

Anaphora Resolution In Reading Among Malaysian L2 English Speakers: An Eye-Tracking Investigation

Resolusi Anafora Dalam Pembacaan Dalam Kalangan Penutur Bahasa Inggeris L2 Malaysia: Penyiasatan Penjejakan Mata

Shamita Chantherasarathy Naido
¹Nurjanah Mohd Jaafar

Pusat Kajian Bahasa dan Linguistik
Fakulti Sains Sosial dan Kemanusiaan
Universiti Kebangsaan Malaysia

Correspondence e-mel: ¹nurjanah@ukm.edu.my

ABSTRACT

In this study, eye tracking was utilized to investigate online reading behaviour and comprehension among a less studied L2 speakers of English. The main purpose of this study was to investigate anaphora resolution in reading among Malaysian L2 English speakers, focusing on three objectives: (1) to identify L2 English speakers' gaze behaviour when reading sentences with anaphors; (2) to identify the relationship between processing and comprehension of anaphors; and (3) to investigate the influence of word-frequency on the processing of anaphor. This study employed the quantitative method to measure the statistical relationship between the variables. This study involved 14 participants who were instructed to read the texts silently and their eye movements were recorded using the Tobii X300 non-intrusive eye-tracker with 300 Hz sample rate. The results indicated that generally the gaze behaviour of Malaysian L2 English speakers when reading anaphora texts corresponded to past studies using L1 and L2 English speakers. Next, there were no apparent correlation observed between the processing of anaphor sentences with comprehension performances of the participants. Finally, the word-frequency effect on the whole is insignificant. Overall, advanced Malaysian L2 English speakers do not appear to show significant difficulties when reading anaphors. This study has provided some noteworthy findings on the reading domain among L2 English speakers focusing on the processing and comprehending during anaphora resolution since scarcely any research has been conducted locally on this topic. This study added on and substantiated existing results from previous studies on L2 acquisition while filling the gap in the aspects of the roles of word-frequency during anaphora resolution. Moreover, this study has incorporated eye-tracking technology to validate findings in measuring reading fluency and comprehension which no local studies have included when investigating anaphora resolution.

Keywords: *anaphora; anaphora resolution; eye-tracking; second language; comprehension*

ABSTRAK

Dalam kajian ini, alat penjejak mata digunakan untuk menyiasat tingkah laku dan pemahaman para penutur bahasa Inggeris. Kajian ini bertujuan untuk menyiasat resolusi anafora dalam bahasa Inggeris, dengan memberi tumpuan kepada tiga objektif: (1) untuk mengenal pasti tingkah laku visual ketika membaca ayat dengan anafora; (2) untuk mengenal pasti hubungan antara pemprosesan dan pemahaman anafora; dan (3) untuk mengkaji pengaruh frekuensi kata pada pemprosesan anafora. Kajian ini menggunakan kaedah kuantitatif untuk mengukur hubungan statistik antara pemboleh ubah. Ia melibatkan 14 peserta yang diminta untuk membaca teks secara senyap sambil pergerakan mata mereka direkodkan menggunakan alat pengesan mata Tobii X300. Hasil kajian menunjukkan bahawa secara amnya, tingkah laku mereka ketika membaca teks anafora bertepatan dengan kajian-kajian terdahulu. Seterusnya, tidak ada hubungan yang jelas antara pengolahan perkataan anafora dengan pencapaian pemahaman peserta. Akhir sekali, kesan kekerapan perkataan pada keseluruhannya tidak signifikan. Secara keseluruhan, penutur bahasa Inggeris yang cekap tidak menunjukkan kesukaran yang ketara ketika membaca anafora.

Kata kunci: *anafora; resolusi anafora; pengesanan mata; Bahasa kedua; kefahaman*

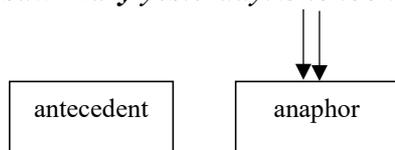
1. Introduction

In this digital era, good English proficiency is a necessity. Proficiency in reading, especially, is essential in enhancing language ability. Readers typically use vocabulary, reading strategies, grammatical and linguistic knowledge along with their background knowledge to understand and achieve their reading purposes (Peregoy & Boyle, 2001). Reading in a second language is a complex cognitive process. It is not surprising, therefore, to encounter worrisome statistics that suggest below average reading skills of Malaysian youths. In MUET (Malaysian University English Language Test), for instance, the total number of candidates who achieved Band 3 was the highest with 38.3% followed by Band 2 (26.2%) (Malaysian Examinations Council, 2019). The MUET scoring is separated into six bands in terms of the level of English proficiency. Bands 6 and 5 are the highest, indicating proficient users of the language. Next, Bands 4 and 3 indicate that candidates are independent users of the language and followed by Bands 2 and 1 for basic users. The present study intends to explore reasons underlying such unsatisfactory English reading skills among Malaysian ESL speakers by focusing on anaphora resolution (AR) skill. This skill refers to the ability to associate between current and previous information (Levine, Guzmán & Klin, 2000). More specifically, AR refers to the process of relating antecedents to its anaphors (McDonald & MacWhinney, 1995). To examine AR skill, the researcher employed the eye-tracking methodology as it would allow the identification of reading patterns that underly poor reading proficiency (Swan & Walter, 2017; Swerling, 2015).

