Read Writ (2007) 20:815–831
DOI 10.1007/s11145-006-9049-3

Unconventional word segmentation in Brazilian children's early text production

Jane Correa · Julie E. Dockrell

Received: 13 July 2006 / Accepted: 20 December 2006 / Published online: 11 February 2007

Abstract An important element of learning to read and write at school is the ability to define word boundaries. Defining word boundaries in text writing is not a straightforward task even for children who have mastered graphophonemic correspondences. In children's writing, unconventional word segmentation has been observed across a range of languages and contexts with more occurrences of hyposegmentation (failure to separate two or more written words with a space) than hypersegmentation (written words are split into more than one segment). However, it is still unclear how frequent these errors are and the relationships of these written error patterns to the child's development in oral language, spelling and reading remains relatively unexplored. To address these issues, unconventional written lexical segmentations in Brazilian Portuguese children's text production during their first years at primary school (Year 1 to Year 3) were examined in relation to different cognitive and linguistic measures and patterns of spelling errors. Results reveal that in Portuguese the establishment of word boundaries in written text is not explained by visuospatial skills or limitations in processing resources (working memory). In contrast higher occurrences of hyposegmentation patterns were associated with lower levels of reading, vocabulary, verbal ability and morphological awareness whereas hypersegmentations were rarer and related to lower levels of reading and morphological awareness and typically only occurred in the initial stages of learning to write (Year 1). Occurrences of hyposegmentations as well as hypersegmentations were also related to spelling errors which reflected children's poor phonological skills.

J. Correa (✉)
Experimental Psychology, Federal University of Rio de Janeiro, Av Canal de Marapendi 1400 BL 1, AP 902 – Barra da Tijuca, Rio de Janeiro, RJ 22631-050, Brazil
e-mail: janecorrea@rjnet.com.br

J. E. Dockrell
Institute of Education, University of London, 25 Woburn Square, London WC1H 0AA, UK
e-mail: j.dockrell@ioe.ac.uk

Keywords Unconventional word segmentation · Hyposegmentation · Hypersegmentation · Early text production

Introduction

The ability to produce written text is a central component of educational achievement. Prior to school entry children are exposed to written notations and create their own notations (Tolchinsky, 2003). These early attempts to 'write' provide a basic structure for later written productions. Yet the ability to produce written text is an extended process. Text writing is a complex activity involving integration of many different skills (Swanson & Berninger, 1994). In the early stages writing can be broken down into two key constituent parts: transcription and text generation (Berninger et al., 2002). Transcription skills include awareness of orthographic rules and conventions in writing. At a basic level, it is important for children to realize that, unlike oral language, written text requires that words be separated either by blank spaces or punctuation. The current study examines the developmental changes in the establishment of word boundaries in written text in the early phases of formal schooling for Portuguese speaking children. Specifically we consider whether children whose texts contain unconventional word segmentations are characterised by different cognitive and linguistic profiles.

Separating words by blank spaces in writing is regarded as natural by literates; it is not, however, obvious from linguistic, historical or ontogenetic views (Ferreiro, 1999). The major development towards current textual organization in terms of paragraphs, sentences and written word boundaries occurred between 400 and 1400 AD (Günther, 1997). Historically the use of blank spaces or punctuation to separate words in texts was introduced to facilitate the processes of reading and copying (Günther, 1997). However, what is separated by blank spaces is language dependent; in Inuit (Eskimo language), for example, only proper names are segmented in a sentence whereas in Hebrew prepositions appear in a sentence attached to the content words as bound morphemes (Tolchinsky, 2006). There is, thus, no clear-cut definition for a written word (Ferreiro, 1999). Word segmentation is not required in the production of oral text therefore oral language experience does not support the identification of written word boundaries. Neither can the number of constituent phonemes or syllables define word boundaries. In contrast suprasegmental features and morphological classes inform the identification of word boundaries.

