Applied Psycholinguistics **32** (2011), 1–17
doi:10.1017/S0142716410000263

Evolution of the missing-letter effect among young readers between ages 5 and 8

DENIS FOUCAMBERT
Université du Québec à Montréal

JACQUES BAILLÉ
Université P. Mendès France, Grenoble

Received: September 16, 2008 Accepted for publication: July 14, 2009

ADDRESS FOR CORRESPONDENCE
Denis Foucambert, Département de linguistique, Université du Québec à Montréal, C.P. 8888, succ. Centre-Ville, Montréal, QC H3C 3P8, Canada. E-mail: foucambert.denis@uqam.ca

ABSTRACT
In light of the numerous studies on the detection of target letters among adults, it is generally accepted that the missing-letter effect depends both on a given word's frequency in its language and on its role (function vs. content) in a sentence. Following a presentation of several models explaining these observations we analyze the results of a letter-detection task given to 886 French students from kindergarten to second grade. The purpose of the present study is to determine the moment when the sensitivity to content/function word distinction emerges. The results of this study reveal that even if word frequency plays a role in letter detection, the emergence of an ability to extract sentence structure, along the lines of the structural model of reading, is significantly linked to the initial stages of explicit reading instruction.

In experimental psychology, letter-detection tasks have been used to study attention and awareness for many years. Recently, the missing-letter effect (MLE; Corcoran, 1966; Healy, 1976) has been gaining considerable attention in research on the reading process. Participants are typically asked to read a given text at normal speed while crossing off a specific letter (target letter) whenever encountered. They are also instructed to not turn back to cross off letters that they think they might have forgotten.

Research shows that readers fail to detect a certain number of target letters, and that the distribution of these omissions is not random, but rather appears to be governed by certain rules. Along these lines, Healy's work has consistently revealed statistically significant elevated rates of letter omission in English high-frequency words such as *and*, *the*, or *for*. To explain these results, the researcher and her colleagues developed the unitization model, a framework in which the reader simultaneously processes text on several levels. These processes are based on visual feature, letter, syllable, and word stimuli (Healy, 1976, 1994; Healy & Drewnowski, 1983). Once the reader identifies a unit through one of these analyses, the other analyses are stopped even if they have not been carried out to completion. In the most general terms, the unitization model is a multicued process, with word frequency as one cue; the omission of letters is interpreted by Healy as the result of rapid, global recognition of a word because of its frequency (Drewnowski & Healy, 1982).

Research conducted by Greenberg and Koriat has resulted in the creation of an entirely different explanatory framework for the MLE (Greenberg, Koriat, & Shapiro, 1992; Koriat & Greenberg, 1991, 1993, 1994, 1996; Koriat, Greenberg, & Goldshmid, 1991). Recognizing that the most frequent words are often grammatical or function words, the authors claimed that it is actually their structuring nature that induces the MLE. Their position resulted in the creation of what is known as the structural account of the reading process, which is based on the assumption that differences in letter omissions are because of the syntactic role that certain words play in the organizational structure of sentences rather than their frequency. According to this model, the reader processes prominent function words, such as prepositions, articles, and conjunctions in order to establish a temporary skeletal frame for a given phrase or sentence onto which meaning derived from content words are subsequently integrated. These functional words then recede into the background as the integration of meaning unfolds within the preestablished structure.

However, these two models are not mutually exclusive. In 2004, an article was published in which Greenberg, Healy, Koriat, and Kreiner incorporated certain aspects of their respective models into a larger model capable of integrating all the seemingly divergent observations made in previous studies. This model, the guidance–organization (GO) model, is based on the assumption that highly frequent function words are identified through both parafoveal vision and the processes originally laid out by Healy in her unitization model. These words are then used as pivots around which a reader organizes the syntax of a given sentence. This syntactic organization then guides the reader's attention to the content words, thus allowing for the analysis of the semantic information of the sentence and its integration into the structural frame established by the reader (Greenberg et al., 2004).

Integrating relevant elements such as bottom-up processes derived from the unitization account (*the unitization assumption* and *the parafoveal processing assumption*) and top-down processes derived from the structural account (*the structural precedence assumption* and *the guidance assumption*), the GO model allows for a more detailed analysis of the reading process. The last assumption, the guidance assumption, lies at the crux between bottom-up and top-down processes and explains how the MLE is attributed to the predictability of the word in a text, this predictability being a function of the contextual constraints that allow for anticipation of sentential slots where the syntactic organizing units are likely to be positioned.

One of the strengths of the GO model is its capacity to be experimentally tested. In an effort to find results congruent with the GO model, Roy-Charland and Saint-Aubin (2006) investigated the interaction (in both French and English) between word class and frequency in the MLE. Using a new experimental procedure, Foucambert (2008) confirmed the importance of parafoveal information in the construction of syntactic structures, shedding new light on the contextual constraint assumption. Finally, Greenberg et al. (2004) found several similarities between eye-movement experiments and GO model predictions, thus calling for more research along the same lines. Recently, Roy-Charland, Saint-Aubin, Klein, and Lawrence (2007) published a particularly important article in which they carried out a systematic study of GO model predictions on eye movement. Although their results were congruent with certain GO model predictions, it is important to note that predictions bearing on response times were not. Contrary to GO model predictions, the authors observe shorter fixation durations on lexical words than on functional words. From these observations, Roy-Charland et al. (2007) elaborated a new explanatory model for the missing letter effect: the attentional–disengagement (AD) model. Inspired by information processing models, the AD model posits that "the missing-letter effect reflects the allocation of attention during reading" (Saint-Aubin & Klein, 2008, p. 133). Letter omissions occur when readers disengage their attention from a word, which happens more often with frequent words than rare words, thus more often with function words than lexical words.

