# The effect of text difficulty on the missing-letter effect

Danielle Vien

Laurentian University

## The effect of text difficulty on the missing-letter effect

When reading for comprehension while searching for a target letter, errors occur. These errors are not randomly distributed. More precisely, target letters are more often omitted in frequent function words than in less frequent content words (Healy, 1976, 1994; Greenberg, Healy, Koriat, & Kreiner, 2004; Koriat & Greenberg, 1991; Moravscik & Healy, 1995; Roy-Charland, Saint-Aubin, Klein, & Lawrence, 2007; Roy-Charland, Saint-Aubin, Lawrence, & Klein, 2009). The discrepancy between the number of omissions for low-frequency content and high-frequency function words is known as the missing-letter effect, and it has been proposed to help understand cognitive processes involved in reading.

Research has documented the role of several factors important to producing the missing letter effect. The majority of the research has explored the role of word related factors that influence omission rates of a target letters, such as word frequency (e.g. Assink, van Well, & Knuijt, 2003; Moravcsik & Healy, 1995; Roy-Charland & Saint-Aubin 2006; Saint-Aubin & Klein, 2004), word function (e.g. Koriat & Greenberg, 1991; Saint-Aubin & Poirier, 1997), target letter position within a word (e.g. Assink & Knuijt, 2000), and prosody and stress pattern (see e.g. Mohan, 1978, Read, 1983; Schneider, Healy & Gesi, 1991). Of all these factors, two have been extensively examined and received the most support: word frequency and word function.

The unitization account and the processing time hypothesis were put forward to explain the missing-letter effect by calling upon word frequency as the main factor (Healy, 1994; Moravscik & Healy, 1995). According to this model, there are different levels of processing involved in reading, and the information gathered at each level gets passed from the lower levels to the higher levels - for example, from features of letters and letter levels to the higher levels of

words and phrases. However, when processing at a given level is completed, lower level processing, whether or not completed, is terminated. Frequency allows words to be processed faster as a whole, thus interrupting letter processing more quickly. Therefore, a word of high frequency has a higher chance of becoming unitized and bypassing the lower levels of processing, including the letter level, which leads to omissions of a target letter embedded in them. This model was supported by results from Roy-Charland et al. (2007) who made use of nine three-letter content words beginning with the letter "*d*". The frequency of these words varied between two and 2602 occurrences per million (New, Pallier, Brysbaert, & Ferrand, 2004). Results revealed a correlation between omission rates and frequency for the set of content words. It supported the importance of frequency by showing that, even if all words are content words, the higher their frequency, the higher their omission rates.

The structural account of the missing-letter effect (Koriat & Greenberg, 1991; 1994) emphasizes the important role of grammar in reading. Function words assume a structure-supporting role and act as anchors in a sentence. For this reason, less cognitive resources are required for function words to be processed during reading as the words are more predictable than contents words. Content words form the actual ideas and are less predictable in a sentence, and therefore require more resources. Since cognitive resources are more present with content words, the likelihood of a target letter being identified is higher than for function words. This was supported by Saint-Aubin and Poirier (1997), who made use of the French word "or", which can be used as a conjunction meaning whereas, and as a noun meaning gold. In a letter detection task, more omissions were made when "or" was used as a conjunction rather than when it was used as a noun. This is consistent with how function words require less attention than content words.

Since the importance of word frequency and word function was demonstrated in numerous studies, the previous models by themselves were unable to account for the importance of both factors. Recent models such as the GO model (Greenberg et al., 2004) and the Attentional Disengagement (AD) model (Roy-Charland et al., 2007, 2009) were proposed and incorporated the importance of both word frequency and word function. It is beyond the scope of this thesis to discuss the differences between these two models and furthermore, they both propose an interaction between word frequency and word function. The GO model is based on the belief that the cause of the missing-letter effect centers around the time spent processing the target word. Frequent words are identified faster such as function words, which can serve as cues for the structural organization of the sentence. This, in turn, guides attention to content words. Roy-Charland and Saint-Aubin (2006) tested the GO model by making use of two texts – one with two high-frequency words, one content and one function – it and at. The second text contained low-frequency words, also one content and one function – helmet and amidst. The position of the letter t in the word was the same for both word pairs. Overall, there was a higher omission rate of the target letter for high-frequency words than for low-frequency words but more specifically. However, while the letter embedded in high-frequency function words are omitted much more than high-frequency content words, there was no difference for low-frequency function and content words.

