One Missing-Letter Effect: Two Methods of Assessment

Jean Saint-Aubin, Université de Moncton
Raymond M. Klein, Dalhousie University

Abstract When participants search for a target letter while reading, they make more omissions if the target letter is embedded in frequent function words than in less frequent content words. This effect is usually observed with a paper and pencil procedure. It has been shown that a similar pattern is observed using a rapid serial visual presentation procedure in which words appear one at a time on a computer screen. It has been questioned, however, whether the two methods tap the same cognitive processes. Item-based correlations between the paper and pencil and the rapid serial visual presentation procedure were high and not significantly different from reliability estimates of either procedure. It is concluded that both procedures highlight the same cognitive processes that are responsible for the missing-letter effect.

When readers are asked to circle all instances of a target letter while reading a prose passage for comprehension, they miss more target letters embedded in frequent function words than in less frequent content words (Corcoran, 1966; Healy, 1994; Koriat & Greenberg, 1994). This well-replicated phenomenon, termed the missing-letter effect, has been used to investigate the cognitive processes involved in reading. As of yet, this phenomenon has been intensively studied with a paper and pencil procedure. In an effort to evaluate the major competing theoretical proposals accounting for the missing-letter effect, investigators have begun to study it using different procedures (Hadley & Healy, 1991; Healy, Oliver, & McNamara, 1987; Saint-Aubin & Klein, 2001; Saint-Aubin, Klein, & Roy-Charland, 2003). Among these procedures, the rapid serial visual presentation (RSVP) procedure in which words are presented one at a time at the centre of a computer screen has been found very promising for such an evaluation.

The RSVP procedure was initially introduced in the field by Healy et al. (1987) to examine the effects of display size on the missing-letter effect. Subsequently, based on RSVP and other findings, the parafoveal processing hypothesis was proposed to account for the influence of eye movements on the missing-letter effect (Hadley & Healy, 1991). The parafoveal processing hypothesis is based on the observation that frequent function words are more likely to be skipped during reading than control content words of the same length (O’Regan, 1979; Saint-Aubin & Klein, 2001). When a word is not fixated, it is assumed to be identified in the parafovea, during fixation of the previous word (see, e.g., Rayner, 1999). Because resolution decreases as the printed material is further away from the fovea, letter identification would be harder for nonfixated words. Because frequent function words are more likely to be skipped during reading, there would be more omissions of their target letter.

In support of this hypothesis, Healy et al. (1987) used an RSVP procedure in which words were presented one at a time at the centre of a computer screen. Because all words are presented at fixation, omission rate should be similar for frequent function words and less frequent content words. Although their results confirmed this prediction, Saint-Aubin and Klein (2001) subsequently showed that this was due to their usage of a too slow presentation rate (between 500 ms and 1,500 ms per word). With presentation durations closer to the usual fixation durations in reading (200, 250, and 350 ms per word), the usual missing-letter effect was found with a variety of target words and in both French and English (Saint-Aubin & Klein, 2001; Saint-Aubin et al., 2003).

Despite the observation of a large missing-letter effect with an RSVP procedure, it could be argued that the missing-letter effect in this procedure is due to different processes. Indeed, there have been some doubts about its validity for distinguishing between explanations of the missing-letter effect. For example, it has been argued that while with a usual paper and pencil procedure, readers are assumed to move on to the next text segment only after completing identification of the word currently in view, with an RSVP procedure, readers might be exposed to the next word before completing the identification of the word currently in view. Based on this logic, some mechanisms have been suggested to produce a missing-letter effect that would be an artifact of the RSVP procedure and not a by-product of the cognitive processes highlighted by the missing-letter effect (see Healy’s suggestion in Saint-Aubin & Klein, 2001). The aim of the present paper was to
examine the appropriateness of the RSVP procedure. Several methods might be applied to this question. In this paper we will determine whether the two methods of measuring target letter omissions produce similar profiles of performance using the same text with both procedures. This test was implemented by a re-analysis of existing data from an RSVP procedure (Saint Aubin et al., 2003) together with newly collected data from a paper and pencil procedure using the same text.

Method

Participants

Sixty-eight unpaid undergraduate students (53 women, 15 men) attending an introductory psychological course at Université de Moncton volunteered to participate in the experiment. French was the first language for all the participants.

Materials

The des text used by Saint-Aubin et al. (2003) was used here. The text comprised 593 words, and the target letter was d. There were 24 instances of the French plural indefinite article des, with a frequency count of 15,877 occurrences per million, and 24 instances of three-letter control content words beginning with the letter d, with an average frequency count of 64 occurrences per million (e.g., don [gift or donation], dit [sayls], vix [trem], des [back], dino [duet]) (Baudot, 1992). The 24 occurrences of the control content words were composed of nine different words for which there were between one and four occurrences. In addition, the target letter d was embedded in 14 non-test words. Within the text, each word containing the target letter, be it critical or not, was separated from the previous and the following one by at least four filler words without the target letter. The test words were not included in the first and the last sentence of the text, they were never at the beginning or the end of a sentence, and they were never adjacent to a punctuation mark.