2. Anaphora Resolution in a Second Language

Anaphora Resolution (AR) entails a reader relating antecedents to its anaphors (McDonald & MacWhinney, 1995). An antecedent is usually a noun phrase mentioned earlier in a sentence or discourse which provides the interpretation for a second expression, commonly pronouns - the anaphor. Two examples are provided below.

(1) *John saw **Mary** yesterday. **She** looked tired that day.*



(2) *Yesterday **the minister** left London after reminding **himself** about the letter.*



Seddik and Farghaly (2014) expressed that in natural language processing, AR is a crucial aspect of reading comprehension because a text would not be understood correctly if the roles and meaning of the antecedents and anaphors are not correctly assigned. It is one of the major language facets that should be investigated given the complexity and its interconnection to reading comprehension skills. The complex and mentally demanding nature of these processes (Joseph, Bremner, Liversedge & Nation, 2015; Nobre, 2011) - (1) identifying the anaphor; (2) identifying the potential antecedent; (3) analysing the semantic link; (4) and then determining the anaphor while sustaining a precise and orderly mental model of the ongoing discourse – suggest that it involves skills that require time to develop.

A large number of English speakers - both native and non-native - face processing difficulties despite their advanced proficiency in English (e.g. Contemori, Asiri & Irigoyen, 2019; Joseph, Bremner, Liversedge & Nation, 2015). The study by Çokal, Sturt and Ferreira (2018) revealed that advanced Turkish ESL participants demonstrated online processing limitations as it requires higher proficiency indicating inadequate naturalistic input. The total reading times of anaphors observed for L2 speakers are significantly longer than L1 speakers with similar results on second-pass reading times on anaphors. The study by Contemori, Asiri and Irigoyen (2019) also revealed similar results but this study used an offline method (pen and paper task). The statistical analysis showed that L2 speakers have a lower mean score when the sentence contained an ambiguous pronoun with two antecedents. The authors concludes that L2 speakers face processing difficulties when the complexity of the discourse increases. Past research also indicated that there is a lack of evidence to establish a concrete conclusion on AR in the context of L2 acquisition (e.g., Cunnings, Fotiadou & Tsimpli, 2017; McDonald, 2006). Much research has focused on L1 English speakers, hence, research that provides insight to the nature of AR among L2 English speakers is much needed.

The role of word-frequency during AR has also been extensively examined. Nevertheless, there is still a disagreement in deducing the role of word frequency during AR with previous studies showing contrasting results. Some studies (e.g., Van Gompel & Majid, 2004) observed that pronouns with low-frequency antecedents were read faster but in other studies (e.g., Lago, 2014), high-frequency antecedents were read faster. The present research attempts to fill the gaps in literature by examining the nature of AR among Malaysian L2 English speakers. Moreover, the research attempts to resolve the disagreement with regard to the role of word frequency on AR by examining its effects during the processing and comprehension of anaphor. Furthermore, this research, incorporated the eye-tracking data to allow the researcher to demonstrate the underlying cognitive processes involved during AR. In recent years, there is a surge of interests among researchers to incorporate the eye-tracking technology to investigate reading processes including anaphora resolution (e.g., Whitford & Titone 2012; Winke, Gass, & Sydorenko 2013). Eye-tracking allows the researchers to examine readers' eye gaze and movements while reading, thereby providing an indication of cognitive processes underlying such a complex cognitive task.

To guide the study, three research questions were formulated:

1. How do L2 English speakers read sentences with anaphors?
2. What is the relationship between the processing and comprehension of anaphors?
3. What is the influence of word frequency on the processing of anaphors?

3. Methods

3.1 Participants

There were three criteria in selecting the participants for this study; (1) L2 English speakers; (2) undergraduate students; and (3) having normal to corrected vision. 14 UKM undergraduates who were all L2 English speakers participated in the study. Their ages ranged from 21 to 25 ($M = 23$; $SD = 1.17$); and their L1 were either Malay ($n = 7$), Mandarin ($n = 1$) or Tamil ($n = 6$). The participants were from different faculties and were in their first, second or third year of their respective courses - English Language Studies ($n = 7$), Law ($n = 3$) and one each from Civil Engineering, Electrical and Electronics Engineering, Computer Science, and Mass and Digital Communication. Next, participants' MUET exam scores were inclusive of Band 3 to Band 5, indicative of independent to proficient users of English (Malaysian Examinations Council, 2019). On average, participants rated their English speaking, comprehension, reading, and writing skills on a scale of 1 to 6 (with 1 being very low and 6 being native-like), ranging from intermediate to native-like. All participants had normal to corrected vision with none having any diagnosis related to reading problems. The sample size (i.e. 14) was small due to pandemic restrictions imposed by the government and the university. Movement restrictions had limited data collection. Several participants approached were under self-quarantine when they returned to campus; and others were not allowed to leave their respective hostel areas when cases increased within the campus. Subsequently, the university had also denied access to the eye-tracking laboratory which called for discontinuation of data collection. Consequently, the research had to proceed with data analysis.