Nonetheless, the development of these word- and morpheme-processing skills on the basis of the syntactic role of words or their frequency of appearance in the written language remains largely incomplete (Koriat & Greenberg, 1991). Some studies have attempted to link the MLE with the reader's age; most notably, Greenberg, Koriat, and Vellutino (1998) showed that the more readers advance in age the more letter omission responds to certain categories of words (function vs. content). For these authors, this trend reflects a growing awareness of the sentence structure and the role that certain categories of words play within that structure. However, although their study compares students from a relatively wide age range (7 to 13 years old), it does not answer the question of when this sensitivity to the role and nature of words in sentence construction develops. Saint-Aubin and Klein (2008) wrote a critical review of the work examining the MLE among young readers. In general terms, the MLE becomes progressively more prominent with age (Cunningham, Healy, Kanengiser, Chizzick, & Willitts, 1988; Drewnowski, 1978, 1981; Saint-Aubin, Klein, & Landry, 2005). However, the results in the literature reviewed by Saint-Aubin and Klein (2008) are difficult to interpret because of some uncertainties generated by the experimental designs of the reported studies, for example, the uncontrolled position of the letter in the word *the* versus *mother* in Drewnowski (1978) or the repetition of a task using the same textual material (Drewnowski, 1978, 1981), a condition that is known to either reduce the MLE (Healy, Oliver, & McNamara, 1987) or make it disappear altogether (Saint-Aubin, Roy-Charland, & Klein, 2007).

Two recent studies have revisited this question around the relationship of the factors that accompany the emergence of the MLE: the age and skill level of young readers. First, the results reported by Saint-Aubin et al. (2005) shed light on questions concerning the relationship between selective letter omission and learners' introduction to explicit reading instruction. In their experiments, they observed an evolution of the MLE in the frequency and class of the word in which the letters appear among children from first to seventh grades. Their main conclusion suggests that if a missing-letter differential is observed starting at the end of kindergarten, it cannot be attributed to the syntactic or lexical role of the word before the end of fourth grade. Second, Saint-Aubin and Klein (2008, p. 135) published an article with the objective of being "the first comprehensive examination of the relationship between the missing-letter effect and degree of acquisition of reading skills." Evaluating the young readers using the Wide Range Achievement Text—Third Edition test, they demonstrated that from first to fourth grade the MLE is positively correlated to the students' reading skill level. However, for the older students (Grade 5), the difference between good and poor readers only solidifies, in terms of the MLE, with the function word *the*. In light of these results, the authors conclude that the distinction between good and poor readers lies in the capacity of the former to organize the structure of a sentence being read, such a conclusion concurring with the structural side of both the GO (Greenberg et al., 2004) and the AD (Roy-Charland et al., 2007) models. However, this skill only emerges in the fifth grade. Moreover, this result is congruent with the preliminary results from Saint-Aubin et al. (2005), which demonstrate that the role of word class does not emerge before the fourth grade in the explanation of the MLE.

The present study aims to build on previous work done by Saint-Aubin et al. (2005), adding two distinguishing elements. First, we wanted to run an initial evaluation before the onset of explicit reading instruction, then we wanted to verify the same hypotheses using authentic reading material (a children's book). Concerning the first point, Saint-Aubin et al. (2005) administered their tests at the end of the academic year, 9 months after the beginning of explicit reading instruction, which eliminates the possibility of comparing postresults with the results of the students' performance prior to explicit reading instruction. Before even entering the first grade, children have already been more or less involved in the discovery of writing, mainly through the readings done by their teachers in the lower grades and by the practice of reading at home within the children's families. Two studies (Baker, Fernandez-Fein, Scher, & Williams, 1998; Bus, van IJzendoorn, & Pellegrini, 1995) have revealed an important link between numerous aspects of children's knowledge of written language and culture, including a better understanding of the alphabet, and the role of writing and vocabulary. Under these conditions, the influence of word frequency on the MLE could be observed even before the start of explicit reading instruction. Second, when using an authentic text to measure the MLE, it seemed important to replicate these measures in the most authentic of conditions, especially with young children. However, compared to the more classic studies on the MLE, the experimental material runs the risk of losing internal validity as a result of a lack of control of certain elements that would normally be considered in other studies. This is why, in the present study, the characteristics related to word size, and especially letter position within words, were integrated into statistical models as control variables to the extent that many studies have established a link between letter detection and the visual configuration of the words which contain them (Nazir, Jacobs, & O'Regan, 1998; White, Rayner, & Liversedge, 2005). Moreover, considering the linguistic material, and the extent to which students are in the process of consolidating their control over grapheme–phoneme correspondences, it would be interesting to introduce this dimension into the analysis of analyzed material (Gross, Treiman, & Inman, 2000).

The objective of the present study is to analyze the results of a letter-detection task among younger students (ages 5 to 8) in an attempt to answer the following question: when does sensitivity to content/function word distinction emerge? It would be reasonable to claim that a lack of knowledge of written language could produce a randomizing effect on the MLE for the youngest of kindergarten students. In other words, the MLE would play a less important role on letter recognition, rendering the pattern of omitted letters more random. In contrast, the relative familiarity that older students (first and second grades) have with words and sentences should progressively produce a more regular pattern in the MLE in terms of the influence exerted by the syntactic and semantic role of the observed words.