An instruction sheet was stapled in front of the text. The instructions encouraged the participants to read the text for comprehension at their normal reading speed. They were told that whenever they came across a d (either in upper- or lower-case) they were to circle it. They were also warned not to slow down their reading speed to detect all des, and not to backtrack to circle a d they had missed. To familiarize participants with this task, they were asked to read the instruction sheet again and to circle all ds.

Procedure

The participants were tested in two groups of about equal size. The experimenter read the instructions aloud, while the participants read them silently. Before the participants started practice on the instruction sheet, they were informed that they would have to answer three multiple-choice comprehension questions after the text. This procedure was used to promote reading for comprehension.

RSVP Methods and Data

Data from the des text of Experiment 1 from Saint-Aubin et al. (2003) were re-analyzed to provide omission rates for individual items using the RSVP procedure. In Experiment 1 of Saint-Aubin et al., the des text was presented with only one word displayed at a time, on the center of a computer screen linked to a Macintosh computer. Words were presented with normal capitalization, and punctuation marks were presented along with the word to which they were attached in the text. Each word was presented for 250 ms. Participants pressed the space bar of the keyboard to respond. Response latencies were measured from the onset of a word containing the target letter until the participant pressed the space bar. A response was scored correct if it occurred during the first 900 ms after the onset of a test word. Participants were tested individually in a private room. The instructions were the same as those used here, except that participants were also told to respond carefully because both their speed and accuracy were scored. Here, only omission data were used for the re-analysis.

Results and Discussion

Paper and Pencil

The mean percentages of omission errors for both des and its control content words were computed for each participant by dividing the number of omissions by the number of occurrences of the text word (namely 24). The analysis of variance (ANOVA) with word function (des vs. control content words) as the only factor revealed that there were significantly more omissions for des (33.5%) than for its control content words (17.0%). $F(1,67) = 51.93, MSE = 177.87, p < .001$, $F(1,67) = 21.99, MSE = 148.27, p < .001$. The observed missing-letter effect is of about the same magnitude as the one observed by Saint-Aubin et al. (2003) in Experiment 1, with an RSVP procedure with 41.3% of omissions for des and 21.3% for its control content words. Item-based analysis (the eta squared ($\eta^2$) statistic) revealed that the main effect of role accounted for 32% of item variance. A re-analysis of Saint-Aubin et al.'s data revealed that role also accounted for 32% of item variance when the RSVP procedure was used.

Besides the high similarity of results with both procedures, and before moving to the additional analyses further investigating their relationship, one empirical result deserves comment. More specifically, both
Figure 1. Mean z-score for omission rate for the 24 occurrences of *des* (left panel) and for the 24 occurrences of the control content words (right panel) as a function of their order of presentation in the text. The upper panel presents data from both halves of the sample undertaking the paper and pencil procedure; the middle panel presents data from both halves of the sample undertaking the RSVP procedure, and the lower panel presents data from the entire sample with the paper and pencil procedure as well as data from the RSVP procedure (the *des* text used in Experiment 1 of Saint-Aubin et al., 2005).

Inspection of Figure 1 and η² results just presented suggest that much of variance in letter detection is not captured by word function. This is best explained by the most recent model of the missing-letter effect: the GO model (Greenberg, Healy, Koriat, & Kreiner, in press). According to the GO model, at least two other factors can influence omission rate over and above the role of the target word. Based on previous work by Healy and her associates, it is assumed one of these is word frequency (see, e.g., Moravcsik & Healy, 1995). Here, omission rate variations among content words can be partly attributed to word frequency because content words were of various frequencies. Indeed, with the RSVP data used here, Saint-Aubin et al. (2005) found a significant positive correlation of .54 between the log frequency of the 24 occurrences of control content words and their omission rate. For the paper and pencil data of the current study, the corresponding correlation is .63. Whereas word frequency can account for variance among content words, it cannot account for variance among the 24 occurrences of *des*, because it is always the same word. However, according to the third assumption of the GO model, contextual constraints are critical in helping to identify sentential slots where structure supporting units like function words are likely to reside. As a consequence, those constraints would influence omission rate among function words. Here, it can be assumed that over sentences, contextual constraints were not always suggesting with the same strength the presence of a function word, which would explain the variation in omission rate among the 24 occurrences of *des*. 

