



Unraveling linguistic sexism & sex role stereotyping in Malaysian English Language textbooks: The Wordsmith Tools way

YUEN CHEE KEONG¹, MOHD. SUBAKIR MOHD. YASIN¹,
KESUMAWATI ABU BAKAR¹, AZHAR JALUDIN¹
& BAHYAH ABDUL HAMID²

ABSTRAK

Analisis secara kuantitatif terhadap seksisme linguistik dan penonjolan peranan stereotaip gender yang wujud di dalam dokumen-dokumen bercetak seperti buku teks sekolah lazimnya dijalankan dengan menggunakan kaedah pengiraan peratusan. Dalam kajian lepas yang telah dijalankan di Malaysia seperti yang dilakukan oleh Saedah Siraj (1990) ke atas buku teks bahasa Melayu dan oleh Chandran dan Mardziah Abdullah (2003) ke atas buku teks bahasa Inggeris telah menggunakan kaedah pengiraan kekerapan secara tradisi iaitu secara manual di dalam menentukan prejudis gender. Kaedah pengiraan secara manual ini bukan sahaja mudah mengakibatkan ketidaktepatan penemuan dan analisis bahkan memakan masa yang lama untuk disiapkan, dan dengan tidak langsung akan mempengaruhi kesahihan dan integriti penemuan kajian berkenaan. Untuk mengatasi kelemahan-kelemahan di dalam analisis data ini, satu kaedah yang lebih mantap dan menjimatkan masa perlu digunakan. Berpandukan kepada satu kajian yang sedang dijalankan yang membuat analisis kandungan, lebih khususnya, analisis teks, makalah ini akan memaparkan bagaimana instrumen Wordsmith Tools 4 (Oxford University Press 2004) boleh dimanfaatkan untuk menganalisis dokumen bercetak dan dapat memberikan keputusan yang lebih sahih dan cekap dalam mengenal pasti seksisme linguistik dan penonjolan peranan stereotaip gender di dalam buku teks Bahasa Inggeris di Malaysia.

INTRODUCTION

Analysis of linguistic sexism and sex role stereotyping in written documents may have a long history in first world countries (Luara Madson 1999) but it is not the case with third world countries in particular countries that have first

world infrastructure but third world thinking. Such a development is indeed alarming in today's world of multiliteracies where the realities of local diversity and global connectedness (Cope & Kalantzis 2000) implore us to "fit in" with the demands of the new capitalism (Paul Gee 2000). Therefore, it is important for Malaysians to examine linguistic sexism and sex role stereotyping to keep abreast with the physical development of the nation and the best place to begin such an endeavour is in school textbooks, the very source of image molding for a fledging nation on the verge of gaining first world status such as Malaysia.

Although attempts have been made to study linguistic sexism and sex role stereotyping in Malaysian textbooks, such studies were far and few. Until recently, only three known studies have been undertaken: one on Malay school textbooks (Saedah Siraj 1990), another on English language textbooks (Sanda Chandran & Mardziah Abdullah 2003) and the last on the perception of teachers who used English language textbooks in the classroom (Sanda Chandran 2005). These studies were not only limited in scope, since only textbooks of only one form was studied instead of all the textbooks in either the six levels of primary school or the six levels of secondary school, but also in depth as most were focused on the perception rather than the actual description of the written evidence. One possible reason for the dearth of studies on linguistic sexism and sex role stereotyping in Malaysia in general and in English language textbooks in particular is that these studies were conducted manually.

However, in this day and age of globalization, the traditional manual approach may neither be practical nor effective given its limitations stated above. Hence, there is an immediate need to use new and innovative methods to study linguistic sexism and sex role stereotyping. One such way is the use of the Wordsmith Tools (WST) (Oxford University Press 2004) software programme.

This paper discusses the research methodology that combines the use of the quantitative research tool WST and the qualitative Contrastive Discourse Analysis (CDA) approach based on an on-going Fundamental Research Grant Scheme study entitled "Linguistic Sexism and Sex Role Stereotyping in Malaysian KBSR and KBSM English Language School Textbooks: Implications to Nation Building". This two-year research beginning November 2006 was undertaken under the 9th Malaysia Plan to determine linguistic sexism and sex role stereotyping in Malaysian English language textbooks and how they could affect nation building. In order to process the large quantities of text in the study an appropriate methodology which was not only efficient but also precise had to be developed. The efficiency and accuracy of the content analysis was crucial because the findings of the study could affect the development of the country and its people especially human capital as outlined in the 9th Malaysia Plan.

