

Comparing Useful Words in Second Language Adult Learners' Writings with BNC Wordlist: A Prototype Approach

**Nor Arifah Mohd Nor^{1*}, Wardatul Akmam Din²
Irma Wani Othman³ & Norazah Mohd. Suki⁴**

*¹⁻³Centre for the Promotion of Knowledge and Language Learning,
Universiti Malaysia Sabah*

⁴Faculty of International Business and Finance, Universiti Malaysia Sabah

ABSTRACT

This research focusses on the preliminary investigation of vocabulary knowledge on teachers undergoing the Teacher Graduate Programme (PPG) in Universiti Malaysia Sabah (UMS). By conducting this research it is possible to determine whether these teachers are familiar with the same set of words listed in BNC therefore have sufficient vocabulary knowledge which they most probably impart to their students. Forty-four adult students were asked to write an English essay based on the topic given by the teacher. Soft copies of their essays were compiled and converted to .txt files which were then transferred and analyzed using the Simple Concordance Program 4.0.9 List of words in BNC (refer rank frequency order cited in ucel.lancs.ac.uk/bncfreq/flists.html, Oct. 23, 2014) and list of words based on frequency gathered from the essays of the PPG students were then compared. It is found that the adult students' competence in using words that are common in BNC wordlist.

Keywords: word list, writing skill, adult students, concordance

Introduction

Academic Wordlist, University Wordlist and British National Corpus are some of the available wordlists aim to compile words for different purposes and for different target groups. These compiled words help teachers and second language learners ascertain what

* Corresponding author: Nor Arifah Mohd Nor, Centre for the Promotion of Knowledge and Language Learning, Universiti Malaysia Sabah . E-mail: arifah_mnor@yahoo.com

words are important to be learned. Why is it important to select what words to learn? According to Laufer (2014, 38) “Vocabulary selection is of vital importance since L2 learners do not have the time and the opportunity to learn words in the same way native speakers do, via language input”. As such, highlighting words that are useful might be helpful in teaching a language to second language learners. As recommended by Kaur (2013) word knowledge is vital because it helps to ensure sufficient literacy of the target language.

In Malaysia, English language went through various positioning (Thirunsaku & Melor, 2014). From a language for the elite during the British era to the reintroduction of English as a medium of instruction in technical subjects (Thirunsaku & Melor, 2012) and the recent restrengthening of English policy in 2012. The latest planning to bring in 300 US Fullbright scholars in Malaysia to help improve the English proficiency in urban and rural areas in 2015 showed how urgent it is to make sure that English language proficiency in schools are raised. Similarly, Teacher Graduate Program (PPG) was introduced in order to upgrade the qualifications of Malaysian school teachers in attaining at least a degree. All these efforts done by the Ministry of Education in Malaysia are to ensure that Malaysians will become a knowledge-based society (Thirunsaku & Melor, 2014).

With all these steps taken it is a question on whether the Malaysian government is approaching the issue of raising English language proficiency among language learners and teachers realistically. What should Malaysian learners learn and what should our teachers teach? Investigating vocabulary knowledge among second language learners is a small step to determine whether policies and measures taken are after all worth the pain.

Purpose of the Study

Specifically, the aim of this study is to conduct a preliminary investigation of vocabulary knowledge on teachers undergoing the Teacher Graduate Programme (PPG) in Universiti Malaysia Sabah (UMS). By conducting this research it is possible to determine whether these teachers are familiar with the same set of words listed in BNC therefore have sufficient vocabulary knowledge which they most probably impart to their students.

Research questions

The following research questions will be investigated:

- 1) What words do these PPG teachers use in their writings?
- 2) Do these PPG teachers demonstrate significant frequency of preferred words compared to British National Corpus wordlist?

Significance of the study

This study acknowledges the level of vocabulary knowledge these PPG teachers have. With the current measures taken by the Ministry of Education (see Thirusanku & Melor, 2014) it might be discriminatory to imply that teachers are not proficient enough because the state of English language that Malaysia holds can vary according to the target group background, the location of the language being used and purpose of using the English language to name a few. Although it is not the focus of this study to measure the level of proficiency of these PPG second language learners, it is significant to reveal the ability of these teachers in using similar frequently used words as compred to BNC wordlist.

Literature review

Vocabulary knowledge deals with many spectrums. It can be simply familiarity towards such words, ability to use them and being able to explain how they can be used (Laufer & Goldstein, 2004). Wordlists are basically words counted based on frequency. The words are ranked by frequency of occurrence grouped either by alphabetical order, theme, purposes, subject matter etc. which in a way assists vocabulary learning. A word list by frequency "provides a rational basis for making sure that learners get the best return for their vocabulary learning effort", (Nation, 1997). Nation (1997) added that it is mainly intended for course writers, not directly for learners.