The current study will focus on explanations of the AD model, as it is the most recent instantiation aiming to explain the missing-letter effect. As attention is engaged on a word, information pertaining to its component letters is accumulated, and when attention disengages, the mental representation of accumulated information decays. The AD model indicates that readers' attention would be engaged on a content word longer than on a function word, as well as

longer on high frequency words rather than low frequency words due to attention. For this reason, omission rates will be increased for high frequency function words, rather than low frequency content words.

All the previously mentioned studies and models examined and explained the impact of individual word characteristics (frequency and function) on the omission rates. However, there are fewer studies that have examined the impact of the whole passage characteristics. This is extremely important since the main goal of reading is related to discourse and whole passage message communication. In other words, while single word recognition is essential to reading, the communication process requires whole passage processing. Camblin, Gordon, and Swaab (2007) describe how lexical processing alone is not enough as words seldom appear in isolation, but rather in a whole passage. A reader is able to interpret the meaning of a sentence or passage by gathering the information from information in preceding sentences.

To the best of our knowledge, in the missing-letter effect field, text familiarity is the only text-related factor that has been explored. Saint-Aubin, Roy-Charland and Klein (2007) examined the impact of text familiarity by having participants read a text multiple times. Overall, it was observed that the missing-letter effect is unaffected by repetition. In other words, participants made more omissions of target letters in frequent function words than in less frequent content words both when reading a text for the first time or for the third time. Furthermore, the size of the difference between omission rates for frequent function words and less frequent content words remains the same.

Saint-Aubin and Roy-Charland (2012) tested for familiarity effects on the missing-letter effect using self-generated texts. Self-generated texts are a more powerful manipulation than rereading a text as it will produce a level of familiarity that cannot be achieved through multiple

readings. The participants were required to create a text, and then perform the letter detection task with their own and anther participant's text. The missing-letter effect was still present in both the familiar and unfamiliar texts, but overall, the omission rates were reduced when searching in their own text. These results are not consistent with the previous study on familiarity, but could be due to the way that self-generated texts have a level of familiarity that cannot be achieved through rereading a text.

While text familiarity is an important component of discourse, it is not the only text-related factor than could affect communication. Other text related factors, such as text difficult, may play a role in the missing-letter effect. Text difficulty varies in everyday life - from reading newspapers to scientific journals, novels to text books, there are different levels of difficulty. By examining the impact of text difficulty on the missing letter effect, it will inform us about the impact of difficulty on cognitive processes in reading. The missing-letter effect is a great method of measuring mental processes as it does not interfere too much with a natural reading process, but is manipulating enough that we do not have to infer what is happening or having to rely upon subjective measurements.

The current study will examine the impact of text difficulty on the missing letter effect, which has not been looked at previously. With regards to speed, easy texts should be read more quickly than difficult texts, based on general research that has been conducted on reading . In addition, based on the trade-off effect in cognitive research (Healy, 1976) and on the conclusions about familiarity effect (Saint-Aubin et al., 2007; Saint-Aubin & Roy-Charland, 2012), an easier text should have overall lower omission rates since less resources are required for reading, thus freeing resource for the detection task. However, difficult texts would have more difficult content words than easy texts, which indicates content words should be omitted less in difficult

texts than in easy texts based on the attentional disengagement model (Roy-Charland et al., 2007, 2009).

## Methods

### Participants

Sixty-three students (53 female and 10 male) from Laurentian University participated in this study. All participants had English as their native primary language. Their ages ranged from 17 to 45 with an average age of 20.746. Most did not have a second language (37 participants), 24 had French as a second language, 1 had Italian, and 1 had Ojibway. The participants were recruited from classes at Laurentian University and with the help of the Cognitive Health Research Laboratory. The students received extra bonus marks for participating if their professor allowed it.

### Materials

Three passages were assembled. The first was a practice passage about Coca-Cola – its history and how it got to where it is today. The next two texts were taken from the North American Norming Study (NANS) with one being easy and the other difficult. The American norms were collected at the University of Illinois in Chicago, Illinois (UIC) and the Canadian norms were collected from Laurentian University in Sudbury, Ontario. The easy text selected for this study, entitled “Talk” was the easiest in Canadian norms (difficulty rating of 2.4 out of 10) and the 3rd easiest by American norms (difficulty rating of 2.8 out of 10) out of 32 texts. The text discusses the different ways that boys and girls communicate and their use of metamessages. Finally, the third text which is the difficult text, entitled “Social Investing” was the most difficult according to both Canadian (6.7 rating) and American norms (6.5 rating). It discusses the social investing community and a woman named Joan Bavaria molded the face of the social investing

community with her firm called Trillium. All texts had 6 comprehension questions associated with them.