STATEMENT OF THE PROBLEM

The FRGS study mentioned above required a sound methodology that could process a large quantity of raw data since 22 Malaysian English language school textbooks taken from all levels, from Primary One to Secondary Five of the Malaysian education system and from schools all over Peninsular Malaysia, had been identified for the research. One option was to adopt the existing traditional method of frequency count by hand. Another was to design a new and more reliable and efficient method to undertake the above challenge.

PURPOSE OF STUDY

The objective of this study was to design an innovative and reliable method in the examination of large quantities of written documents. More specifically, it was a study on the linguistic sexism and sex role stereotyping in Malaysian English language textbooks using the WST software programme and CDA.

A secondary aim of the study was to determine the effectiveness of WST and CDA over the traditional manual approach in document analysis.

RESEARCH QUESTIONS

Based on the above objectives, the following research questions were formulated:

- (1) Which is the most reliable and efficient methodology in processing and analyzing large quantities of written texts? In other words, what is a good methodological framework for text analysis?
- (2) How can the methodology or framework assist in processing and analyzing linguistic sexism and sex role stereotyping in Malaysian English language textbooks?

STUDIES USING MANUAL FREQUENCY COUNT IN ANALYZING WRITTEN DATA

Sanda Chandran & Muhammed Shahriar Haque (2005) looked at whether English language teachers were aware of gender bias in the textbooks they used and how they dealt with the issue, if they were. Sanda Chandran & Muhammed Shahriar Haque's research method was a questionnaire administered to 30 teachers from three different secondary schools in Seremban, the state capital

of Negeri Sembilan. Their findings revealed that English language teachers in Malaysia were unaware of gender bias in English language textbooks. For those who were, there was no open discussion because of work pressure. This study conducted quantitatively using a questionnaire appears to be fairly objective. However, the data collected from the questionnaire was analyzed manually which might have limited the range and depth of the findings of the study. Hence, only surface findings of teachers' perception of gender biasness were extracted from the data.

The traditional manual approach in the data analysis might have curbed opportunities to exploit the data since the analysis of the data is highly restricted. The scope and scale of the study were also limited. 30 teachers in three schools can hardly be representative of the views of English language teachers in Malaysia. This shortcoming seems to raise the question of validity and accuracy of the findings in the study. In addition, the study was not backed by concrete examples of gender bias extracted from textbooks, the real aim of the research. These limitations could have been avoided if a more innovative method was used in analyzing the data.

In another gender biasness study by Saedah Siraj (1990), the manual analysis was also to analyze Malay language elementary reading textbooks. She conducted the study on five Malay language textbooks. Her findings revealed that the Malay textbooks were littered with gender biasness for men. Males dominated as main characters in the reading texts and they were also described as more competent intellectually, having better personal development and social interpersonal skills than their female counterparts. Besides that, the males were disciplined, patient, energetic, prolific writers and powerful leaders. Saedah Siraj concluded that the gender stereotyping of male and female characters in Malay reading textbooks might affect children's attitude towards themselves and others.

The above results were derived using a manual frequency count of the linguistic forms that represented gender biasness in the textbooks. The forms were "bits and pieces" of words and phrases extracted from the reading textbooks. There were no comprehensive listing of all the words found in all the five elementary textbooks. Her findings, therefore, may not be as accurate as revealed in her study.

Yet another gender bias study in Malaysian English language textbooks was conducted by Sanda Chandran & Mardziah Abdullah (2003) who looked at four Form 4 secondary textbooks used by schools in the capital of one the states in Malaysia. The textbooks were analysed for four aspects: male and female characters, terms of address, social roles, and activities associated with male and female characters. A secondary aim of the study was to determine the relationship between the frequency of occurrence of gender bias in textbooks and the gender of writers.