The British National Corpus (BNC) is a 100 million word collection of samples of written and spoken language from a wide range of sources, designed to represent a wide cross-section of British English from the later part of the 20th century, both spoken and written. The latest edition is the BNC XML Edition, released in 2007. The written part of the BNC (90%) includes, for example, extracts from regional and national newspapers, specialist periodicals and journals for all ages and interests, academic books and popular fiction, published and unpublished letters and memoranda, school and university essays, among many other kinds of text. The spoken part (10%) consists of orthographic transcriptions of unscripted informal conversations (recorded by volunteers selected from different age, region and social classes in a demographically balanced way) and spoken language collected in different contexts, ranging from formal business or government meetings to radio shows. (<http://www.natcorp.ox.ac.uk/corpus/>, retrieved Oct 24, 2014)

Methodology

Prototype Theory by Eleanor Rosch underlies the methodology of this research. This theory asserts that every category is associated to a set of networking criteria and that for each there is a 'core' and 'periphery' boundaries (Tsohatzidis, 1992). In this research it is hypothesized that the participants share a common preferred words which they might probably use if they were to be given the same stimulus which, in this case, the same topic to write. Therefore, in this research the 'core' or preferred word of the group will be identified based on how frequent words are being used in their writings. Since prototype theory focuses on an ideal exemplar for instance chair is considered as a better chair than stool (<http://www.statemaster.com/encyclopedia/Prototype-Theory>, Oct. 25), the concept of sharing a mutual preference towards a category (in this research words) will be used to discuss the collected data.

Forty-four PPG students majoring in TESL participated in this research. As part of their assessments students were asked to write an essay based on the topic "A Day at Work" which generally looked at how they used some tenses in their writings. In this research however, identifying what words the PPG group commonly used was the aim. Students were asked to send their works via email to the instructor. Soft copies of their essays were compiled and converted to .txt files which were then transferred and analyzed using the Simple Concordance Program 4.0.9 List of words in BNC (refer rank frequency order cited in ucel.lancs.ac.uk/bncfreq/flists.html, Oct. 23, 2014) and list of words based on frequency gathered from the essays of the PPG students were then compared.

Discussion

The first objective of this research is to investigate what words the PPG teachers use in their writings. The following screen capture displays part of the output showing words that were used in the collected essays and their frequency of occurrences. It should be clarify here that since this is a small-scale study it is not possible to use all words listed in BNC wordlist. As such only ten most frequent words in BNC wordlist are identified to be compared to ten most frequent words used by the PPG group.