3.2 Materials

The stimuli comprised of 2 sets of short texts in English which are adapted from journal articles. Each set consists of 20 experimental texts of 20 different items. There are two conditions for each item: (1) Pronoun, High-frequency antecedent; and (2) Pronoun, Low-frequency antecedent. However, each participant only read one version of the same item. The average length of the anaphora is 5 to 7 characters long as previous studies found significant word skipping behaviour when the target words consist of 8 letters or more (Rayner & McConkie, 1976). Meanwhile, the average word count for each short text is 15 to 40 words as shown in the samples from Egusquiza, Navarrete and Zawiszewski (2014), Garrod, Freudenthal and Boyle (1994) and Kreiner, and Sturt and Garrod (2008). Next, each short text is followed by a comprehension question (true/false) on the same slide as the respective texts. Thus, there is a total of 40 short texts and 40 comprehension questions for the experimental items. Besides, 20 filler texts containing non-experimental items were randomly included for both sets. Hence, both sets consisted of the same filler texts. Below is the sample of sentences from short texts that was used. Each participant only read either one of the two sentences below.

High frequency antecedent	Low frequency antecedent
1. [A minister] criticized the queen during the speech of yesterday. [He] disapproved the monarchy.	2. [A senator] criticized the queen during the speech of yesterday. [He] disapproved the monarchy.

3.3 Procedure

Participants will be tested individually in the eye-tracking laboratory. Participants will sit in front of a 19-inch screen and their eye movements will be recorded using a Tobii X300 non-intrusive eye-tracker with 300 Hz sample rate. Firstly, the participants will undergo a standard 9-point calibration process before proceeding to the real experiment. This is to ensure that the eye movements are being recorded accurately. After that, they will be presented with the instructions on the screen and proceeded with the experimental and filler sentences. The experiment will begin with a filler text to familiarize participants with the experimental procedure. The presented texts are in Times New Roman 18 font, randomly arranged, with no two experimental items adjacent. Participants will be informed to read carefully at his or her normal rate. The participants are given a mouse to click on to continue with the next texts at their own pace. Optional short breaks will be given to prevent fatigue. At the end of each shorttext, comprehension questions (true/false) will be given. The comprehension questions will be placed at the bottom the same page after the text. The participants will orally provide their answers to the researcher.

3.4 Apparatus and eye-tracking metrics

Participants' eye movements will be recorded using a Tobii X300 non-intrusive eye-tracker with 300 Hz sample rate. The apparatus is suitable for its features: (i) has the flexibility in allowing slight head movements; (ii) the ability to carry out experiments for a long period of time without causing tiredness of the eyes; (iii) and high sensitivity in tracking and detection of word-to-word eye movements.

Three eye-tracking metrics were used to examine the participants' anaphora resolution:

- (1) first fixation duration (FFD) – “the duration of the first fixation on the word independent of the number of fixations on the word” (Rayner & Well, 1986, p. 505);
- (2) total fixation duration (TFD) which is defined as “the sum of all of the fixations made on a target word prior to any movement away from the target word” (Rayner & Duffy, 1986, p.505); and
- (3) fixation count (FC) which measures the total number of fixations on a target word (Rayner & Well, 1986).

3.5 Data analysis

The raw data obtained from the in-built software were transferred to *Statistical Package for the Social Sciences* (SPSS) for data analysis. For RQ1, descriptive statistics was used, specifically, mean (M) and standard deviation (SD) to measure FFD, TFD and FC. Fixation duration below 100ms or above 2500ms were excluded from the analyses. Analyses included two critical regions the anaphor and the antecedent regions (Kreiner, Sturt & Garrod, 2008; Egusquiza, Navarrete & Zawiszewski, 2014; Cokal, Sturt & Ferreira, 2016). An example of AOI definition is illustrated in Figure 1.