METHOD

Participants

A total of 886 students from 45 different classes participated in this series of experiments. The participant sample consisted of 116 kindergarten students (mean age = 6.1 years, $\sigma = 33$ weeks), 420 first-grade students (mean age = 7 years, $\sigma = 24$ weeks), and 350 second-grade students (mean age = 8.1 years, $\sigma = 25$ weeks).[1] This distribution is normal and is typical of the population distribution in French schools.

Materials

A two-page booklet was distributed to the students. A very short text was given on the first page, allowing them to become familiar with the task, but using a different target letter. The second page contained a 116-word text, *Les baisers du Loup* (Solotareff, 2001), which served as the base for the experiment. The texts (see Appendices A and B) were printed in 24-point Garamond type with single-line spacing. The target letter *u* appeared 36 times in the text (without counting the first and last sentences of the text). However, to be more closely aligned with the normal conditions of the task, 7 occurrences were deleted because the letter *u* was capitalized (*Une, Un*) or because it appeared at the beginning or the end of the line (*leur* and *plume*) or sentence (*poulailler, cour, loup*). In the end, 29 occurrences were observed.

Procedure

The experiment was conducted in class at the end of the school year in June. The teachers did not read or present any part of the two texts. The following is a copy of the instructions provided to the students: “With a pen, and while browsing the text, while reading it, to the best of your ability, and at your own speed, without going back, you are going to cross off the letter *u* (the letter is written on the board) whenever you encounter it.”[2] These instructions were repeated before starting the test on the second text.

Measurements

For every word in which the target letter is found, we established several variables characterizing it.

- Word size: This variable was introduced into the model to control for the well-documented effects of word length extraction from parafoveal information (Kennedy, 2000; White et al., 2005) and the relatively frequent skipping of short words by the reader (Rayner, 1998). Previous studies on eye movement have shown that function words (articles, conjunction, prepositions) are omitted from eye fixations during reading (Carpenter, Just, & Rayner, 1983), as well as that three-letter words are skipped more often when they are functional (*the*) than when they are verbs (O'Regan, 1979). Consequently, frequent words, as is the case with small words, would often be identified in the parafoveal vision (by the peripheral zone of the retina) during fixation on a preceding word, whereas the less common words would be processed by the foveal zone of the eye (Rayner, Pollatsek, & Coltheart, 1987).
- The position of the letter in the word: To represent this value, we calculated the ratio between the position of the letter and the size of the word. The closer this value is to 1, the closer the letter is to the end of the word. Prior studies have shown that the eye frequently positions itself toward the end of the first third of the word (O'Regan & Jacobs, 1992; Vitu, 1991; Vitu, O'Regan, Inhoff, & Topolski, 1995).
- The frequency of the word in which the letter appears: This frequency is derived from the database FRANTEXT,[3] which provides information on contemporary French words from 1950 to 2000 (New, Pallier, Ferrand, & Matos, 2001). In order to normalize the distribution of this variable (that ranged from 0.3 to 11,469 per million) in the following analyses, we use the normal logarithm of the different values.
- Whether the letter *u* studied represents the phoneme [y] in the word (*pur* vs. *jour*).
- The type of word in which the letter is present (function vs. content): This variable was based on word categories. Nouns, personal pronouns, verbs, adjectives, and adverbs are considered content words in that they carry information about a character or an action in the text. Their printed frequencies ranged from 57.8 to 11,468.6 per million (M = 5,329.3, SD = 4,274.6). Function words, on the other hand, organize the structure of content words and carry little or no semantic information. Conjunctions, articles, relative pronouns, and prepositions are examples of function words. Their printed frequencies ranged from 0.3 to 4,519.8 per million (M = 867.8, SD = 1,448.9).

RESULTS

The results were analyzed from two angles. We intended to determine whether the analyzed data would reveal a difference according to the length of formal exposure to written text in school (operationalized by grade level) and whether they would indicate, within the same grade level, which of the five variables best explains the MLE.

Table 1. *Means (standard errors) for percentages of omissions according to grade level, pronunciation, and word class*

	Grade Level		
	Kindergarten	Grade 1	Grade 2
Pronunciation			
Yes	30.80 (1.20)	22.12 (1.89)	23.24 (2.46)
No	29.82 (0.91)	21.65 (1.42)	24.08 (1.85)
Word class			
Content	29.80 (0.98)	16.51 (1.46)	17.11 (1.91)
Function	30.81 (1.25)	27.27 (1.97)	30.21 (2.58)
All words	30.31 (0.7)	21.89 (1.2)	23.67 (1.6)

Analysis 1

For an initial examination by item, a 2 × 2 × 3 mixed-design analysis of covariance was conducted with phonemic representation and word class as between-item factors; grade level as a within-item factor; and frequency, letter position, and word length as covariants. We also introduced into the model the interaction between grade level and all of the other variables of the model.

The results presented in Table 1 demonstrate a significantly higher letter-omission rate among kindergarten students (30%), in comparison with the two other levels (21% and 23%), F (2, 46) = 3.61, $p < .04$, $\eta^2 = 0.14$. The post hoc tests specify that only the difference between kindergarten and the two other grades is significant (unequal N honestly significant difference [HSD]: $p < .0002$ for the other two grades).