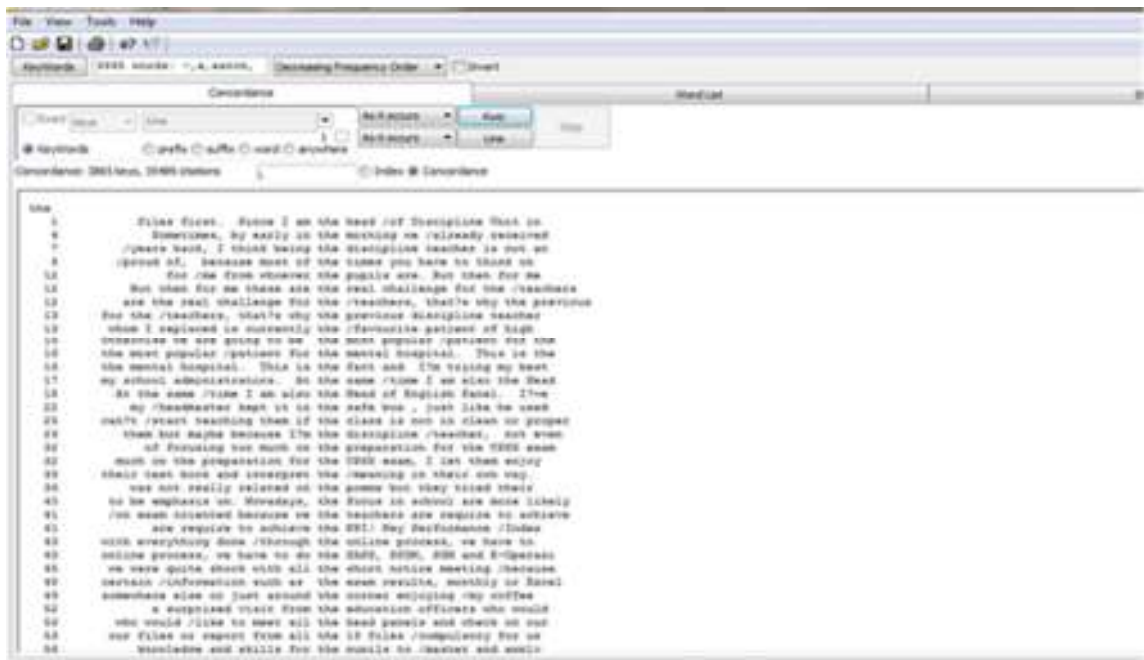
Word List	Frequency
110 the	2200
179 a	1100
407 that	800
101 is	700
178 with	600
103 people	500
111 he	450
106 not	400
176 am	350
100 however	300
102 by	250
112 English	200
113 an	180
107 time	170
104 something	160
89 on	150
77 out	140
71 your	130
86 competition	120
85 challenge	110
82 have	100
89 stop	90
85 too	80
81 great	70
89 exact	60
88 beach	50
86 son	40
85 but	30
82 season	20
82 work	10
81 place	10
85 started	10
25 to	10
27 would	10
34 some	10
34 not	10
1204 in	1100
630 my	1100
611 see	1100
309 school	1100
249 day	1100
224 go	1100
202 have	1100
184 this	1100
187 from	1100
162 out	1100
150 it	1100
110 there	1100
111 also	1100
107 need	1100
96 do	1100
89 some	1100
75 write	1100
70 know	1100
67 about	1100
65 be	1100
62 you	1100
57 been	1100
54 who	1100
51 us	1100
49 good	1100
47 well	1100
46 more	1100
45 I	1100
42 baby	1100
42 has	1100
41 she	1100
40 year	1100
39 team	1100
38 always	1100
34 course	1100
33 as	1100
1142 I	1100
888 at	1100
809 for	1100
204 an	1100
217 that	1100
112 he	1100
103 was	1100
104 teacher	1100
104 year	1100
103 about	1100
104 class	1100
114 because	1100
112 but	1100
104 a	1100
99 can	1100
79 in	1100
73 like	1100
69 just	1100
68 more	1100
65 then	1100
62 before	1100
64 really	1100
63 nothing	1100
63 even	1100
60 see	1100
60 activity	1100
60 something	1100
60 did	1100
60 parent	1100
60 time	1100
60 some	1100
60 few	1100
60 stop	1100
60 give	1100
60 work	1100
60 did	1100
888 not	1100
887 so	1100
880 is	1100
107 at	1100
105 they	1100
105 all	1100
105 did	1100
103 them	1100
100 will	1100
100 was	1100
100 because	1100
110 that	1100
100 when	1100
100 one	1100
88 out	1100
76 other	1100
76 as	1100
68 up	1100
68 what	1100
68 why	1100
60 get	1100
60 being	1100
60 still	1100
60 if	1100
60 next	1100
60 happy	1100
60 remember	1100
60 his	1100
60 program	1100
61 know	1100
60 sometimes	1100
30 his	1100
27 go	1100
26 same	1100
24 one	1100
23 help	1100

Based on the data it was found that ‘the’ is the most preferred word used by the subjects with 2200 occurrences and ‘for’ received 409 occurrences. The following table shows the list of ten most preferred words and the number of occurrences of each word being used by the PPG learners in their essays.

As mentioned before the main objective of the task given to the L2 learners was to test them on their grammar. It is therefore clearly seen that such activity stimulate the use of many grammatical components such as articles, pronouns, prepositions, past tenses etc.

No	Word	Frequency of Occurrences
1	the	2200
2	to	1304
3	I	1162
4	and	986
5	a	679
6	my	653
7	of	599
8	that	427
9	was	411
10	for	409

In order to look at how the word ‘the’ and ‘to’ are being used in their writings the following screen captures give insights of such occurrences. However, it is not the intention of this paper to analyze how these frequently used words are being used in L2 writings.





The second objective of this study is to investigate whether the PPG teachers demonstrate significant frequency of preferred words compared to British National Corpus wordlist? The following is a comparison in terms of ranking between twenty most frequently used words by the PPG learners and the first twenty words listed in BNC wordlist.