The texts were arranged, justified so that the text was aligned with the left and right margins. The text was in size 11 Arial font. It was adjusted so that the target letter *t* would not appear in the left and right margins. Some words containing the target letter were not being considered in the analysis as to follow certain rules established by Roy-Charland, Saint-Aubin, McLaughlin, and Magee (currently in press). All words in which the target letter was embedded twice or in capitals, as well as words containing a hyphen, an apostrophe or with an ambiguous role were excluded in our analysis.

Each passage was followed with six multiple choice questions to test the participant for comprehension. The package also included a 30 word vocabulary questionnaire that consisted of choosing the best definition out of 5 options. For example, the word ascend has the options of A) to go up or mount, B) consent, C) improve with time, D) to leave behind or E) to replace a leader. A bilingual questionnaire was also included to indicate the participants' primary language and native language. Participants are also asked to indicate reading, writing and speaking abilities on a scale of 1 to 5 and at what age they learned this language, followed by if the participant has a second language and also asks the age at which it was learnt and their reading, writing and speaking ability on a scale of 1 to 5.

Finally, a demographic questionnaire was used. The questions targeted gender, the date of birth, age, and what year of university – be it first, second, third, fourth or graduate years.

**Procedure**

Participants were tested in groups (ranging in sizes from 1 to 20) in sessions lasting approximately 30 minutes. Each participant received a package. The first pages consisted of the

consent form. This was reviewed together and clarifications were made if participants did not understand any part of it. The instructions were read aloud to the group after which they were able to begin. Participants were instructed to read the texts for comprehension at their normal reading speed and to also circle the target letter *t*, either in upper or lower case. They were informed that they should not slow down their reading speed to get every *t*, and if they were ever to notice a *t* that was missed in a previous word, they should not retrace their steps to circle it. In addition, a clock was projected at the front of the room and the participants had to indicate the time, including the seconds, before they started reading and again when they had completed the passage. In order to promote reading for comprehension, they were informed that they would have to answer six multiple-choice comprehension questions following each reading task. All participants read both the easy and difficult texts after the practice text. Presentation order of the texts was counterbalanced across the participants, with half reading the difficult text first and half beginning with the easy text. Following the texts and comprehension questions was the vocabulary questionnaire, bilingual questionnaire and demographics questionnaire.

## Results

### Comprehension

Performance on comprehension questions was analyzed first. Results, shown in Table 1, revealed that participants were more accurate for the easy than the difficult text, [$F(1,62) = 14.09$, $\eta^2_p = .19$]. However, 10 of the participant did not respond to at least 50% of the questions (3 of 6) correctly for the easy text. Since comprehension is an essential component for the missing-letter effect, another analysis was performed after removing these participants. When excluded, again participants were more accurate for the easy than the difficult text [$F(1,52) =$

39.90, $\eta^2_p$ = .43]. It should be noted that results for the difficult text remained similar even when removing the participants with lower scores on the easy text.

**Reading Time**

Reading speed was also analyzed. Results, shown in Table 2, revealed that participants read the easy text more quickly than the difficult text, [$F$(1,62) = 104.01, $\eta^2_p$ = .627]. The same results were present for the 53 participants [$F$(1,52) = 96.65, $\eta^2_p$ = .65]. It should be noted that results for the difficult text remained similar even when removing the participants with lower scores on the easy text.

**Missing-Letter Effect**

Mean proportions of omissions for function and content words are presented in Table 3 as a function of text difficulty. A first analysis was computed on the complete sample. Inspection of this table reveals that, for both the easy and difficult text, participants made more omissions of the target letter when it was embedded in the function words rather than in the content words. A 2X2 repeated-measures ANOVA with text difficulty and word function as within-subjects factors revealed a significant main effect of word function [$F$(1,62) = 113.07, $\eta^2_p$ = .65] but neither the main effect of text difficulty [$F$(1,62) = 3.28, $\eta^2_p$ = .50] nor the interaction ($F$ < 1) were significant. Another analysis was conducted on the 53 participants that excluded the 10 participants who did not answer at least 50% of the comprehension questions correctly for the easy text (see Table 2). The pattern of results is identical to that of the complete sample. More precisely, results revealed a significant main effect of word function [$F$(1,52) = 79.74, $\eta^2_p$ = .61] but neither the main effect of text difficulty nor the interaction were significant (both $F$s < 1).

