



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

T-Test for Visualizing Frequently Used Arabic Words

¹R.J.R. Yusof, ¹R. Zainuddin, ¹M.S. Baba and ²Z.M. Yusoff

¹Faculty of Science Computer and Information Technology,
University of Malaya, 50603, Kuala Lumpur, Malaysia

²Academy of Islamic Studies, University of Malaya, 50603, Kuala Lumpur, Malaysia

Abstract: The aim of visualizing the frequently used words is to solve the problem of reading comprehension. This is referring to the case of the non-Arabic speakers of the Muslim community, reading or reciting extensively an Arabic document (the Quran) without comprehension. This study outline an experiment testing whether there is any significant difference on the level of comprehension when images are used as part of the reading material of the Arabic text. It was found that using text only translation, resulted in no significant difference of the level of comprehension and the expected values. However, there is significant difference on the level of comprehension between Arabic text translation of the frequently used words and the text image of the frequently used word.

Key words: Reading comprehension, text visualization, frequently used words, Arabic language and translation

INTRODUCTION

Written text is comprehended by the process of reading. Therefore, the most important reason why people read is to comprehend the text. Comprehending text requires mental power and it is related to the cognitive load of an individual. If the cognitive load is reduced, then it is more likely that the individual will understand the written text. One way of reducing the cognitive load of a human individual while reading is to present the text in the visual form.

There are many work on text visualization systems with the intention to assist the users to become aware of the content of the text document or comprehension of the meaning of the text such as by Fortuna (2005), Weippl (2001), Wise *et al.* (1995), Fang (2006), Weber (2007), Yeap *et al.* (2005), Abbassi and Chen (2007) and others. However, work on Arabic visualization system is in the infancy stage. This study tries to present an experimental result to be used as a basis of development of an Arabic visualization system. Particularly, to solve the problem of users that can read in Arabic but could not comprehend the meaning. It applies to the case of reading the Holy book (to the Muslim) called Al-Quran. The objective of the experiment is to test whether there is any significant difference on the level of comprehension when images are used as part of the reading of the Arabic text. The strategy of using the frequently used words are base on theories of reading comprehension related to word identification such as by Perfetti *et al.* (2007), Lupker

(2007) and Frost (2007). Word identification is a skill and if acquired will help in reading comprehension.

MATERIALS AND METHODS

The materials used in the experiment were the Arabic text and its translation selected from the Chapter 114 or Surah Al-Nas of the Quran. The Surah consist of six ayah or verses.

The translation by M. T Hilal/ Khan from DivineIslam's Qur'an Viewer software version 2.9 is as below:

- i **Say:** I seek refuge with the lord and cherisher of mankind
- ii The king (or ruler) of mankind
- iii The God (or judge) of mankind
- iv From the mischief of the whisperer (of evil), who withdraws (after his whisper)
- v (The same) who whispers into the hearts of mankind
- vi Among jinns and among men

To achieve the objective, randomly selected readers were asked to read one of the three instruments such as in Fig. 1, Surah Al-Nas (114), ayah 1-6 tabulated with all the translation in text form (I_i), Fig. 2, Surah Al-Nas (114), ayah 1-6 tabulated with the translation in text form for the most frequently used words (I_{ii}) and Fig. 3, Surah Al-Nas (114), ayah 1-6 tabulated with the translation in visual form for the most frequently used words (I_{iii}). In Fig. 3,

1	قُلْ	أَعُوذُ	بِ	رَبِّ	النَّاسِ
	Say	I seek refuge	with	the Lord and Cherisher	Of Mankind
2	مَلِكِ	النَّاسِ			
	The King (or Ruler)	of Mankind,			
3	إِلَهِ	النَّاسِ			
	The god (or judge)	of Mankind			
4	مِنَ	شَرِّ	الْوَسْوَاسِ	الْخَفَّاسِ	
	From	the mischief	of the Whisperer (of Evil)	who withdraws (after his whisper)	
5	الَّذِي	يُوسِّسُ	صُدُورَ	النَّاسِ	
	(The same) who	whispers	into the hearts	of Mankind	
6	مِنَ	الْجِنَّةِ	و	النَّاسِ	
	Among	Jinns	and	men	