This analysis of covariance yielded no effect for word length, F (1, 23) = 2.79, $p < .11$, for the phonetic presence of the sound [y], F (2, 23) = 0.01, *ns*, or for letter position, F (1, 23) = 0.21, *ns*. However, we observed a significant word class effect for the entire population, F (1, 23) = 19.73, $p < .001$, $\eta^2 = 0.46$, with an average omission rate of 29.4% for function words in comparison to 21.1% for content words and a significant word frequency effect, F (1, 23) = 11.26, $p < .003$, $\eta^2 = 0.33$, with an increase of omissions for more frequent words.

We calculated several essential interactions that shed light on the activation of these different elements throughout the students' progression from kindergarten to second grade. We initially noticed that interaction between the phonetic presence of [y] and grade level does not have an effect on the MLE, F (2, 46) = 0.19, *ns*. There was also a lack of effect with the interaction between grade level and word frequency, F (2, 46) = 2.04, $p < .15$, which illustrated the constant influence of frequency in the MLE. In addition, neither interaction between grade level and word length, F (2, 46) = 0.8, *ns*, nor interaction between grade level and letter position, F (2, 46) = 0.35, *ns*, were significant. The role of word class, however, evolves significantly throughout the cycle from kindergarten to second grade, F (2, 46) = 7.91, $p < .002$, $\eta^2 = 0.25$. The post hoc comparisons (Tukey HSD for unequal N) show that at the end of kindergarten the omission patterns in content words are similar to that in function word ($p < .99$) and that the omission

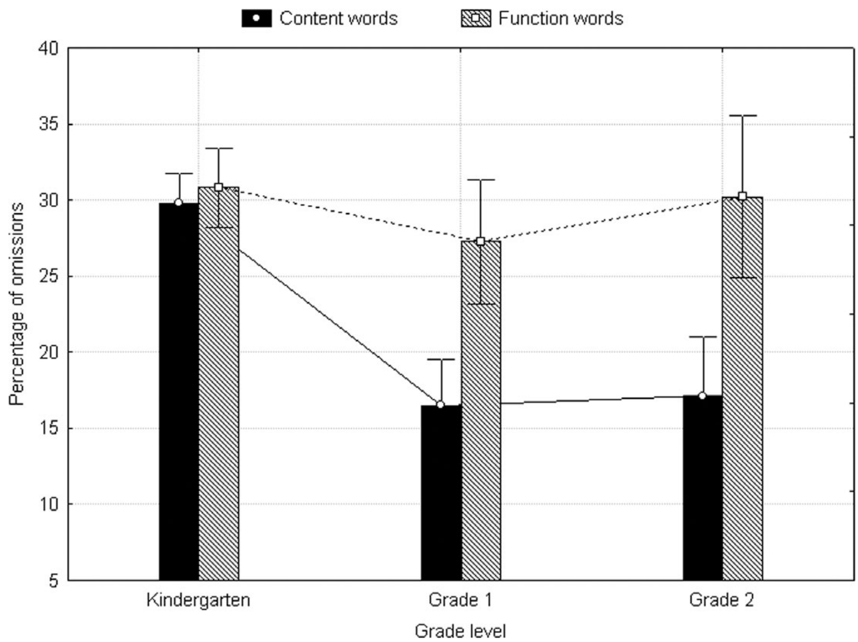

Figure 1. Mean percentages of omissions according to word class and grade level. The error bar represents the 95% confidence intervals.

patterns begin to differentiate significantly depending on word class starting from the students' initial exposure to writing (all ps $< .02$).

Analysis 2

To enhance our investigation, a 3×2 mixed-design analysis of variance was conducted with grade level as the between-subjects factor and the percentage of letter omissions in content and function words as the repeated factor. Corroborating the item analysis, the results show a word class effect, F (1, 883) = 240.76, $p <$ .0001, $\eta^2 = 0.21$, and a grade level effect, F (2, 883) = 28.97, $p < .0001$, $\eta^2 =$ 0.06. Finally, the interaction between grade level and word class is significant as well, F (2, 883) = 56.62, $p < .0001$, $\eta^2 = 0.11$.

Figure 1 illustrates the change in the role of word class throughout the 3 years of schooling. Note that the post hoc tests (Tukey HSD for unequal N) reveal that variation according to word class is insignificant in preschool. Significance emerges in the first grade ($p < .04$), only to become stronger by the end of the second grade ($p < .02$).

Analysis 3

These analyses allow us to have a clearer understanding of what is constructed throughout the progression from kindergarten to second grade. However, they do not allow us to evaluate the respective weight of the various factors introduced into the explanatory model. In order to find a response to this question, we conducted three (M1, M2, M3) multiple regression analyses (complete models), each one using percentages of letter omission in the three grade levels as a dependent variable. The independent variables remain the same as in the preceding analysis, with the exception of grade level. If Model M1 shows that the MLE is not significant in kindergarten (see Table 2), the portion of the variance explained by the independent variable (note that >60% of this phenomenon is explained by Models M2 and M3) and the multiple correlation coefficient ($R_{M2} = .82$ and $R_{M3} = .78$) justifies an in-depth investigation of the contribution of these variables to the MLE.

Table 2 presents the regression coefficients, standardized coefficients, *t* tests, and level of significance of these coefficients for the variables or modalities introduced into the explanatory models. It confirms the previous analyses. Both significant models present two highly correlated variables, if all other things are equal, in the MLE in words. The role of the first variable, the type of word in which the letter appears, is nonexistent in kindergarten ($\beta_{M1} = -0.13$, $p < .6$) but emerges robustly in the first grade ($\beta_{M2} = -0.5$, $p < .001$) and becomes an even stronger force in the second grade ($\beta_{M3} = -0.6$, $p < .001$). Word frequency follows a similar evolution because its significant contribution to the MLE emerges in first grade ($\beta_{M2} = 0.68$, $p < .002$) and continues into second grade ($\beta_{M3} = 0.47$, $p < .05$). This evolution can also be observed within the three sample groups in Figure 2, which presents the correlations and the partial correlations between log frequency and the percentage of omissions contingent on word type.