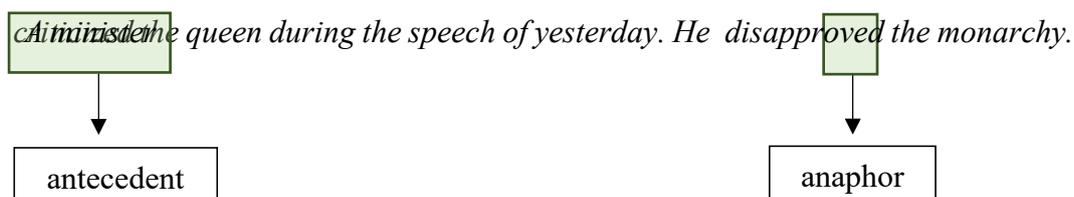


FIGURE 1. AOI definition

4. Findings and Discussion

4.1 How do L2 English speakers read sentences with anaphors?

Table 1 summarizes the descriptive data of the 40 experimental sentences from the two sets.

TABLE 1. Means and standard deviations of eye tracking metrics according to frequency and area of interest (AOI)

Eye-tracking metrics	Area of interest (AOI)	Word frequency	<i>n</i>	Mean (<i>M</i>)	Std. Deviation (<i>SD</i>)
First Fixation Duration (FFD)	Antecedent	High	20	0.18	0.04
		Low	20	0.2	0.04
	Anaphor	High	20	0.11	0.06
		Low	20	0.12	0.06
Total Fixation Duration (TFD)	Antecedent	High	20	0.51	0.13
		Low	20	0.57	0.2
	Anaphor	High	20	0.21	0.18
		Low	20	0.23	0.16
Fixation Count (FC)	Antecedent	High	20	2.49	0.6
		Low	20	2.72	0.89
	Anaphor	High	20	0.98	0.81
		Low	20	1.02	0.6

The data revealed that participants made reasonably longer fixations when the AOI is a low-frequency antecedent for both FFD and TFD. On average, participants looked at low-frequency antecedents for 0.20s ($SD = 0.04$) and 0.57s ($SD = 0.2$) for FFD and TFD, respectively. Furthermore, FC demonstrated that the number of times the participants had fixated on antecedents is greater for low-frequency ($M = 2.73$, $SD = 0.89$) antecedents compared to high-frequency antecedents ($M = 2.49$, $SD = 0.6$). Next, all three metrics showed that participants fixated relatively longer and more frequently for anaphors of low-frequency antecedents than anaphors of high-frequency antecedents. The data reported that the mean for low-frequency anaphors for FFD and TFD are 0.12s ($SD = 0.05$) for the former and 0.23s ($SD = 0.156$) for the latter. The mean FC is 1.02 times ($SD = 0.91$) for low-frequency anaphors. In contrast, the means for high-frequency anaphors with regard to the three metrics are 0.11s (FFD), 0.21s (TFD) and 0.98 times (FC).

As hypothesised, participants in the present study demonstrated longer FFD and TFD with greater FC for anaphora sentences, particularly for low-frequency antecedents and its anaphors. This is parallel to the findings of previous studies, such as Hyönä and Olson (1995), Rayner and Duffy (1986) and Schilling, Rayner and Chumbley (1998), all of which explained that high-frequency antecedents are processed faster as the retrieval of these antecedents is more sensitive to lexical frequency as suggested by the full-access hypothesis. In general, the results of this present study suggest that gaze behaviour of Malaysian L2 English speakers in this study replicated those of native and non-native speakers of English from the aforementioned studies. The findings also support the notion made by researchers (e.g., Almor, 1999; Carpenter & Just, 1977; Klin, Weingartner, Guzman, & Levine, 2004, who asserted that texts involving anaphora resolution commonly

produce longer reading times and frequent rereading as these texts require higher level encoding and merging of words during sentence processing.

4.2 What is the relationship between the processing and comprehension of anaphors?

Table 2 shows the results from the Pearson correlation coefficient test to investigate the relationship between the processing and comprehension of anaphors.

TABLE 2. Correlations between eye tracking metrics and comprehension score

Eye tracking metrics	Word frequency	Reading comprehension ability			
		Antecedent		Anaphor	
		<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
FFD	High	-0.132	0.580	0.139	0.560
	Low	-0.226	0.338	-0.201	0.396
TFD	High	0.169	0.477	0.017	0.945
	Low	-0.354	0.125	-0.308	0.187
FC	High	0.109	0.648	0.083	0.727
	Low	-0.413	0.070 ⁺	-0.282	0.228