Fig. 1: Surah Al-Nas (114), ayah 1-6 tabulated with the translation in text form

1	قُلْ	أَعُوذُ	بِ	رَبِّ	النَّاسِ
					Of Mankind
2	مَلِكِ	النَّاسِ			
		of Mankind,			
3	إِلَهِ	النَّاسِ			
		of Mankind			
4	مِنَ	شَرِّ	الْوَسْوَاسِ	الْخَفَّاسِ	
	From		of the Whisperer (of Evil)		
5	الَّذِي	يُوسِّسُ	رِصْدُ	النَّاسِ	
		whispers		of Mankind	
6	مِنَ	الْجِنَّةِ	و	النَّاسِ	
	Among			men	

Fig 2: Surah Al-Nas (114), ayah 1-6 tabulated with the translation in text form (for the most frequently used words)

1	قُلْ	أَعُوذُ	بِ	رَبِّ	النَّاسِ
				رَبِّ	النَّاسِ
2	مَلِكِ	النَّاسِ			
		النَّاسِ			
3	إِلَهِ	النَّاسِ			
		النَّاسِ			
4	مِنَ	شَرِّ	الْوَسْوَاسِ	الْخَفَّاسِ	
	From				
5	الَّذِي	يُوسِّسُ	صُدُورَ	النَّاسِ	
				النَّاسِ	
6	مِنَ	الْجِنَّةِ	و	النَّاسِ	
	Among			النَّاسِ	

Fig 3: Surah Al-Nas (114), ayah 1-6 tabulated with the translation in visual form (for the most frequently used words plus the meaning of رَبِّ النَّاسِ)



Fig. 4: Surah Al-Nas (114), ayah 1-6 outlined with only the Arabic words (control instrument)

preposition words are included and they are left in the form of text since it is quite impossible to represent visual image of the words. Figure 4 is the control instrument (Ic), containing only the Arabic text of Surah Al-Nas (114), ayah 1-6.

There were 46 participants involved with Ii, 36 with Iii, 27 with Iiii and 46 with Ic. Participants were among those who can read the Quran but could not speak the Arabic language. Mixture of participants from Malaysia and Iran were asked to read the instruments Ii, Iii, Iiii or Ic after which they had to answer a few questions. However, they were not given the instruments to refer to while answering the question.

The questions are open ended. If the participants could not answer the questions, they were asked to leave it blank. The questions are:

- What are the activities described in the Surah? List them out
- Who are involved? List them out with some description (if possible)

The answers to the questions should be as follows:

The activities involve are:

- Say
- Seek refuge
- Whisper into the heart
- Withdraws after whisper

The characters involve are:

- Allah (the lord and cherisher, the king of mankind)
- Mankind
- The whisperer/the devil (whispers evil into the heart)
- Jinns

Expected scores: The expected scores of instruments Ii, Iii, Iiii were calculated. For questions number i) described

Table 1: The Probabilities and the Expected Total Score of each Instrument, I_i-I_{iii}

Instrument	Probabilities S (the score for each instrument)	Expected total score (Ets)
I _i	P(Si = 0), P(Si = 8) = 1/256 P(Si = 1), P(Si = 7) = 8/256 P(Si = 2), P(Si = 6) = 28/256 P(Si = 3), P(Si = 5) = 56/256 P(Si = 4) = 70/256	Ets (Ii) = 4
I _{ii}	P(Sii = 0), P(Sii = 3) = 1/8 P(Sii = 1), P(Sii = 2) = 3/8	Ets (Iii) = 1.5
I _{iii}	P(Siii = 0), P(Siii = 4) = 1/16 P(Siii = 1), P(Siii = 3) = 4/16 P(Siii = 2) = 6/16	Ets (Iiii) = 2

earlier, the full mark, Fi is 4 (1 for each answer item) and question ii), the full mark, Fii is also 4. Therefore for a perfect comprehension of the Surah, the total score, Ts should be 8 resulting in the formula below:

$$Ts = Fi + Fii = 8$$

Assuming the scores that will be obtained follows a normal distribution, the probabilities and the expected scores of each instrument can be estimated. For Ii, since all the translated word is displayed, the total expected scores that can be obtained is 11 so therefore, the probabilities of getting scores zero to eight is estimate as in Table 1. For Iii, since only frequently used words are used, the possible scores are zero to three but in Iiii the possible scores are zero to four hence the probabilities as in Table 1. The same method is used to find other probabilities. The probabilities can be obtained as follows:

$$P(Ts = X) = \frac{{}_8C_x}{{}_8C_0 + {}_8C_1 + \dots + {}_8C_7 + {}_8C_8} \text{ where } X = 0, 1, \dots, 8$$