In terms of the respective weight of the two variables, we observe a crossing of their influence between the end of first grade and the end of second grade: although frequency is the most influential variable at the end of first grade ($\beta_{\text{M2-word class}} = -0.5$ vs. $\beta_{\text{M2-freq.}} = 0.68$; $z = -5.13$, $p < .001$),[4] word class is the most dominant variable by second grade with the greatest causal power in explaining letter omissions ($\beta_{\text{M3-word class}} = -0.6$ vs. $\beta_{\text{M3-freq.}} = 0.47$; $z = -4.02$; $p < .001$).

DISCUSSION

In an attempt to refine the reports from Saint-Aubin et al. (2005), the present study was designed to investigate elements necessary to help us determine the moment when students begin to omit letters in letter-detection tasks in a nonrandom manner. Our experimental plan differed from those of the majority of studies on the observation of the MLE in that we chose to use an authentic text from French children's literature. This approach required us to statistically control à posteriori certain factors that influence letter omission. If this decision resulted in the loss of a certain level of rigor in comparison with specifically constructed experimental material, it brought about a more authentic situation and quite likely more interesting texts, especially with younger participants, from which it is important to observe whether the obtained results follow the same patterns as more classic studies. Among the results of this study, we find a certain quantity of classic data from the letter-detection tasks corroborating the principal models used to explain this phenomenon.

In kindergarten, we observed that letter omission is equivalent in frequent function words and in more rare lexical words and that this omission is 30%

Table 2. *Explanatory models for percentages of omissions according to grade level*

	Models Explaining Missing-Letter Effect According to Grade		
	Kindergarten (M1)	Grade 1 (M2)	Grade 2 (M3)
Multiple correlation (R)	.49	.82	.78
Propor. explained variance (R^2)	.24	.68	.61
Significance test for R^2	.25	<.001	<.001

Variable	β	*B*	*t*	*p*	β	*B*	*t*	*p*	β	*B*	*t*	*p*
Word class (ref. = content)	−0.13	−0.50	−0.61	<.55	**−0.5**	**−5.38**	**−4.15**	**<.001**	**−0.6**	**−6.56**	**−3.75**	**<.001**
Letter pronunciation (ref. = no)	−0.12	−0.49	−0.64	<.53	−0.03	−0.23	−0.19	<.85	−0.04	−0.42	−0.54	<.6
Word frequency (log freq.)	0.49	0.73	1.59	<.13	**0.68**	**2.48**	**3.43**	**<.002**	**0.47**	**2.03**	**2.03**	**<.05**
Word length	0.19	0.27	0.63	<.54	0.38	1.34	1.97	<.1	0.21	0.85	0.72	<.4
Letter position in word	−0.24	−4.48	−1.11	<.28	−0.08	−3.41	−0.54	<.6	0.03	1.66	0.04	<.9
Intercept		**27.57**	**5.41**	**<.0001**		2.74	0.34	<.8		6.33	0.71	<.6

Note: All of the effects described in the table are "pure"; that is, all of the other variables are controlled. For example, in the coefficient column of the Grade 1 model (M2), we see that because the target letter is present in a content word (qualitative variable) it "contributes" to the decrease in the percentage of omissions from 5.38 points and that (quantitative variable) each additional letter in the word that contains the target letter increases the proportion of omissions from 1.34%. The values in bold indicate statistically significant coefficients of regression ($p < .05$).

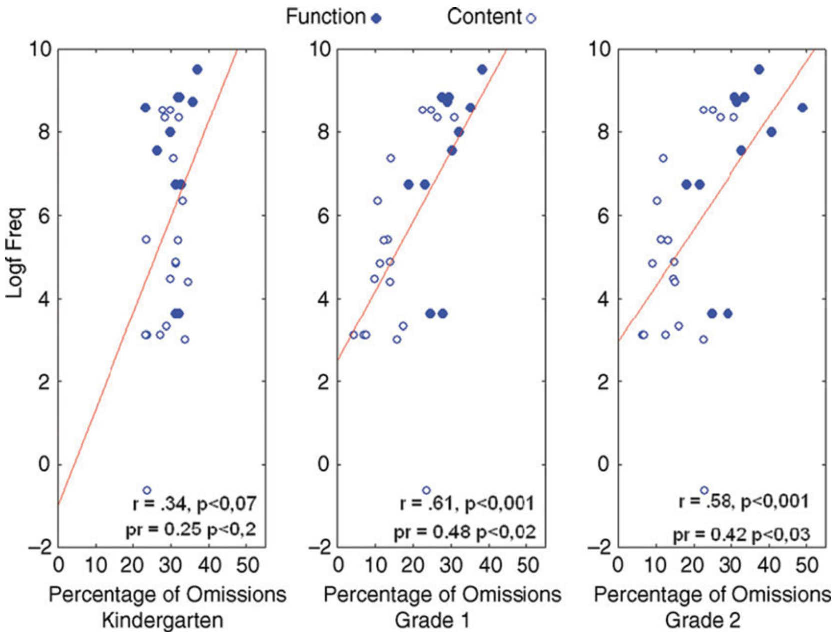

Figure 2. Mean percentages of omissions as a function of the log frequency. The partial correlations controlled for the influence of word class. [A color version of this figure can be viewed online at journals.cambridge.org/aps]