Based on the results, it can be stated that the processing of high- and low-frequency antecedents and anaphors do not explain the participants' comprehension performance. In general, there is no significant correlation between the processing of high- and low-frequency antecedents and anaphors with comprehension of the sentences. To illustrate, the results demonstrated that both conditions are either positively correlated with comprehension scores, nevertheless insignificant with $p > 0.05$, or negatively correlated. Overall, only 5 conditions show positive correlations with comprehension scores while others are negatively correlated, all of which are non-significant. However, one clear observation that can be seen is that low-frequency antecedents and its anaphors are all negatively correlated with comprehension scores. In terms of antecedent, the correlation values (*r*) for FFD, TFD and FC are -0.23, -0.35 and -0.41, respectively. Meanwhile, *r* values for FFD, TFD and FC with regard to anaphors are -0.2, -0.31, and -0.28. On the other hand, FFD is the only metric which showed a negative correlation for high-frequency antecedents with the *r* value of -0.13. Other metrics for high-frequency antecedents (*r* = 0.17 for TFD; *r* = 0.11 for FC) and its anaphors (*r* = 0.14 for FFD; *r* = 0.02 for TFD; *r* = 0.08 for FC) are positively correlated with comprehension scores. Nonetheless, the results showed the FC for low-frequency antecedents as marginally significant as $p < 0.1$. Hence, FC is fairly associated with comprehension performance when processing low-frequency antecedents. Generally, word-frequency effect is not linked to comprehension performance at large since the data summarizes that the participants' comprehension scores do not correlate significantly with processing of anaphors.

Generally, these findings conflicted with those of past studies (e.g., Contemori, Asiri, & Perea Irigoyen 2019; Joseph, Bremner, Liversedge, & Nation 2015) which found L2 speakers having difficulty with anaphora resolution. The present study contradicts Roberts, Gullberg and Indefrey (2008) despite also having ambiguous pronoun and two entities in the experimental texts because they found L2 speakers being incorrect significantly more often than L1 speakers. Nonetheless, this study is aligned with previous studies, for

example, Masrai (2019) and Mulder, van de Ven, Segers, Krepel, de Bree, de Jong and Verhoeven (2021). Masrai (2019) stated that readers are able to comprehend a text adequately, if not accurately, if 98% of the words belong to the high- and mid-frequency groups. This may explain the findings of the present study as the level of difficulty is 72.3 according to the Flesch Reading Ease (FRE) readability score (Jindal & MacDermid 2017). Therefore, the experimental texts on average are fairly easy to read since FRE score range of 71-80 is estimated to be appropriate for Grade 7 in the United States which equates to Form 1 students in Malaysia. The readability score in comparison with the MUET scores of participants also suggest that anaphora resolution was not obvious as most participants are either Band 5 or Band 4. This suggests that further research use more challenging materials, suitable for proficient or independent users of L2 English.

Although the participants in the present study typically processed high-frequency antecedents and its anaphors faster (shorter FFD and TFD with lesser fixation count), however, their comprehension scores showed the opposite – i.e., the participants scored better on comprehension questions with low-frequency antecedents and anaphors compared to sentences with high-frequency antecedents and anaphors. These are aligned with previous findings (e.g., Çokal, Sturt & Ferreira 2016; Rayner 1995) with L1 English speakers which deduced that for majority of the participants, their comprehension performances are neither predicted by the length of fixations nor frequency of fixations. Based on discourse salience (Grosz, Joshi & Weinstein 1983), word frequency effects influence the saliency of words, in this context, low-frequency words will be more salient than high-frequency words. Hence, this explains the findings observed in the present study according to discourse salience which claims that low-frequency words are easier to access as they are encoded more strongly in the memory due to greater processing load (Van Gompel & Majid 2004). Meanwhile, FC for low-frequency antecedents indicate that it is marginally significant. Thus, the findings in this study do not resemble the claims that L2 English speakers tend to have a processing disadvantage compared to L1 speakers (Çokal, Sturt and Ferreira 2018) which denoted that anaphora resolution is a rather complicated skill as a large number of participants - both L1 and L2 English speakers - faced processing difficulties despite their advanced proficiency in English (e.g., Contemori, Asiri & Irigoyen 2019; Joseph, Bremner, Liversedge & Nation 2015).

One other conclusion that can be made from the findings is that Malaysian L2 English speakers who are advanced users of English (Band 4 and Band 5 based on MUET) do not appear to encounter much difficulties when reading anaphors. To illustrate, six out of 14 participants achieved a score of 80 and above in the comprehension test, all of whom were advanced users of the English language. On the other hand, the participant with Band 3 only scored 7.1%. This suggests that differences in language proficiency contribute to variation in comprehension performance. Stanovich (1986) and Just and Carpenter (1980) argued that higher proficiency speakers are able to comprehend anaphors better than those with lower proficiency because of the fact that they tend to have greater familiarity with less common words, perhaps because of their increased exposure to text. In addition, they observed that poor readers are more likely to be less successful in reading texts with anaphora resolution. Thus, they were not able to understand the whole text accurately. Besides, Pretorius (2005) investigated the relationship between skills in anaphoric resolution, academic performance and language proficiency. The findings show that ability to resolve anaphoric expression predicts academic performance. To illustrate, students with poor academic performance had great difficulties with anaphora resolution. The rationale proposed by researchers (Chateau & Jared 2000; Tainturier, Tremblay & Lecourse 1992;) is that as readers receive higher level of education, they are exposed to greater vocabulary. Thus, they tend to be less sensitive to lexical factors in reading,

including word-frequency (Butler & Hains 1979).