**Correlations**

Correlations were also computed between vocabulary scores and the MLE for the difficult text (omissions function – omissions content), the MLE for the easy text (omission function – omission content), omissions for content words and function words in the difficult text and omissions for content and function words in the easy text. None of the correlations were significant, all $rs < -.27$, $p > .06$.

## Discussion

There are many studies that have explored individual word characteristics, such as frequency and function, and their impact on the missing-letter effect. While single word recognition is essential to reading, one must be able to process the entire passage to gain understanding of a coherent message. Therefore, it is important to observe the entire passage characteristics rather than single word characteristics. The main goal of the current study was to determine whether or not text difficulty of a passage would have an effect on the missing-letter effect.

An inspection of the results for comprehension questions indicated that 10 participants did not score at least 50% on the comprehension questions for the easy text. This was problematic as it demonstrates that they may not have been reading while performing the letter detection task, but instead were only searching for instances of the letter *t.* Therefore, their packages would not establish reliable results as they were not performing the dual task of reading and searching. Therefore, second analyses for comprehension scores, reading times and omission rates were performed after removing these participants.

The results for the analyses of comprehension questions were fairly similar before and after removing the participants. It was revealed that, on average, participants responded more accurately to questions about the easy text than the difficult text. Results of the second analysis

compared to the first revealed that the average score for the easy text's comprehension questions increased but the average score for the difficult text stayed fairly similar to the average of the whole sample (see Table 1). This indicates that the readers may have adopted different reading and letter detection strategies for each text. The speed-accuracy trade-off effect can account for this. The participants may have used less cognitive resources on the detection task and more for the reading task, lessening their comprehension, for the easy text. But for the difficult text, they may have used more resources on the reading task than on the detection task, increasing their comprehension.

As expected, participants were slightly faster at reading the easy text than the difficult text. The gain in reading speed was expected as the words of the easy text would be disengaged from more quickly than those of the difficult text, as they are of higher frequency. In keeping with the AD model, content words in the easy text were likely more familiar and frequent than the difficult text's content words, causing attention to be disengaged more quickly and therefore, read faster. After removing the 10 participants who did not perform well on the comprehension for the easy text (see Table 2), the means of the reading times were slightly higher. The readers who were removed from the analysis may have been fast readers, which hindered their reading abilities, and therefore, their comprehension.

One main thing that was evident in this study would be that participants made more omissions when the target letter was embedded in frequent function than in less frequent content words while reading both the easy and difficult texts. That is to say, readers demonstrated the missing-letter effect. In addition, as shown by the non-significant interaction between word type and text difficulty, the size of the missing letter effect is of similar magnitude for both the easy

and difficult texts, indicating no impact of text difficulty on the MLE. The results did not confirm either of our hypotheses.

The attentional disengagement model indicates that attention plays a key role in the difference of omission rates between frequent function words and less frequent content words. The speed at which attention would be disengaged from the content words in the easy text compared to the difficult text may have been overestimated, or at least the extent to which this would cause fewer omissions. Future studies could take this into consideration.

Future studies could take into consideration different reading, as it was evident after removing the 10 participants, that some may adopt different reading and searching strategies for each text. The level of difficulty of texts could be considered as well – by making use of texts that are of much different difficulty levels could more clearly demonstrate whether or not there is an effect of text difficulty.

## Conclusion

As previously stated, whole passages are important to observe rather than individual words. In the current study, it was found that text difficulty does not affect omission rates during letter-detection tasks.