The expected score can be obtained as follows:

$$ETs(I_x) = \sum_{j=0}^k JP(I_x = j)$$

Where:

j = Score

k = Maximum score

Using $Ets(I_i) = 4$, $Ets(I_{ii}) = 1.5$, $Ets(I_{iii}) = 2$ and the rest of the expected result is as listed in Table 1, the expected level of comprehension, Ec can be determined as $Ec = Ets/Ts$, hence:

- For I_i , $Ec = 0.5$
- For I_{ii} , $Ec = 0.19$
- For I_{iii} , $Ec = 0.25$

The closer the value of Ec to 1 indicates higher level of comprehension of the individual. For each instrument used, the Ec will be compared to the real level of comprehension, Rc found from the samples taken. Finally, a one-tailed t-test were applied to test whether there is any significant difference on the level of comprehension when images are used as part of the reading text.

RESULTS

Firstly, the set of results of Ec for each instruments I_i , I_{ii} and I_{iii} were normalized. The mean value of I_c (0.22) was

subtracted from the Ec found for all of the participants. Table 2 shows the one-sample statistics for instruments I_i , I_{ii} , I_{iii} and I_c showing the population, mean, standard deviation and standard Error Mean.

While Table 3 shows the results of one-sample T-test for instruments I_i (with test value = 0.5), I_{ii} (with test value = 0.19) and I_{iii} (with test value = 0.25). It shows that for I_i , there is evidence that no significant difference in the expected real comprehension value and the tabulated value (since $t(45) = -0.477$, $p > 0.05$). While for both I_{ii} and I_{iii} , there is evidence that there is significant difference in the expected real comprehension value (since $p < 0.05$) and the calculated value found is greater than the tabulated value (For I_{ii} , $t(35) = 6.105$, $p < 0.05$, for I_{iii} , $t(26) = 4.418$, $p < 0.05$).

Table 4 shows the group statistics of results of an Independent t-test for instrument I_{ii} and I_{iii} . In Table 5 it shows that (for the independent sample test between instrument I_{ii} and I_{iii}) there is evidence showing significant difference between the two groups ($t(61) = 4.423$, $p < 0.005$).

DISCUSSION

Yaxley and Zwaan (2007) found that readers mentally simulate the visibility of an object during language

Table 2: One-Sample Statistics for instruments I_i , I_{ii} , I_{iii} and I_c (population, mean, SD and SEM)

Sample category	N	Mean	SD	SEM
Real comprehension level of the control group I_c	39	0.2244	0.29827	0.04776
Normalized real comprehension level of I_i	46	0.4920	0.11447	0.01688
Normalized real comprehension level of I_{ii}	36	0.3668	0.17377	0.02896
Normalized real comprehension level of I_{iii}	27	0.3935	0.16879	0.03248

Table 3: Results of one-sample t-test for instruments

Sample category	t	df	Sig. (2-tailed)	Mean difference	95% Confidence interval of the difference	
					Lower	Upper
Normalized real comprehension level of I_i Test value = 0.5	-0.477	45	0.636	-0.00804	-0.0420	0.0260
Normalized real comprehension level of I_{ii} Test value = 0.19	6.105	35	0.000	0.17681	0.1180	0.2356
Normalized real comprehension level of I_{iii} Test value = 0.25	4.418	26	0.000	0.14352	0.0767	0.2103

I_i (with test value = 0.5), I_{ii} (with test value = 0.19) and I_{iii} (with test value = 0.25)

Table 4: Group Statistics for instrument I_{ii} and I_{iii}

	Type	N	Mean	SD	SEM
Real comprehension level	Samples from Instrument I_{iii}	36	0.5868	0.17377	0.02896
	Samples from Instrument I_{ii}	27	0.3935	0.16879	0.03248

Table 5: Independent samples test between instrument I_{ii} and I_{iii} , measuring the real comprehension level

	Levene's test for equality of variances	t-test for equality of means								
		F		Sig.		Mean difference		95% confidence interval of the difference		
		Lower	Upper	t	df	Lower	Upper	Lower	Upper	
Real comprehension level	Equal variances assumed	0.377	0.542	4.423	61	0.000	0.19329	0.04370	0.10590	0.28068

comprehension thus linguistic simulation of the object properties is one of the ways that could help the reader to comprehend. Similar work by others Zwaan *et al.* (2004), Gosselin and Schyns (2004), Richardson *et al.* (2003) also support this evidence. A related work such as by Harber and Myers (1982) found that there is greater accuracy in remembering pictograms compared to words. These theories suggest that images may be used as a basis to solve the problem encountered by readers of the Arabic document (the Quran, especially those who can read in Arabic but could not comprehend the meaning).