higher than in the two other grades. In contrast, starting at the end of first grade and increasingly important throughout second grade we observe a considerable presence of the MLE significantly associated to frequency and word class. These results agree with the hypotheses developed by the two most prominent models explaining the MLE. The GO model developed by Greenberg et al. (2004) as well as the AD model developed by Roy-Charland et al. (2007) predict both a word frequency and a word class effect. For the AD model, the omissions are the result of a decrease in attention paid to easily anticipated high-frequency words and functional words carrying a light semantic load (Roy-Charland et al., 2007). In terms of the GO model, it unites the two older models in that the structural model explains letter omission according to the role of a given word in the very rapid construction of a syntactic skeleton and the unitization account attributed letter omission to the effects of the lexical proximity between the reader and the syntagm (frequency and familiarity).

If the results of the present study do not allow for a clear choice between these two explanatory models, it does allow us to characterize the emergence of letter omission. In the first place, the results confirm the early emerging role of word frequency in terms of explaining the MLE because we see its effect emerge at the end of the first year of explicit reading instruction. In this sense, they point in the same direction as the overall results of work exploring the development of the

MLE (Cunningham et al., 1988; Drewnowski, 1978, 1981; Greenberg et al., 1998; Saint-Aubin et al., 2005).

We aimed to test the level of engagement of the structural model throughout the process of learning to read. This model claims that the reader conducts two operations while reading: the reader extracts the grammatical skeleton of the sentence, then fills it in with the meaning obtained from the semantic information of the individual words in the sentence. In this sense, the words responsible for organizing the structure of the sentence are relegated to a background position, and the reader's attention, as a consequence, is predominantly focused on the content words. The findings of this present study appear to support this model. Many of our results reveal that initiation to written language is accompanied by the emergence of simultaneous attention to the effects of the role that words play in a given sentence. Our first analysis (cf. Figures 1 and 2) shows that as soon as children begin to master reading techniques (i.e., at the end of first grade) the distribution of letter omissions is no longer random and the difference between function and content words becomes very sensitive. The most striking observation was the decrease of letter omissions in content words, compared to kindergarten students (Figure 1). This observation illustrates the greater availability of semantic information provided by these words as the eye moves across the sentence (Koriat & Greenberg, 1996). Consequently, this significant decrease in the percentage of missing letters in content words starting from the earliest stages of the reading learning process reflects the early development of a set of skills that young readers use to establish syntactic structural frames (skeletons) into which they integrate semantic information. In contrast, our third analysis reveals the developing importance of word class in the MLE, which exerts a less important influence than frequency in first grade and supplants it at the end of the second grade.

These results are noticeably different than those from previous studies. The data from Greenberg et al. (1998) show an effect of word class starting only in the fourth grade. This is similar to the results from Saint-Aubin et al. (2005) and Cunningham et al. (1998), who did not find a difference in omissions for the letter *a* before the third grade when it appeared as its own entity (as an article) or when it was embedded in familiar content words. Our results on the strong decrease in the percentage of letter omissions in content words between children in kindergarten and those in the first grade illustrate skills developed by young readers that enable them to establish syntactic structural frames onto which they integrate semantic information. Moreover, it is striking to note that from the end of the second grade it is word class that becomes more important, indicating that the capacity to construct the syntactic skeleton that is acquired very early seems to develop with age whereas the influence of frequency remains relatively constant in letter omission. What differentiates our results from previous work in the field is not the general profile of the factors influencing letter omissions; it is the shift in the time of emergence of syntactic sensitivity at the end of the first year of reading instruction. However, only one single prior developmental study used French as the support language. Nevertheless, our results indicate once again an earlier emergence of the role of word class in the MLE. One possible explanation for these results among young readers may lie in the differences in teaching approaches between the two school districts from which our populations come:

the young readers observed in Saint-Aubin et al. (2005) learned to read in New Brunswick, Canada, whereas ours were educated in the French system. There are at least two explanations for this situation. First, this variance could be attributable to differences in reading skill levels between the New Brunswick and the French students. International evaluation tools such as the PISA (OCDE, 200, 2001) show that the French-speaking children in New Brunswick are weaker in reading that their French homologues: the young French-speaking New Brunswickers obtained an average score of 478 (513 for the English speakers), whereas the young French students obtained an average of 505 points (general average = 500, standard deviation = 100). Second, given these conditions, there may be a significant difference in skill levels between the students tested by Saint-Aubin et al. (2005) and ours. In this case, our results, which were obtained with a different protocol and a larger sample population, tend to be more in line with those obtained by Saint-Aubin and Klein (2008) insofar as our students would be among the strongest whereas the young French-speaking New Brunswickers would be positioned on the side of the weakest readers. However, given the discrepancy between the age of our students and those tested by the PISA, this is not the only plausible explanation. This is why this simple difference raises questions about the role of teaching practices in the observation of reading skills among young children.

Future studies should explicitly take into consideration teaching methods and pedagogical practices in order to better understand how specific reading skills are developed, because, independent of observations at the moment when this syntactic skill can begin to be observed, the problem of its teaching and development remains unsolved. Grammar instruction, which starts more or less early according to individual curricula, is characterized by two opposing tendencies: one is largely based on syntactic criteria, whereas the other is based on uniquely semantic criteria. One can teach students to recognize a noun either because it can be preceded by a determiner or because it designates a thing. Furthermore, certain curriculums, or certain teachers, consider that the automatic recognition of the most frequent words is an important precondition for the mastery of reading skills; thus, they tend to favor explicit teaching of the most frequent words of the French language, of which many are function words. Articulation of these two elements is probably not possible without considering the relationship to the observations that we can make of the MLE within the youngest of children. For example, the data presented here imply that the ability to create a syntactic framework is developed from one's first contact with written language, but they do not indicate whether this development is simply the result of exposure to written language or whether it deserves to be supported by specific pedagogical interventions. If the answer to this question proves to be positive, it would need to be taken into account in the daily life of schools.