Like Saedah Siraj, Sanda Chandran & Mardziah Abdullah utilised frequency count in their data analysis. However, they did include a co-rater to address issues of reliability and validity of the data analysed. Their findings revealed that male dominance in gender was evident in all the four identified categories and in all the four textbooks irrespective of the gender of the writers.

The methodology used giving only percentages of the identified linguistic forms without any reference to the total number of all the words found in all the four textbooks but instead the percentages were calculated based on the total number of words identified by the researchers. This method of calculation appears to affect the reliability of the study since it does not give a complete picture of the actual situation. The researchers could have resorted to selective word choice focusing on only words that they wanted to identify and leaving out the rest which could have revealed gender bias. A possible reason for the selective word choice could be attributed to the fact that the manual frequency count of the identified linguistic forms was time consuming. Such a methodology, therefore, could have an effect on the real findings of their study.

In conclusion, all traditional manual frequency count studies identified above had very limited data samples. Saedah Siraj's (1990) used only five elementary reading textbooks while Sanda Chandran & Mardziah Abdullah's (2003) had only four. Such small scaled studies which are confined to isolated locales in Malaysia cannot be used by these researchers to justify the issue of gender biasness for the whole nation. Moreover, in Saedah Siraj's study, the selection of elementary reading textbooks for the data samples meant that not much textual content was available for analysis given the nature of children textbooks where illustrations and pictures form a major component in attracting young readers. Secondly, the frequency counts were converted into percentages manually raising the question of reliability and validity, in addition to the laborious task of conversion. Because of these concerns, a more reliable methodology which could handle enormous quantity of texts in the shortest time possible had to be adopted.

STUDIES USING WST IN ANALYZING WRITTEN DATA

Bahiyah Abdul Hamid & Kesumawati Abu Bakar's (2006) study examined the construction of adolescent identities in personal advertisements in magazines in Malaysia. This quantitative and qualitative study used document analysis as its research instrument. The data was analyzed quantitatively using WST and qualitatively using Fairclough's (1989) 3-stage CDA. WST provided the researchers a frequency list of all the words in the 1000 adolescent personal advertisements, the collocation of the words from the data and the key words of

unusually high frequency of occurrence. Having determined these three categories of data from the analysis of the advertisements, the data was further subjected to the 3 stage CDA for adolescent identity patterning. The findings revealed that Malaysian adolescents were able to create unique individual discourses about themselves through personal advertisements. Furthermore, they were able to recognize the need to belong to a group and of being accepted in it because peer group offered them security, attention and dignity.

Nevertheless, the overall findings of the study belied the rich data it extracted from the personal advertisements. From the wordlist, concordance and key words generated from the data with WST, the researchers were able to categorize the male and female adolescents' identities into six aspects of self, namely, gender, interests/hobbies, ethnicity, positive social/relational characteristics/affective behaviour/psychological traits, physical appearance and religion. In each aspect, descriptors (words, phrases and clauses) and their frequency of occurrence were identified making the interpretation of the data easier. Given the capability and capacity of WST in generating a variety of the data and also in its detailed forms, it has helped the study to provide findings that were not only valid but meaningful.

In addition, the study analysed 1000 adolescent personal advertisements from three annual publications of a popular magazine in Malaysia. The researchers would not have produced the rich results, especially the corpus of personal ads from adolescents and youths, if the traditional method of counting manually was used for their study.

Yuen Chee Keong (2007) in another study utilized the WST to examine the written business correspondence of Malaysian business executives in a financial organization. Although this study did not look at linguistic sexism and sex role stereotyping, it is related to this current study in terms of the research methods used. Yuen utilized two research instruments: document examination to collect the data and the Wordsmith Tools to analyze the data. These were the same instruments used in this study. Because of this, a review of Yuen's research is important.

In Yuen's ethnographic case study, 20 business executives were selected as participants who produced 126 pieces of written business correspondence comprising letters, memos, reports and e-mails over a four-month period. The findings of the study revealed that localized or indigenized language forms were prevalent in the written business correspondence of the Malaysian executives. The WST was used to determine the quantitative aspect of the research. It assisted in the frequency count of the localized language forms in written data. The frequency count was crucial in screening out idiosyncratic usage from the language norms of the participants. Without WST it would have been a laborious task counting the frequency of occurrence of a particular word by hand.