From an ontogenetic perspective, research on written language acquisition indicates that young children's early spontaneous writings take the form of continuous writing (*scriptio continua*) with no internal segmentation or further textual organization (Ferreiro & Teberosky, 1982). Thus an important element of learning to read and write at school is the ability to define word boundaries. Given the complexity of the process it is expected that word segmentation errors will be found in children's early texts. In the transition from *scriptio continua* to conventional organization of text, children initially segment words which are semantically salient such as nouns, adjectives and verbs (Tolchinksy, 2003). Two types of errors are possible: hyposegmentation and hypersegmentation. Hyposegmentation occurs when there is a failure to separate two or more words with a space. While hypersegmentation occurs when words are split into more than one segment, that is, a blank space is imposed within a conventional word. Based on preschoolers' spontaneous texts, which in many cases take the format of continuous writing (Ferreiro & Teberosky, 1982), hypersegmentations are not expected to occur as frequently as hyposegmentations.

Even for children who have mastered graphophonemic correspondences defining word boundaries in text writing is not a straightforward task. Unconventional word segmentation has been noted across a range of languages and contexts (Ferreiro & Pontecorvo, 1996), with more occurrences of hyposegmentation than hypersegmentation in children's writing (Ferreiro & Pontecorvo, 1996; Tolchinsky & Cintas, 2001). When writing a well-known fairy tale, children in Spanish, Italian and Portuguese produced high percentages of unconventional word segmentation (95% for Mexican, 76% for Uruguayan, 58% for Italian and 77% for Brazilian children). The occurrence of unconventional segmentation in children's writing has also been reported in Hebrew (Sandbank, 2001) and in Mayan (Pellicer, 2004); with errors noted in English children's spontaneous writing although no systematic description of the children's written productions has occurred (Nunes, 1999). Children appear to have specific problems separating closed words, including prepositions, articles and conjunctions, from the nearest content word (noun, verb, or adjective) (Ferreiro & Pontecorvo, 1996; Tolchinsky & Cintas, 2001). When words are not semantically salient such as auxiliary verbs, clitics and prepositions in Spanish, separating words accurately can remain difficult in some cases until high school (Tolchinsky, 2003).

The genre of the text produced also influences unconventional word segmentation (Tolchinsky & Cintas, 2001) and the linguistic choices afforded by genres are different (Tolchinsky & Cintas, 2001). Unconventional word segmentations are more common in children's narratives than in descriptions. In descriptive texts children are required to portray beings, things or places focusing on their attributes and qualities. In contrast the production of narratives requires a greater production of events. As such descriptions which contain more nouns and adjectives are more likely to contain single units whereas the high proportion of verbs, adverbs, conjunctions and prepositions found in narratives is likely to result in a greater proportion of instances of hyposegmentation.

These segmentation errors suggest that there are morphemic features that are related to the ways in which children segment text. Recent research has highlighted the importance of children's understanding of morphology to the development of their spelling abilities for English (Bryant, Nunes, & Bindman, 2000; Treiman & Cassar, 1996) and more transparent orthographies such as Portuguese (Bryant & Nunes, 2004). As children develop a greater competence with morphological aspects of the written language, segmentation errors ought to decrease, although cross-sectional data for these errors are rarely reported.

Both vocabulary knowledge and reading may serve to support the development of accurate segmentation. Children's vocabulary has been associated with early literacy skills, especially with decoding (Strattman & Hodson, 2005) and lexical knowledge is likely to support the development of conventional word boundaries in their writing. Reading is also implied in spelling as spelling skills include, besides the ability to write words correctly, the ability to recognize whether a word was spelt correctly. Decoding strategies used in reading provide children with useful information for spelling (Marsh, Friedman, Welch, & Desberg, 1980). Children might take advantage of the spelling-to-sound correspondences used to decode words in reading to help them to write (Waters, Bruck, & Seidenberg, 1985). Increased decoding skills therefore provide children with the competence to deal with unstressed morphological units such as prepositions, articles and clitics. Vocabulary and reading skills are thus likely to influence the occurrence of hyposegmentation, hypersegmentation and the nature of the errors the children make independent of age. In their initial attempts to introduce blank spaces to mark word boundaries in written text, children are likely to be guided by their semantic knowledge, resulting in increased occurrences of hyposegmentation.