APPENDIX A

The letters in italic below were not included in the analysis because the letter *u* was capitalized or because it was in a word that begins or ends a sentence or a line. In addition, the first and last lines are not considered.

Les baisers d*u* lo*u*p

*U*n lo**u**p vo**u**lait se marier.

Il choisit la pl**u**s adorable, la pl**u**s blanche des oies d'**u**n po*u*lailler. Il l**ui** fit la co*u*r. Elle fit d'abord des manières mais accepta assez rapidement de le s**u**ivre po**u**r l'épo**u**ser, ne sachant pas ce q**ue** fait le lo**u**p a**u**x oies.

Après être passes devant le maire p**u**is à l'église d**u** village, les de**u**x mariés, impatients de se retro**u**ver se**u**ls, dansèrent pe**u** et mangèrent à peine, p**u**is s'engo**u**ffrèrent dans le*u*r no**u**velle petite voit**u**re électriq**ue** et filèrent dans les bois, droit a**u** repaire d**u** lo*u*p. *U*ne fois chez l**ui**, le lo**u**p, emporté par ses baisers, mangea l'oie j**u**sq**u**'à la dernière pl*u*me.

Évidemment.

Pas étonnant q*u*'on dise : bête comme *u*ne oie.

APPENDIX B

Frequency, word class, and omission rates by grade for the studied words

Word	Freq. (Log)	Word Class	Omission K	Omission 1	Omission 2
Loup	3.106	Content	23.79	6.93	6.44
Engouffrèrent	−0.616	Content	23.90	23.52	22.74
Loup	3.105	Content	23.50	4.38	12.58
Voulait	5.418	Content	23.66	13.28	11.29
Loup	3.105	Content	27.33	7.66	6.87
Plus	8.520	Content	27.93	22.55	22.58
Lui	8.349	Content	28.53	30.85	30.73
Électrique	3.314	Content	29.04	17.37	16.05
Lui	8.349	Content	32.21	26.42	27.17
Épouser	3.006	Content	33.91	15.84	22.58
Retrouver	4.830	Content	31.33	11.31	9.01
Nouvelle	4.874	Content	31.33	13.87	14.87
Voiture	5.399	Content	32.04	12.41	13.15
Seuls	4.472	Content	29.90	9.85	14.59
Plus	8.520	Content	29.90	24.74	25.15
Deux	7.351	Content	30.75	14.23	12.02
Suivre	4.384	Content	34.61	13.87	15.02
Peu	6.334	Content	33.32	10.58	10.30
Puis	6.729	Function	32.79	18.76	18.15
Jusqu'	3.624	Function	31.35	27.74	29.18
Puis	6.729	Function	31.35	23.14	21.59
Jusq**u'**	3.624	Function	32.21	24.45	24.89
Pour	8.732	Function	35.91	29.12	31.59
Aux	7.554	Function	26.48	30.22	32.88
Au	8.580	Function	23.48	35.04	48.93
Un	9.514	Function	37.08	38.25	37.60
Du	8.837	Function	32.04	29.34	30.90
Que	7.998	Function	29.90	32.12	40.77
Du	8.837	Function	32.46	27.52	33.61

ACKNOWLEDGMENTS
The authors thank Michael Zuniga for his assistance in this work and three reviewers for their comments on an earlier version of this paper.

NOTES
1. The present article uses the American equivalents of the French education system: kindergarten is the equivalent of "la grande section de maternelle," Grade 1 corresponds to "le cours préparatoire," and Grade 2 is "le cours élémentaire 1ère année."
2. Vous allez prendre le stylo et en parcourant le texte, en le lisant, chacun comme il peut, à sa vitesse, sans revenir en arrière, vous allez barrer la lettre u à chaque fois que vous la rencontrez.
3. Various analyses were conducted by calculating word frequency with the database NOVLEX (Lambert & Chesnet, 2001), a tool used to estimate the extensiveness and lexical frequency of the written vocabulary given to French elementary school students. However, we decided to keep the FRANTEX database for three reasons: the correlation between the two databases is very significant ($r = .97, p < .0001$), the results obtained with the two databases are identical, and the nature of its content is more considerably extensive (31 million items vs. 417,000).
4. The z score is obtained using the formula

$$z = \frac{\beta_{i1} - \beta_{i2}}{SE^2_{\beta i1} + SE^2_{\beta i2}},$$

as recommended by Cohen and Cohen (1983, p. 111) to "test the null hypothesis that two independent Betas (i.e., coming from different samples 1 and 2) are equal by using their respective standard errors."

REFERENCES
Baker, L., Fernandez-Fein, S., Scher, D., & Williams, H. (1998). Home experiences related to the development of word recognition. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 263–386). London: Erlbaum.

Bus, A. G., van IJzendoorn, M. H., & Pellegrini, A. D. (1995). Joint book reading makes for success in learning to read: A meta-analysis on intergenerational transmission of literacy. *Review of Educational Research*, *65*, 1–21.