The WST methodology used by Yuen (2007) facilitated the processing of the various forms of business correspondence of the participants in the study while the traditional method used by respective researchers mentioned above did not even specify the various genres found in school textbooks. Such an omission might have an effect on the findings of those studies. In addition, the generated wordlist helped Yuen to overcome a crucial step in his analysis. He was able to filter out idiosyncratic usage of a particular word with the wordlist as the objective of his study was to identify localized language forms in the written business correspondence of his participants. This step provided added validity and reliability to his study.

From the literature review of the methodologies used in text analysis, a comparison highlighting the advantages and disadvantages of both methods is made in the Table 1 below:

TABLE 1: Comparison between traditional Manual Frequency Count & Innovative WST

Features	Method	
	Traditional Manual FrequencyCount	Innovative WST software
Words for analysis	Manual generation of words	Facilitates the building up of a corpus of data from the computer generated wordlist
Functions	--	Allow concordance of words and the distribution of key words
Scope of study	Small scale	Can process large quantities of written texts or documents
Validity and reliability	Unreliable	Valid and reliable
Efficiency	Time consuming	Save time
Use	Laborious manual counting	Easy to use as it uses current computer technology

METHODOLOGY FOR THE FRGS STUDY

Findings from previous research on gender biasness were used to formulate a methodology which was not only reliable but also capable of processing large quantities of written texts, for the larger on-going FRGS study. Given the limitations of the traditional manual frequency count in text analysis above, it was important that a new and innovative methodology be devised.

The methodology of the study had to rely on both the quantitative and qualitative approaches since document examination was its main objective. 22 Malaysian English language school textbooks currently in use from Primary One to Secondary Five of the FRGS research study needed to be examined to determine whether linguistic sexism and sex role stereotyping exist in these textbooks. The methodology was divided into two parts: the first was a quantitative analysis of the textbooks using the WST software and the second was a qualitative analysis of the quantitative data using CDA. This methodology was adopted to keep in line with the secondary objective of the study which was to determine the capability of modern technology over the traditional manual approach in document examination.

STAGES OF THE WST APPROACH

The first step involved the scanning of every page of the textbooks into word document files. Then the files were converted into text document file because the WST could only read text files.

At the third stage, a wordlist of all the words in the text files was generated using the Wordlist function of the software. The wordlist provided the frequency and percentage of each word in descending order found in all the 22 textbooks. This step is actually the production of a corpus. For the FRGS research, a corpus of KBSR and KBSM English language school textbooks needed to be developed which is the output of the study too. In order words, the text files are the corpus of data and WST actually facilitates the building up of a corpus. Without a corpus, there will be no study. Furthermore, WST can only work effectively if the corpus is large enough to get a decent salient output especially from wordlist. For instance, Bahiyah Abdul Hamid & Kesumawati Abu Bakar (2006) had a corpus of personal ads from adolescents and youths while Yuen Chee Keong (2007) developed a corpus of written business correspondence.

For the next stage, words which reflected linguistic sexism and sex role stereotyping were identified from the wordlist. The identified words were then put through a concordance procedure using the Wordsmith Tools in the fifth

stage. The concordance of the identified word would provide the collocation of the word or the type of company the identified word kept.

The sixth and final step was to generate the keywords found in the textbooks using the Key Word function of the Wordsmith Tools. With the Wordlist, Concordance of the identified words, and the Keywords generated by the Wordsmith Tools, the analysis of the data extracted from the written document could begin. This completed the quantitative part of the methodology. A summary of the WST methodology is illustrated in the Table 2 below:

TABLE 2: Stages of the WST methodology

Stage	Function
1.	Scan every page of the textbooks into word document files.
2.	Convert word document files into text document files.
3.	Generate a wordlist giving the order and frequency of occurrence with percentages of every word in textbooks.
4.	From the wordlist, words which reflected linguistic sexism and sex role stereotyping were identified.
5.	Concordance of the identified words is carried out with the Concordance function.
6.	Generate the keywords found in the textbooks using the Key Word.