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
participants with the paper and pencil procedure and 60 with the RSVP procedure). Omission rate for each of the 48 occurrences of the target words (24 *des* and 24 control content words) was computed separately for each half and procedure. The normal z score of omission rate for each occurrence of the 48 target words for each half are presented in the two upper sections of Figure 1. As can be seen, for both procedures, and for both *des* and its control content words, the pattern of results is remarkably stable over the 24 occurrences of each target. For instance, there is no trend toward an increment in omission rate as one moves from the 1st to the 24th occurrence as could be expected, for example, if the missing-letter effect were due to a semantic satiation (see Smith & Klein, 1990) or repetition blindness (see Kanwisher, 1987) effect. More importantly, for both procedures, the curves representing the data of each random half of the participant sample are very similar, which translates into a significant item-based correlation of .77 for the paper and pencil procedure and of .80 for the RSVP procedure. Even after controlling for the influence of word function – by assigning a function role to all 24 occurrences of *des* and a content role to all 24 control content words – the correlation remains significant at .68 for the paper and pencil procedure and at .73 for the RSVP procedure.

An item-based correlation was then computed between, on the one hand, mean omission rate for each of the 48 target words obtained with the total sample of 68 participants undertaking the paper and pencil task, and, on the other hand, the 60 participants of Saint-Aubin et al.'s (2003) first experiment undertaking the letter detection task with the same test, but with an RSVP procedure. The normal z scores of omission rate for both procedures are presented in the lower section of Figure 1. Inspection of this figure reveals a pattern of results as similar for both tasks as it is for both halves of the paper and pencil and the RSVP procedures. This trend was confirmed by a significant item-based correlation of .75 between both procedures and of .69 after controlling for word function. As can be seen in Figure 2, correlations and partial correlations across tasks are very similar to correlations between both halves of the same task. This was confirmed by a series of z-tests performed to verify if the correlation between both procedures was as large as it could be given our reliability estimates. Z-tests revealed that the correlation between procedures was not significantly different from the separate reliability estimates (z = 0.11, *p* = .81, and z = 0.35, *p* = .73, compared to the reliability estimates of the paper and pencil and RSVP procedures, respectively), and the same was true for the partial correlations controlling for word function (z = 0.19, *p* = .81, and z = 0.82, *p* = .41, respectively).

**Figure 2.** Split-half reliability correlations for the paper and pencil and RSVP procedures and the correlations between these two procedures are shown as vertical bars. These and the 95% confidence intervals for the cross-procedure correlations are referenced to the left Y-axis. The horizontal lines represent the upper limit for the cross-procedure correlations adjusted for attenuation (Nunnally & Bernstein, 1994; see text for explanation) which are referenced to the right Y-axis. The upper panel presents correlations and the lower panel presents partial correlations controlling for the influence of word function.

**Item-Based Analyses**

Similarity in overall effect size, as observed with the ANOVAs, is consistent with the assertion that similar underlying causal mechanisms are operating in the two situations. A similar profile of performance across the individual items, however, would provide a much more compelling basis for this claim. Before moving to the analyses designed to compare performance between the paper and pencil and the RSVP procedures, it is important to assess the reliability of the letter detection task with both procedures. These reliability estimates will be used as the baseline against which the strength of the relationship between both procedures will be evaluated. Reliability was estimated by randomly splitting into two equal halves the two original samples (68
These findings are consistent with the fact that the split-half reliability correlations fall well within the 95% confidence intervals for the cross-procedure correlations (Figure 2).

Despite the fact that significant full and partial correlations were observed between both procedures and the fact they are not significantly different from the correlations observed between the random halves of each procedure, it could be argued that there is a more direct test. In effect, it is possible to compute an estimate of the correlation between the two procedures if they were both perfectly reliable (Nunnally & Bernstein, 1994). This correction, called correction for attenuation, revealed that if both procedures were perfectly reliable, the expected correlation between the paper and pencil and the RSVP procedures would be .95, while the partial correlation controlling for word function would be .91 (see Figure 2, right axis). The closeness of these corrected correlations to a perfect correlation of 1.0 strongly supports the conclusion that both procedures are highlighting the same processes.

Conclusion

Proposals about the source of the missing-letter effect have emphasized the relative likelihood that words are fixated while reading (the parafoveal-processing hypothesis, Hadley & Healy, 1991), the relative processing time for word versus letter identification (the processing-time hypothesis, Moravcsik & Healy, 1995, 1998), and the relative salience of lexical items in the reader’s mental representation of the text (the structural account, Koriat & Greenberg, 1991, 1994, 1996). The RSVP procedure is particularly well suited for exploring these hypotheses. It not only ensures that all items will be at fixation and allows the experimenter to control item exposure duration, instead of leaving those parameters under the participant’s control, but also allows the measurement of target detection time (Saint-Aubin & Klein, 2001; Saint-Aubin et al., 2003).