4.3 What is the influence of word-frequency effect on the processing of anaphors?

A Wilcoxon signed rank test was carried out to investigate whether word-frequency effects influenced the processing of anaphors under two different conditions – high- and low-frequency antecedents. A non-parametric test is used due to the small sample size and the non-normal distribution of data. The results are presented in Table 3.

TABLE 3. Test statistics (z , p , r) of FFD, TFD and FC based on Wilcoxon signed ranks test

	Antecedent			Anaphor		
	FFD	TFD	FC	FFD	TFD	FC
Z	-1.66	-1.78	-0.63	-0.89	-0.71	-2.51
Significance (p)	0.096	0.07	0.53	0.37	0.48	0.01
Effect size (r)	-0.37	-0.4	-0.14	-0.2	-0.16	-0.56

The results indicate that none of the 40 texts generated significant influence ($p > 0.05$) on the differences in reading durations (FFD and TFD) between high- and low- frequency antecedents and its anaphors. Thus, low- and high- word frequency have fairly similar effects on FFD and TFD. Nevertheless, p values for antecedents in terms of FFD ($z = -1.66$, $p = 0.096$, $r = -0.37$) and TFD ($z = -1.78$, $p = 0.07$, $r = -0.4$) can be deemed as marginally significant. This indicates that low-frequency antecedents have a larger influence than high-frequency antecedents. Evenso, the results demonstrated that only the differences in the number of fixations (or FC) on anaphors made by participants was significant, $z = -2.51$, $p = 0.01$, $r = -0.56$. This suggests that influence of word-frequency effect was evident, thereby hinting that word-frequency may explain the differences in processing of anaphors with respect to FC. Furthermore, data on effect size (r) – based on Cohen’s conventions – reported a medium effect of word-frequency effect on the processing of anaphors with respect to FC. Meanwhile, r values for FFD (-0.37) and TFD (-0.4) with regard to antecedents revealed that the differences are reasonably visible with $0.2 < r < 0.5$, indicating a medium effect size. Hence, word-frequency effect had a moderate influence on the differences in FC when processing anaphors, and somewhat moderate effects on FFD and TFD for antecedents among the participants.

The word-frequency effect observed is, by and large, insignificant, since the AOIs of the experimental items did not generate significant influence on the differences in reading durations (FFD and TFD) between high- and low-frequency antecedents and its anaphors. However, the p values for antecedents in relation to FFD and TFD approached significance ($0.05 > p > 0.10$). Besides, the r values for FFD (-0.37) and TFD (-0.4) with regard to antecedents also revealed reasonably visible differences with $0.2 < r < 0.5$, indicating a medium effect size. This indicates that low-frequency antecedents influenced the processing of anaphors more than high-frequency antecedents. Accordingly, an earlier hypothesis that participants will make longer FFD and TFD for low-frequency antecedents and its anaphors was supported. Online reading times on the AOIs are faster for high -frequency words antecedents and its anaphors during anaphora resolution is consistent with previous research with L1 speakers focusing on silent reading (e.g., Lago 2014; Rayner & Duffy 1986; Schilling, Rayner & Chumbley 1998; Whitford & Titone 2017).

Next, this study also observed slightly longer reading times, FFD and TFD for

anaphors of low-frequency antecedents than anaphors of high-frequency antecedents which is in line with the results from the study by Van Gompel and Majid (2004). Nevertheless, the *p* values indicate that it is insignificant despite the presence of differences in processing between anaphors of high- and low-frequency antecedents. Therefore, there is an influence of word-frequency on the processing of antecedents but inconsequential which could be because of the sample participants who were predominantly proficient users of English. Lago (2014) explained that word-frequency effect is less influential as proficiency increases – i.e., poor readers are greatly affected by word-frequency effect followed by average readers. They claimed that good readers on average are not affected due to higher exposure to the language. Moreover, this study corroborates findings on the differences in the number of fixations made on high- and low- frequency anaphors (Inhoff 1984; Schilling, Rayner & Chumbley 1998). Inhoff (1984) and Kliegl, Olson and Davison (1982) justified that HF words commonly receive only one fixation, whereas low-frequency words are more likely to receive more than one fixation. They rationalised that FFD for high-frequency antecedents encompass lexical access and word interpretation which mirror complex processing operations. On the other hand, FFD on low- frequency antecedents predominantly mirror lexical access with additional fixations indicating word interpretation. Thus, more complex cognitive processes exhibit at a later stage for low- frequency words.

The differences in the number of fixations on anaphors (FC) made by participants was significant with a medium effect size although the online reading times were clearly insignificant. Simner and Smyth (1999) also did not report a word-frequency effect on the anaphors in which the participants' reading times were unaffected by the frequency of the antecedent noun. The findings of this study also replicated the study by Egusquiza, Navarrete and Zawiszewski (2014) to examine word-frequency effects on pronoun anaphoric comprehension in which they observed no significant frequency effect in the anaphor region when it is a pronoun.