Table 1

*Means and Standard Deviations of Comprehension Question Responses*

| | Easy Text | | Difficult Text | |
|---|---|---|---|---|
| | M | SD | M | SD |
| N=63 | 4.16 | 1.44 | 3.35 | 1.36 |
| n=53 | 4.62 | 1.00 | 3.42 | 1.35 |

Table 2

*Means and Standard Deviations of Reading Times*

| | Easy Text | | Difficult Text | |
|---|---|---|---|---|
| | M | SD | M | SD |
| N=63 | 342.60 | 70.41 | 392.75 | 80.61 |
| n=53 | 347.98 | 67.54 | 399.81 | 74.03 |

Table 3

*Means and Standard Deviations of Omission Rates*

| | Easy Text (N=63) | | Difficult Text (N=63) | | Easy Text (n=53) | | Difficult Text (n=53) | |
|---|---|---|---|---|---|---|---|---|
| | M | SD | M | SD | M | SD | M | SD |
| Function | .41 | .24 | .43 | .21 | .38 | .23 | .39 | .26 |
| Content | .19 | .13 | .27 | .16 | .18 | .13 | .20 | .15 |

References

Assink, E.M.H., & Knuijt, P.P.N.A. (2000). Reading development and attention to letters in words. *Contemporary Educational Psychology, 25*, 247-262.

Assink, E.M.H., Van Well, S., & Knuijt, P.P.N.A. (2003). Contrasting effects of age of acquisition in lexical decision and letter detection. *American Journal of Psychology, 116*, 367-387.

Camblin, C.C., Gordon, P.C., & Swaab, T.Y. (2007). The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. *Journal of Memory and Language, 56*, 103-128.

Greenberg, S. N., Healy, A. F., Koriat A., & Kreiner H. (2004). The GO model: A reconsideration of the role of structure units in guiding and organizing text online. *Psychonomic Bulletin & Review, 11*(1), 428–33.

Healy, A. F. (1976). Detection errors on the word the: evidence for reading units larger than letters. *Journal of Experimental Psychology: Human Perception and Performance, 2*(2), 235–242.

Healy, A. F. (1994). Letter detection: A window to unitization and other cognitive processes in reading text. *Psychonomic Bulletin & Review, 1*(1), 333–344.

Koriat, A. & Greenberg, S. N. (1991). Syntactic control of letter detection: Evidence from English and Hebrew nonwords. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 17*, 1035–1050.

Koriat, A. & Greenberg, S. N. (1994). The extraction of phrase structure during reading: Evidence from letter detection errors. *Psychonomic Bulletin & Review, 1*, 345–356.

Mohan, P.J. (1978). Acoustic factors in letter cancellation: Developmental considerations. *Developmental Psychology, 14*, 117-118.

Moravcsik, J. E. & Healy, A. F. (1995). Effect of meaning on letter detection. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 82-95.

New, B., Pallier, C., Brysbaert, M., & Ferrand, L. (2004). Lexique 2: A new French lexical database. *Behaviour Research Methods, Instruments, & Computers , 36*, 516-524.

Read, J.D. (1983). Detection of Fs in a single statement: The role of phonetic recording. *Memory & cognition, 11*, 390-399.

Roy-Charland, A. & Saint-Aubin, J. (2006). The interaction of word frequency and word class: A test of the GO model's account of the missing-letter effect. *The Quarterly Journal of Experimental Psychology, 59*(1), 38-45.

Roy-Charland, A., Saint-Aubin, J., Klein, R.M., & Lawrence, M. (2007). Eye movements as direct tests of the GO model for the missing-letter effect. *Perception & Psychophysics, 69*(3), 324-337.

Roy-Charland, A., Saint-Aubin, J., Lawrence, M. A., & Klein, R. M. (2009). Solving the chicken and egg problem of letter detection and fixation duration in reading. *Attention, Perception, & Psychophysics, 71*, 1553-1562.

Roy-Charland, A., Saint-Aubin, J., McLaughlin, M., & Magee, S. (In press). Reading ability and syntactic processing revisited.

Saint-Aubin, J. & Klein, R.M. (2004). One missing-letter effect: Two methods of assessment. *Canadian Journal of Experimental Psychology, 58*(1), 61-66.

Saint-Aubin, J. & Poirer, M. (1997). The influence of word function in the missing letter effect: Further evidence from French. *Memory & Cognition, 25*(5), 666-676.

Saint-Aubin, J. & Roy-Charland, A. (2012). Is familiarity the all-purpose reading tool? The case of the missing-letter effect for self-generated texts. *Scientific Studies of Reading, 16*(1), 35-44.

Saint-Aubin, J., Roy-Charland, A., & Klein, R.M. (2007). The influence of multiple readings on the missing-letter effect revisited. *Memory & Cognition, 35*(7), 1578-1587.

Schneider, V.I., Healy, A.F., & Gesi, A.T. (1991). The role of phonetic processes in letter detection: A reevaluation. *Journal of Memory & Language, 30*, 294-318.