ANALYSIS OF THE DATA

The second part of the methodology was qualitative in nature since this was a quantitative cum qualitative study. A descriptive-interpretive approach using Fairclough's (1989) 3-stage CDA which views language use as a social practice, shaped by and shaping social identities, social relations and systems of knowledge and beliefs of individuals, was used to analyse the data. In the first stage, an analysis of formal textual features was conducted to facilitate formal text description. The interpretation of the relationship between text and social interaction was carried out in stage two. This enabled the analysis of the norms, mental standards of socially accepted behaviour in specific roles/relationships

used to produce, receive, and interpret the text. The last stage explained the relationship between social context and cultural context. The 3-stages CDA approach can be illustrated in Table 3 below:

TABLE 3: The 3-stage CDA

Stage	Function
1.	Analysis of formal textual features.
2.	Interpretation of the relationship between text and social interaction.
3.	Explanation of the relationship between social context and cultural context.

CONCLUSION

This paper adopted a problem solving approach in developing a methodology that could analyze large volumes of texts. Much of the research thus far in gender bias analysis followed the traditional method of frequency count of the linguistic forms by hand. This method had revealed that there were problems in terms of validity and reliability. In order to overcome these problems, a new innovative method which uses WST to analyze the text quantitatively and CDA to interpret the quantitatively derived data qualitatively was designed to process the 22 Malaysian English language school textbooks of the FRGS study. This proposed methodological framework linking quantitative and qualitative methods had been tried out in past research by Bahiyah Abdul Hamid & Kesumawati Abu Bakar (2006) and Yuen Chee Keong (2007) and they had proven that the WST methodology to be reliable and effective in processing voluminous amount of written texts. Hence, the adoption of this new and innovative WST methodology over the traditional manual approach of frequency count for the FRGS research was fully justified given its objective of analyzing large volumes of written texts, more specifically, the analysis of linguistic sexism and sex role stereotyping in Malaysian English language school textbooks.

REFERENCES

- Bahiyah Abdul Hamid & Kesumawati Abu Bakar. 2006. Articulating male and female adolescent identities via the language of personal advertisements: A Malaysian perspective. Paper presented at Globe 2006, Communicating Across Age Groups: Age, Language and Society, Warsaw, Poland, 21-23 September 2006.
- Cope, Bill & Kalantzis, Mary. 2000. Multiliteracies: The beginnings of an idea, in *Multiliteracies: Literacy Learning & the Design of Social Futures*, edited by Bill Cope and Mary Kalantzis. London: Routledge.
- Fairclough, Norman. 1989. *Language & Power*. London: Longman.
- Gee James Paul. 2000. New people in new worlds: Networks the new capitalism and schools, in *Multiliteracies: Literacy Learning & the Design of Social Futures*, Bill Cope & Mary Kalantzis (editor). London: Routledge.
- Madson, Luara. 1999. Does alternating between masculine and feminine pronouns eliminate perceived gender bias in text? [Http://www.findarticles.com/p/articles/mi_m2294/is_1999](http://www.findarticles.com/p/articles/mi_m2294/is_1999).
- Oxford University Press. 2004. *Wordsmith Tools*. London: Oxford University Press.
- Saedah Siraj. 1990. An analysis of gender stereotypes in Malay language elementary reading textbooks: Implications for Malaysian schools. PhD Thesis, University of Pittsburgh.
- Sanda Chandran and Mardziah Abdullah. 2003. Gender bias in Malaysian English language textbooks. In Jayakaran Mukundan (ed.). *Readings on ELT Materials*. Serdang: Universiti Putra Malaysia Press.
- Sanda Chandran & Muhammed Shahriar Haque. 2005. Gender discourse: Are teachers aware of gender bias in Malaysian English language text Books, in *ELT Matters 1: Issues in English Language Learning & Teaching*.
- Yuen Chee Keong. 2007. Ringing the changes: A study of language variation and use in the written forms of communication in a financial institution in Malaysia. PhD research currently being undertaken at Universiti Kebangsaan Malaysia.

¹ Faculty of Social Sciences & Humanities
Universiti Kebangsaan Malaysia

² Centre For Gender Research (CGR)
& School of Language Studies & Linguistics
Univeriti Kebangsaan Malaysia

