywords in Table 2, both Red and Yellow messages focusing on the other side. The Reds discuss the Democrat party (the main party in the current government led by Abhisit), the middle class, the military, and the PAD (the Yellow group that organised demonstrations when Thaksin was prime minister). The Yellows write about Red shirt leaders. Focusing the discussion on the opposing side is more likely to lead to antagonistic discourse than talking about one's own beliefs. Another pattern confirming previous findings is that the Yellows use lexis that increases the strength of their statements: very and please (when used with an imperative as in the Yellow sub-corpus). Finally, there is some indication that the Yellows are predominantly Thai (through the use of we and our) and that the Reds are foreigners (the use of the, since Thais tend to use far fewer articles in English than speakers of other languages).

The quantitative analyses conducted provide 'in-built quality safeguards' (Dörnyei, 2007, p. 56) to the selected threads, being representative of the whole large data. The three threads showed that the Reds were more civil and provided better argumentation than the Yellows and the quantitative analyses can ensure the interpretation of the data.

#### *Solution to Issue 2*

The second issue to be presented is how to deal with a large set of data. First, we need to consider how the data can be made relevant to the research questions. Going back to the main study of this paper, it aimed at finding out the trend of reconciliation between the two opposing sides, Reds and Yellows. Therefore, an overall picture of the development in civility and

argumentation of the whole forum was given by dividing the corpus with 372 postings into two halves.

**Table 3** Development of characteristics of messages through the forum

	First half of forum (postings 1-186)	Second half of forum (postings 187-372)
Number of Red messages	14	42
Number of Yellow messages	126	61
Number of Neutral messages	16	44
Number of Unclear messages	30	39
Average level of civility	2.51	2.97
Average level of argumentation	1.99	2.51

From the qualitative analyses of threads, it was found that the levels of civility and argumentation changed as the threads developed. In some threads, as the argument went on, the discussion became heated. In the others, civility increased as the postings employed more argumentation. Table 3 can give an overview of this trend of this development. It suggests that, as a higher proportion of messages comes from Reds as the forum develops, the overall discussion becomes more civil and contains more argumentation. This could imply that the forum is acting, in a minor way, as a tool for reconciliation. However, there is also some evidence, as shown in other messages, that many of the Red messages were posted by non-Thais, especially Americans, while most Yellow messages were posted by Thais. This commonality can only be discovered from the analyses of the whole set of data quantitatively.

### Conclusion

The general characteristics of how argumentative and civil messages were identified in isolation, between postings and through the discourse. With the quantitative analysis supplemented into the qualitative research design, we could look at the whole forum. The sheer quantity of data suggests a quantitative approach is more likely to highlight patterns. The data length made possible framing of the analysis, chunking the data, subgrouping data and focusing on bigger characteristics. Together, the mixed methods could offer “a complete evaluation of online forums’ deliberative potentialities” (Janssen & Kies, 2004, p. 11). Triangulation and generalisability were achieved through the use of corpora while maintaining the rich and observant qualitative findings throughout the analyses.

The process of employing MMDA in the study has enabled us to achieve the two advancements of research design and research expertise. First, the methods union proved to have provided valid, balanced and comprehensive research findings. Representativeness of the data was achieved by giving the corpus-based data overview. Second, it has enriched and enhanced our expertise as MMD analysts. To combine two methods, we underwent the research cycle – making assumption, generating and testing hypotheses, research design, analysis and making inferences – through the two different paradigms. The experience of trying out the mixed methods has enabled us to be gain methodological skills to handle both data analyses. We also hope that this methods mix will appeal to researchers that are interested in investigating data that are long, multi-layered, dynamic and complex in nature.

The decision about using the MMDA must encompass the pre-defined analysis objective. The objective, accompanied by analysis plan, is based on the research questions. In the analysis plan, mode of analysis must be defined - whether it is to be exploratory, descriptive, or comparative in nature. One of the most important things that the researcher needs address is what MMDA can offer. Despite its complementary strengths, the researcher must be aware that it can be ‘a distraction to serious research’ (Dörnyei, 2007, p. 167). The researcher must take into

consideration the research goal established and theoretical stance adopted and not become overambitious to use anything or everything.

## References

- Biber, D., Connor, U & Upton, T. (2007). *Discourse on the Move: Using Corpus Analysis to Describe Discourse Structure*. Amsterdam: John Benjamins Publishing.
- Brannen, J. (2005). "Mixed methods research: A discussion paper". Retrieved 15 May 2011 from the ESRC National Centre for Research Methods website: <http://www.ncrm.ac.uk/publications/methodsreview/MethodsReviewPaperNCRM-005.pdf>
- Breeze, R. (2011). Disciplinary values in legal discourse: a corpus study. *Ibérica*, 21, 93-116.
- Creswell, J. & Zhang, W. (2009). The application of mixed methods designs to trauma research. *Journal of Traumatic Stress*, 22(6), pp. 612–621.
- Dörnyei, Z. (2007). *Research methods in applied linguistics*. OUP.
- Greene, Jennifer C., Caracelli, Valerie J. and Graham, Wendy F. 1989. "Toward a conceptual framework for mixed-method evaluation design." *Educational Evaluation and Policy Analysis*, 11(3), pp. 255-74.
- Schiffedercker, K.E. & Reed, V. A. (2009). Using mixed methods research in medical education: basic guidelines for researchers. *Medical Education*, 43, pp. 637–64.
- Stubbs, M. (2007) On texts, corpora and models of language. In M Hoey, M Mahlberg, M Stubbs & W Teubert *Text, Discourse and Corpora*. London: Continuum.
- Upton, T. & Cohen, M. (2009). An approach to corpus-based discourse analysis: The move analysis as example. *Discourse Studies*, 11(5), pp. 585-605.

## The authors

Pattamawan Jimarkon works for SoLA KMUTT, Thailand. She received her PhD in Education from the University of Nottingham. Her research interests lie in discourse analysis, spoken language, learner interaction and classroom language.

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.