Limitations in written production may also reflect limitations in cognitive efficiency rather than language limitations per se. Thus any analysis of the word segmentation errors produced by children needs to control for the child's ability to deal with a range of task demands. Otherwise the occurrence of unconventional segmentation may be more parsimoniously explained by the limited cognitive resources available to the young writer (Torrance & Galbraith, 2005) than by language specific factors.

This study aims to address the factors that influence changes in unconventional written lexical segmentations in Portuguese children's text production during their first years at primary school in Brazil. We considered the role of verbal and non-verbal ability, morphological awareness and reading as factors in influencing segmentation patterns. We also explored the connection between single-word spelling and occurrences of unconventional word segmentation in children's texts. Although establishing word boundaries in writing is a complex process for typically developing children, persistent difficulties in mastering written words conventions can result in an obstacle which can impair further development of children's writing skills. Thus we examined whether children who present relatively more occurrences of either hyposegmentation or hypersegmentation in their texts possessed a different profile of cognitive and linguistic skills and spelling performance from children with fewer occurrences of segmentation errors in their writings.

Method

Participants

Seventy-six children participated in the study: 24 children in Year 1 (14 boys and 10 girls) mean age 6:10 (range 6:4–7:3); 28 children in Year 2 (13 boys and

15 girls) mean age 7:7 (range 6:11–8:02); 24 children in Year 3 (9 boys and 15 girls) mean age 8:7 (range 7:10–9:01). The children attended a primary school located in a low-middle class neighbourhood in Rio de Janeiro, Brazil. The children had no identified learning difficulties.

Tasks

Morphology task

The task was developed to measure children's awareness of morphemes. It took the form of an analogy task in which a word is to a second word as a third word is to a fourth word according to specific grammatical transformations. The task consisted of 16 items including derivational morphemes. To control for working memory effects, children were presented with a card containing the first pair of words written on a card's first line, followed by a third word and a blank space in the second line. The words were read to the children and they were asked to produce the correct form of a word to fill the blank space using the same grammatical transformation applied to the first pair of words. The reliability coefficient for the word analogy task was .78 (Cronbach's Alpha).

Reading test

This is a single word reading accuracy test (Stein, 1994) standardized for Brazilian children of school age. It consisted of a list of 68 words displayed according to their level of familiarity, number of syllables and syllabic pattern. The reliability coefficient for the reading test .98 (alpha coefficient).

Nonverbal ability

Raven's matrices scores were taken as measure of children's nonverbal ability. Children are asked to choose among four or six items the one which completes a certain pattern.

Verbal ability

The verbal scale on Wechsler Intelligence Scale for Children-III (WISC-III) was used to evaluate children's verbal ability. In general, the verbal scale measures verbal fluency, word knowledge, general factual knowledge, numerical reasoning in problem solving, short term auditory memory and language expression. It does not require reading or writing. The examiner gives the questions orally, and the child provides a verbal answer. It consisted of six subtests: information, similarities, arithmetic, vocabulary, comprehension and digit span.

Vocabulary

The vocabulary subtest on Wisc-III Verbal Scale was administered to examine children's lexical knowledge. The child is asked to give oral definitions of words.

Working memory

The digit span subtest on Wisc-III Verbal Scale was used to evaluate children's working memory. In the digit span test children were asked to repeat a dictated series of digits. Series begin with two digits and increase in length, with two trials at each length.