Controlling for or measuring fixation is necessary for evaluating a class of proposals subordinating the missing-letter effect to eye-movements (e.g., the parafoveal-processing hypothesis, Hadley & Healy, 1991), whereas controlling for exposure duration or measuring fixation duration is necessary for evaluating proposals about processing time (e.g., the processing-time hypothesis, Moravcsik & Healy, 1995, 1998; the GO-model, Greenberg et al., in press). Finally, measurement of detection time has proven essential for evaluating both the processing-time hypothesis and the structural account.

In conclusion, if there were doubts about the appropriateness of using the RSVP methodology to explore the missing-letter effect or to distinguish between theories that seek to explain it, the nearly perfect similarity between omissions in the RSVP procedure and the traditional paper and pencil procedure should go a long way toward dispelling them.

This research was supported by a discovery grant from the Natural Sciences and Engineering Research Council of Canada to each author and by a grant from the Canadian Language and Literacy Research Network (CLLRN) to both authors. We thank Arnie Jabbert and Isabel Savoie for their assistance in running the experiment.

Correspondence concerning this article should be addressed to Jean Saint-Aubin, École de psychologie, Université de Moncton, Moncton, New Brunswick E1A 3E9. (E-mail: saint-a@umoncton.ca or klein@cr.psychology.dal.ca).

References
Sommaire

Lorsqu’un lecteur doit rechercher une lettre cible tout en lisant un texte continu, il l’omet plus fréquemment lorsqu’elle est incluse dans un mot de fonction très fréquent comme des plutôt que dans un mot de contenu moins fréquent. Ce phénomène expérimental désigné sous le nom d’effet de la lettre omise est l’un des phénomènes expérimentaux les plus robustes dans l’étude des processus cognitifs associés à la lecture. L’effet de la lettre omise est habituellement obtenu avec une procédure de type papier et crayon. Cependant, un effet similaire a également été observé avec une procédure de présentation sérielle visuelle rapide au cours de laquelle les mots sont présentés rapidement l’un après l’autre sur l’écran d’un ordinateur (Saint-Aubin & Klein, 2001; Saint-Aubin, Klein, & Roy-Charland, 2003). La présente étude a pour but d’explorer jusqu’à quel point les deux procédures mettent en lumière les mêmes processus cognitifs. Pour ce faire, les données rapportées par Saint-Aubin et al. à l’expérience 1 avec une procédure de présentation sérielle visuelle rapide sont utilisées. Ces données sont mises en relation avec des données nouvellement recueillies auprès de 68 étudiants universitaires et ce, en utilisant le même texte qui compte 24 occurrences du mot de fonction très fréquent des et 24 occurrences de mots de contenu moins fréquents de trois lettres et débutant par la lettre cible d (par exemple, don, du, dit, diz, dos).

Les résultats recueillis avec la procédure papier et crayon révèlent que, tout comme avec la procédure de présentation sérielle visuelle rapide, les lecteurs ont tendance à ométer la lettre cible d lorsqu’elle est incluse dans le mot des (33,5 %) plutôt que dans les mots contrôles (17,0 %), $t(167) = 51,93$, $MSE = 177,87$, $p < .001$. Qui plus est, tout comme pour la procédure de présentation sérielle visuelle rapide, le rôle du mot cible explique 32 % de la variance du taux d’omissions des items (y²). Par la suite, le patron d’omissions obtenu avec chacune des deux procédures, pour les 48 mots cibles du texte, fait l’objet d’une comparaison systématique. Une corrélation positive significative de .75 est observée et cette corrélation demeure significative même après avoir contrôlé statistiquement pour l’influence du rôle dans la phrase (.64). La valeur de ces corrélations est comparée aux indices de fidélité de chacune des deux procédures. Les indices de fidélité sont obtenus en divisant de façon aléatoire en deux moitiés équivalentes chacun des deux échantillons : l’un avec la procédure papier et crayon et l’autre avec la procédure de présentation sérielle visuelle rapide. Une corrélation basée sur les 48 items est calculée entre le taux d’omissions obtenu pour chacune des deux moitiés. Pour la procédure papier et crayon, la corrélation est de .77, alors qu’elle est de .80 pour la procédure de présentation sérielle visuelle rapide. De même, la corrélation partielle qui contrôle le rôle dans la phrase est de .68 pour la procédure papier et crayon et de .73 pour la procédure de présentation sérielle visuelle rapide. Une série de tests z révèle ensuite qu’il n’y a pas de différences significatives entre la valeur de la corrélation ou de la corrélation partielle entre les deux procédures et celles obtenues entre les deux moitiés de chacune des procédures. A la lumière de ces résultats, il semble que les mêmes processus cognitifs soient à la base de l’effet de la lettre omise observé avec les deux procédures.