Carpenter, P. A., Just, M. A., & Rayner, K. (1983). What your eyes do while your mind is reading. In K. Raynor (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 275–307). New York: Academic Press.

Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Corcoran, D. W. (1966). An acoustic factor in letter cancellation. *Nature, 210*, 658.

Cunningham, T. F., Healy, A. F., Kanengiser, N., Chizzick, L., & Willitts, R. L. (1988). Investigating the boundaries of reading units across ages and reading levels. *Journal of Experimental Child Psychology, 45*, 175–208.

Drewnowski, A. (1978). Detection errors on the word the: Evidence for the acquisition of reading levels. *Memory & Cognition, 6*, 403–409.

Drewnowski, A. (1981). Missing *-ing* in reading: Developmental changes in readings units. *Journal of Experimental Child Psychology, 31*, 154–168.

Drewnowski, A., & Healy, A. F. (1982). Phonetic factors in letter detection: A reevaluation. *Memory & Cognition, 10*, 145–154.

Foucambert, D. (2008). Vision parafovéale et extraction précoce de la structure des phrases durant la lecture. *Psychologie Française, 53*, 259–277.

Greenberg, S., Healy, A. F., Koriat, A., & Kreiner, H. (2004). The GO model: A reconsideration of the role of structural units in guiding and organizing text on line. *Psychonomic Bulletin and Review, 11*, 428–433.

Greenberg, S. N., Koriat, A., & Shapiro, A. (1992). The effects of syntactic structure on letter detection in adjacent function words. *Memory & Cognition, 20*, 663–670.

Greenberg, S. N., Koriat, A., & Vellutino, F. R. (1998). Age changes in the missing-letter effect reflect the reader's growing ability to extract the structure from text. *Journal of Experimental Child Psychology, 69*, 175–198.

Gross, J., Treiman, R., & Inman, J. (2000). The role of phonology in a letter detection task. *Memory & Cognition, 28*, 349–357.

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*, 235–242.

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

Healy, A. F., & Drewnowski, A. (1983). Investigating the boundaries of reading units: Letter detection in misspelled words. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 413–426.

Healy, A. F., Oliver, W. L., & McNamara, T. P. (1987). Detecting letters in continuous text: Effects of display size. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 279–290.

Kennedy, A. (2000). Parafoveal processing in word recognition. *Quarterly Journal of Experimental Psychology, 53*, 429–455.

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

Koriat, A., & Greenberg, S. N. (1993). Prominence of leading factors in function morpheme sequences as evidenced by letter detection. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 34–59.

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

Koriat, A., & Greenberg, S. N. (1996). The enhancement effect in letter detection: Further evidence for the structural model of reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1184–1195.

Koriat, A., Greenberg, S. N., & Goldshmid, Y. (1991). The missing-letter effect in Hebrew: Word frequency or word function? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 66–80.

Lambert, E., & Chesnet, D. (2001). Novlex: Une base de données lexicales pour les élèves de primaire. *L'Année Psychologique, 101*, 277–288.

Nazir, T. A., Jacobs, A. M., & O'Regan, J. K. (1998). Letter legibility and visual word recognition. *Memory & Cognition, 26*, 810–821.

New, B., Pallier, C., Ferrand, L., & Matos, R. (2001). Une base de données lexicales du français contemporain sur internet: LEXIQUE™. *L'Année Psychologique, 101*, 447–462.

OCDE. (2000). *Mesurer les connaissances et les compétences des élèves: Lecture, mathématiques et science: L'évaluation de PISA 2000*. Paris: Organisation de coopération et de développement économiques.

OCDE. (2001). *Connaissances et compétences: des atouts pour la vie. Premiers résultats du programme international de l'OCDE pour le suivi des acquis des élèves (PISA) 2000*. Paris: Les éditions de l'OCDE.

O'Regan, J. K. (1979). Saccade size control in reading: Evidence for the linguistic control hypothesis. *Perception and Psychophysics, 25*, 501–509.

O'Regan, J. K., & Jacobs, A. M. (1992). Optimal viewing position effect in word recognition: A challenge to current theory. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 185–197.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422.

Rayner, K., Pollatsek, A., & Coltheart, M. (1987). Eye movements in reading: A tutorial review. In M. Coltheart (Ed.), *Attention and performance: Vol. 12. The psychology of reading* (pp. 327–362). Hillsdale, NJ: Erlbaum.

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. *Quarterly Journal of Experimental Psychology, 59*, 25–45.

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 and Psychophysics, 69*, 324–337.

Saint-Aubin, J., & Klein, K. (2008). The influence of reading skills on the missing-letter effect among elementary school students. *Reading Research Quarterly, 43*, 132–146.

Saint-Aubin, J., Klein, R. M., & Landry, T. (2005). Age changes in the missing-letter effect revisited. *Journal of Experimental Child Psychology, 91*, 158–182.

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

Solotareff, G. (2001). *Contes d'été*. Paris: Ecole des Loisirs.

Vitu, F. (1991). The influence of parafoveal processing in linguistic context on the optimal landing position effect. *Perception and Psychophysics, 50*, 58–75.

Vitu, F., O'Regan, J. K., Inhoff, A. W., & Topolski, R. (1995). Mindless reading: Eye-movement characteristics are similar in scanning letter strings and reading texts. *Perception and Psychophysics, 57*, 352–364.

White, S. J., Rayner, K., & Liversedge, S. P. (2005). The influence of parafoveal word length and contextual constraint on fixation durations and word skipping in reading. *Psychonomic Bulletin and Review, 12*, 466–471.