The expected values 0.5, 0.19 and 0.25 are values that were expected from the participants by reading Ii, Iii and Iiii. For the case of Ii, there is evidence that using Arabic text and translation leads to no significant difference from the expected results.

For the cases of Iii and Iiii, although the experiments mentioned earlier by Yaxley and Zwaan (2007), Zwaan *et al.* (2004), Gosselin and Schyns (2004), Richardson *et al.* (2003) and Harber and Myers (1982) had different approaches and was used in different context, the results found in this experiment is consistent with the others. It was found that there is significant difference on the level of comprehension when images are used as part of the reading text. Therefore in this experiment images used resulted in a higher comprehension level of the text read.

The result found can be use as a basis for the development of a visualization system to assist non-Arabic speaker who can read the Arabic document (the Qur'an) to comprehend the text.

REFERENCES

- Abbasi, A. and H. Chen, 2007. Categorization and analysis of text in computer mediated communication archives using visualization. Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries, Jun. 18-23, Vancouver, BC, Canada, ACM., New York, pp: 11-18.
- Fang, S., M. Lwin and P. Ebright, 2006. Visualization of unstructured text sequences of nursing narratives. Proceedings of the 2006 ACM Symposium on Applied Computing, Apr. 23-27, ACM New York, USA., pp: 240-244.
- Fortuna, B., M. Grobelnik and D. Mladenìè, 2005. Visualization of Text Document Corpus. *Informatica*, 29: 497-502.
- Frost, R., 2007. Orthographic Systems and Skilled Word Recognition Processes in Reading. In: *The Science of Reading: A Handbook*, Snowling, M.J. and C. Hulme (Eds.). Blackwell Publishing, Oxford, UK., pp: 272-295.
- Gosselin, F. and P.G. Schyns, 2004. A picture is worth thousands of trials: Rendering the use of visual information from spiking neurons to recognition. *Cognitive Sci.*, 28: 141-146.
- Haber, R.N. and B.L. Myers, 1982. Memory for pictograms, pictures and words separately and all mixed up. *Perception*, 11: 57-64.
- Lupker, S.J., 2007. Visual Word Recognition: Theories and Findings. In: *The Science of Reading: A Handbook*, Snowling, M.J. and C. Hulme (Eds.). Blackwell Publishing, Oxford, UK, ISBN: 978-1-4051-6811-3, pp: 39-60.
- Perfetti, C.A., N. Landi and J. Oakhill, 2007. Modeling Reading: The Acquisition of Reading Comprehension Skill. In: *The Science of Reading: A Handbook*, Snowling, M.J. and C. Hulme (Eds.). Blackwell Publishing, Oxford, UK., ISBN: 978-1-4051-6811-3.
- Richardson, D.C., M.J. Spivey, L.W. Barsalou and K. McRae, 2003. Spatial representations activated during real-time comprehension of verbs. *Cognitive Sci.*, 27: 767-780.
- Weber, W., 2007. Text Visualization - What colors tell about a text. Proceedings of the 11th international Conference Information Visualization, Jul. 04-06, IEEE Computer Society, Washington, DC., pp: 354-362.
- Weippl, E., 2001. Visualizing content based relations in texts. Proceedings of the 2nd Australasian Conference on User Interface, Feb. 1-29, IEEE Computer Society, pp: 34-41.
- Wise, J.A., J.J. Thomas, K. Penneck and D. Lantrip *et al.*, 1995. Visualizing the non-visual: Spatial analysis and interaction with information from text documents. Proceedings of the 1995 IEEE Symposium on Information Visualization, Oct. 30-31, IEEE Computer Society Press, pp: 51-58.
- Yaxley, R.H. and R.A. Zwaan, 2007. Simulating visibility during language comprehension. *Cognition*, 105: 229-236.
- Yeap, W.K., P. Reedy, K. Min and H. Ho, 2005. Visualizing the meaning of texts. Proceedings of the 9th international Conference on Information Visualisation, Jul. 06-08, IEEE Computer Society, Washington, DC., pp: 883-888.
- Zwaan, R.A., C.J. Madden, R.H. Yaxley and M.E. Aveyard, 2004. Moving words: Dynamic representations in language comprehension. *Cognitive Sci.*, 28: 611-619.