One possible reason explaining low effect of word-frequency on the pronouns is because processing of pronouns is claimed to be slightly delayed because they require access of a discourse representation, and discourse effects may occur later than lower-level lexical effects. The frequency effect for pronouns did not occur at the pronoun itself. This is consistent with Rayner et al. (1995), who asserted that frequency effects persisted on the second occurrence of a word. In fact, evidence from various eye-tracking experiments have also substantiated that the earliest effect of word-frequency on anaphors ensued after the pronoun region (e.g., Ehrlich & Rayner 1983; Garrod & Sanford 1994). Van Gompel and Majid (2004) argued that the word- frequency effect is overt after pronouns due to reaccess of the antecedent. This is in accordance with the presumption of discourse salience which claims that the frequency of an antecedent influences the saliency of the antecedent, causing easier referring of the pronoun to its antecedent. In addition, Simner and Smyth (1999) concluded that less influence of word- frequency transference on the anaphors is because readers accessing the lemma representation of the antecedent rather than the lexeme representation when processing the pronouns (anaphors). This corresponds to past studies incorporating eye movements behaviour (e.g., Garrod, Freudenthal & Boyle 1994) which found notable word-frequency on the regions after the pronouns. They revealed that the participants usually read pronouns swiftly with frequent skipping. Consequently, further research on anaphora resolution should include the post- anaphor regions to verify the effects of word-frequency for antecedents on its anaphors.

5. Conclusion and Recommendations

In this study, Malaysian L2 English speakers' anaphora resolution in English was investigated. This study analysed eye movement behaviour when reading texts – adapted from journal articles – in a second language. Firstly, the results denoted that generally the gaze behaviour of Malaysian L2 English speakers when reading anaphora texts corresponded to past studies using L1 and L2 English speakers. On average, the means demonstrated that the participants made longer first fixation duration and total fixation duration with greater fixations when encountering anaphora sentences especially sentences consisting of low-frequency antecedents and anaphors. Secondly, there was no apparent correlation observed between the processing of anaphor sentences with comprehension performances of the participants. The results indicate lack of association between the processing and comprehension of anaphors. Next, the word- frequency effect on the whole is insignificant – low influence on the processing of anaphors. Overall, advanced Malaysian L2 English speakers (MUET Band 4 and Band 5 based) in the study did not appear to show significant difficulties when reading anaphors as suggested by previous studies. Nevertheless, this study comprised of a relatively homogeneous group with regard to English proficiency levels. Most of the participants are high proficient speakers who are able to comprehend anaphors more accurately than lower proficient speakers because of their wide exposure to various ranges of texts. Additionally, this study only comprised of 14 participants within the same population sample. Therefore, implementing diversity in a larger study with a wide range of English proficiency levels and educational levels will offer more generalizable findings.

References

- Almor, A. (1999). Noun-phrase anaphora and focus: The informational load hypothesis. *Psychological review* 106(4): 748.
- Carpenter, P. A., & Just, M. A. (1977). Reading comprehension as eyes see it. *Cognitive processes in comprehension* 109-139.
- Chateau, D., & Jared, D. (2000). Exposure to print and word recognition processes. *Memory & Cognition* 28(1): 143-153.
- Çokal, D., Sturt, P. & Ferreira, F. (2018). L2 Speakers' Reference Resolution in Processing and Production. *Cognitive Science*.
- Contemori, C., Asiri, O. & Perea Irigoyen, E. D. (2019). Anaphora Resolution In L2 English: an analysis of discourse complexity and cross-linguistic interference. *Studies in Second Language Acquisition* 41(5): 971-998.
- Cunnings, I., Fotiadou, G. & Tsimpli, I. (2017). Anaphora resolution and reanalysis during L2 sentence processing: Evidence from the visual world paradigm. *Studies in Second Language Acquisition* 39(4): 621-652. 35
- Egusquiza, N., Navarrete, E. & Zawiszewski, A. (2016). Antecedent frequency effects on anaphoric pronoun resolution: Evidence from Spanish. *Journal of psycholinguistic research* 45(1): 71-84.
- Garrod, S., Freudenthal, D. & Boyle, E. (1994). The role of different types of anaphor in the on-line resolution of sentences in a discourse. *Journal of memory and language* 33(1): 39- 68.
- Garrod, S. C. & Sanford, A. J. (1994). *Resolving sentences in a discourse context: How discourse representation affects language understanding*. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (p. 675–698). Academic Press.
- Grosz, B. J., Joshi A. K. & Weinstein S. (1983). *Providing a unified account of definite*