No.	BNC wordlist	PPG frequent words	Ranking
1	The	The	1
2	of	to	6 & 8
3	and	I	17
4	a	and	3
5	in	a	4
6	to (inf)	my	-
7	is	Of	2
8	to (prep)	that	12
9	was	was	9
10	it	for	11
11	for	is	7
12	that	it	10
13	with	school	-
14	he	as	-

15	be	at	20
16	on	with	13
17	I	day	-
18	by	their	-
19	's	they	-
20	at	pupils	-

Since the topic of the essay given was “A Day at Work” some of the words used might not be the first twenty words listed in BNC wordlist. Nonetheless, thirteen out of twenty (65%) of the frequently used words that the PPG learners used are listed in the BNC wordlist. The other seven words (35%) are not listed as the first twenty words in BNC (note: these words are listed in the BNC wordlist but the rank will not be highlighted).

From these two research questions it is found that the PPG teachers revealed that they share similar set of useful words as a group which when compared to BNC wordlist showed that they use words that are mostly ranked the highest. Because the nature of the essay writing is not to investigate category, it was not clear whether the participants actually shared a preference towards an ideal exemplar as proposed in Prototype theory. However, the fact that the set of words used by these teachers fall among words that BNC wordlist deemed important shows that these teachers are using words that are important to be known and used.

Inevitably, these teachers show the competence in using words that are common in BNC wordlist. These words were gathered from samples of written and spoken language of various sources representing British English (www.natcorp.ox.ac.uk, Oct. 24, 2014). Thus, the teachers can be concluded as to be in the right platform as far as the Ministry of Education is concerned.

Future research

There are some limitations to the study which will be good to be carried out later in future research. Firstly, the samples of written works merely represent a one event episode. It will be beneficial to look at several samples of writings from the same group of participants at different occasions. Next, the topic of the essay restricts the use of words that can demonstrate preference towards an ideal exemplar. An essay that uses many adjectives and verbs might be better than an essay that tested on grammar tenses or items. In addition, it is also good to look at how such grammatical items being used by each individual. The word ‘to’ for example can sometimes functions as infinitive or preposition. This study did not separate the function because the main aim was to see what words the participants used more frequent than the other.

Conclusion

As a conclusion, this study is a small step of looking at what teachers have been using which might not be of much significant to be overgeneralized. However, it is significant to help demonstrate the notion that Malaysian teachers are in fact using words that are

common to native speakers. Although the words are not content words, which would be more informative, they are significant words to be known and used. These teachers showed that they prefer using them, competent (though not discussed in this paper) to use them and might be of high probability to be proficient enough to teach these words. It comes back to the question whether it is necessary for Malaysian government to rely on external sources to nurture our future generations. As Thirunsaku & Melor (2014) suggested in their research “an appropriate approach and teaching materials have to be selected” and that “... the government should guide our teachers to perform better based on endonormative model instead of spending money on foreign teachers and an appropriate learning theory relevant to our students...” (p.259). Malaysia is a multi-linguistic and multi-cultural country with its own character and multi-faceted environment. Bringing in expertise who might have little knowledge on these aspects could be of little help. Using local variety might be a better option after all.

References

- Kaur, N. (2013). A case study of tertiary learners' capability in lexical knowledge. *Gema Online Journal of Language Studies*, 13(1), 113-126.
- Laufer, B. & Goldstein, Z (2004) Testing vocabulary knowledge: size, strength, and computer adaptiveness. *Language Learning*, 54(3), 399-436
- Laufer, B. (2014). Vocabulary in a second language: selection, acquisition and testing: a commentary of four studies for JALT vocabulary SIG. *Vocabulary Learning and Instruction*, 3(2), 38-46.
- Nation, P. & Waring, R.(n.d.). Vocabulary size, text coverage and word lists. Retrieved October 24, 2014, from www.fltr.ucl.ac.be/fltr/germ/etan/bibs/vocab/cup.html
- Thirunsaku, J. & Md. Yunus, M. (2012). The many faces of Malaysian English. *International Scholarly Research Network Education*. Vol. 2012, 14 pages
- Thirunsaku, J. & Md. Yunus, M. (2014). Status of English in Malaysia. *Asian Social Journal*. 10:14, 254-260.
- Tsohatzidis (1992). *Meanings and prototypes: studies in linguistic categorization*. London: Routledge. [Urel.lancs.ac.uk/bncfreq/flists.html](http://www.lancs.ac.uk/bncfreq/flists.html). Retrieved October 23, 2014.
- <http://www.natcorp.ox.ac.uk>. Retrieved October 24, 2014
- <http://www.natcorp.ox.ac.uk/corpus/>. Retrieved October 24, 2014
- <http://www.statemaster.com/encyclopedia/Prototype-Theory>. Retrieved October 25, 2014.