- noun phrases in discourse*. In Proc. 21st Annual Meeting of the ACL, pp. 44-50.
- Hyönä, J., & Olson, R. K. (1995). Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 21(6): 1430.
- Inhoff, A. W. (1984). Two stages of word processing during eye fixations in the reading of prose. *Journal of verbal learning and verbal behavior* 23(5): 612-624.
- Jindal, P., & MacDermid, J. C. (2017). Assessing reading levels of health information: uses and limitations of flesch formula. *Education for Health* 30(1): 84. 36
- Joseph, H. S., Bremner, G., Liversedge, S. P. & Nation, K. (2015). Working memory, reading ability and the effects of distance and typicality on anaphor resolution in children. *Journal of Cognitive Psychology* 27(5): 622-639.
- Klin, C. M., Weingartner, K. M., Guzmán, A. E., & Levine, W. H. (2004). Readers' sensitivity to linguistic cues in narratives: How salience influences anaphor resolution. *Memory & Cognition* 32(3): 511-522.
- Kreiner, H., Sturt, P. & Garrod, S. (2008). Processing definitional and stereotypical gender in reference resolution: Evidence from eye-movements. *Journal of Memory and Language* 58(2): 239-261.
- Lago, S. (2014). Memory and Prediction in Cross-Linguistic Sentence Comprehension. Ph.D. thesis, University of Maryland, United States.
- Levine, W. H., Guzmán, A. E. & Klin, C. M. (2000). When anaphor resolution fails. *Journal of Memory and Language* 43(4): 594-617.
- McDonald, J. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgement performance by late second language learners. *Journal of Memory and Language* 55: 381-401.
- McDonald, J. L. & MacWhinney, B. (1995). The time course of anaphor resolution: Effects of implicit verb causality and gender. *Journal of Memory and Language* 34(4): 543-566.
- Malaysian Examinations Council. (2019). 2019 MUET results.
<https://www.mpm.edu.my/en/sumber/penerbitan/laporan-peperiksaan-stpm-dan-muet/2019#> [12 Desember 2020].
- Masrai, A. (2019). Vocabulary and reading comprehension revisited: Evidence for high-, mid-, and low-frequency vocabulary knowledge. *Sage Open* 9(2).
- Mulder, E., van de Ven, M., Segers, E., Krepel, A., de Bree, E. H., de Jong, P. F. & Verhoeven, L. (2021). Word-to-text integration in English as a second language reading comprehension. *Reading and Writing* 34(4): 1049-1087.
- Nobre, N. (2011). Anaphora resolution. Master's thesis, Instituto Superior Técnico-Universidade Técnica de Lisboa.
- Peregoy, S. & Boyle, O. F. (2001). *Reading, Writing, & learning in ESL*. New York: Longman.
- Pretorius, E. J. (2005). English as a second language learner differences in anaphoric resolution: Reading to learn in the academic context. *Applied Psycholinguistics* 26(4): 521-539.
- Rayner, K. & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition* 14(3): 191-201.
- Rayner, K. & McConkie, G. W. (1976). What guides a reader's eye movements?. *Vision research* 16(8): 829-837.
- Rayner, K. & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review* 3(4): 504-509.
- Roberts, L., Gullberg, M. & Indefrey, P. (2008). Online pronoun resolution in L2

- discourse: L1 influence and general learner effects. *Studies in Second Language Acquisition* 30: 333- 357 38
- Schilling, H. E., Rayner, K. & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition* 26(6): 1270-1281.
- Schroeder, S. (2011). What readers have and do: Effects of students' verbal ability and reading time components on comprehension with and without text availability. *Journal of Educational Psychology* 103(4): 877.
- Swan, M. & Walter, C. (2017). Misunderstanding comprehension. *ELT Journal* 71(2): 228-236.
- Swering, L. S. (2015). Common types of reading problems and how to help children who have them. *The Reading Teacher* 69(5): 513-522.
- Seddik, K. M. & Farghaly, A. (2014). Anaphora Resolution. In *Natural Language Processing of Semitic Languages*, pp. 247-277. Springer, Berlin, Heidelberg.
- Simner, J. & Smyth, R. (1999). Phonological activation in anaphoric lexical access. *Brain and Language* 68: 40-45.
- Tainturier MJ, Tremblay M & Lecours AR. (1992). Educational level and the word frequency effect: A lexical decision investigation. *Brain and Language* 43: 460-474.
- Van Gompel, R. P. & Majid, A. (2004). Antecedent frequency effects during the processing of pronouns. *Cognition* 90(3): 255-264.
- Whitford, V., & Titone, D. (2012). Second-language experience modulates first- and second- language word frequency effects: Evidence from eye movement measures of natural paragraph reading. *Psychonomic Bulletin & Review* 19(1): 73-80. doi:10.3758/s13423-011-0179-5
- Winke, P., Gass, S. & Sydorenko, T. (2013). Factors influencing the use of captions by foreign language learners: An eye-tracking study. *The Modern Language Journal* 97(1): 254- 275.