Procedure

The Year 2 and Year 3 children had taken a nonverbal ability test, a verbal ability test and a reading test two months after the beginning of their school year. Year 1 children, at the beginning of the school year, were reported by their teacher as being in a prealphabetic stage for both reading and writing except for one child who was described as being in a partially alphabetic stage. Therefore, Year 1 assessments occurred in the middle of the school year when, according to their teacher, the children would be expected to read and write at least single words or sentences with simple syntactical structure. The children of all grades were assessed individually by one of two researchers with the appropriate qualification to administer the tests. The tests were administered in two sessions of approximately 30 min each.

In the middle of the school year, all children completed the morphological awareness task. The task was administered individually and took 20 min to complete. Finally a sample of children's writing was collected at the end of the school year. This task was carried out in the classroom setting. The children were asked to write an original story.

Children's spelling error analysis

Children's misspellings were classified into four categories which accounted for both the use of phonological and morphological information in spelling.

String of letters

Children wrote a series of letters which did not represent a recognizable word in Portuguese.

Phonologically acceptable misspellings

Children represented a phoneme by corresponding plausible similar sounding letter string in Portuguese such as *soologico* instead for *zoológico* [zoo]—as the letters can be a legal representation for the phoneme /z/ in some contexts.

Phonologically inaccurate misspellings

This category included errors of

(a) illegal letter representation—children represented a phoneme with a letter or group of letter that were not legal representations of that sound in any context in Portuguese such as *priga* instead of *briga* [fight];
(b) letter omissions—children omitted the representation of a phoneme which is present in speech as for *lina* instead of *linda* [beautiful]. Errors related to the omission of the nasality marks were also included in this category as for *bricou* instead of *brincou* [played];
(c) illegal letter order—children represented phonemes in the wrong order as for *crotou* instead of *cortou* [cut—simple past].

Morphological errors

These errors comprised omission (*mora* instead of *morar* [to live]), addition (*uma acidente* instead of *um acidente* [an accident] or substitution (*viverão* – simple future instead of *viveram*—simple past [lived]) of bound morphemes.

There was a high level of agreement between the two judges in their classification of the children's misspellings (86%, Kappa Coefficient).

Results

Data from four 6-year olds regarded as prealphabetic were disregarded from the analyses. The final corpus of data consisted of 72 handwritten texts. Children's texts varied significantly in terms of the numbers of words written. Year 3 children produced texts that were significantly longer than Year 2 children (Year 3 $M = 153.71$, SD = 97.19; Year 2 $M = 90.11$, SD = 50.51; Kolmogorov–Smirnov test, $Z = 1.50, p = 0.01$) and Year 2 produced texts that were significantly longer than children in Year 1 ($M = 25.95$, SD = 24.57; Kolmogorov–Smirnov test, $Z = 2.5, p < 0.001$).

We examined texts and identified children whose texts included examples of either hyposegmentation or hypersegmentation. As shown in Table 1 hyposegmentation was the main form of unconventional word segmentation for Year 2 and Year 3 but not for Year 1. To control for the number of words produced by the children, occurrences of unconventional word segmentations in children's text were expressed in terms of proportions according to the

Table 1 Percentages of different types of unconventional written word segmentation by grade

Grade	Hyposegmentation	Hypersegmentation
1	53	47
2	76	24
3	80	20

number of words written. These scores were then standardized (transformed into z-scores) to permit further comparisons and statistical analysis.

Overall, the occurrences of unconventional word segmentations in children's writing decreased significantly between Year 1 (M = .55; SD = 1.38) and Year 2 (M = –.10, SD = .95; Kolmogorov–Smirnov test, Z = 1.3, p = 0.03, r = .19) but, unexpectedly, not between Year 2 and Year 3 (M = –34, SD = .25; Kolmogorov–Smirnov test, Z = .71, p = .35, r = .09).

Unconventional word segmentations and children's cognitive skills

The children's scores on cognitive and linguistic measures are presented in Table 2. We examined the extent to which children's linguistic skills (reading, morphological awareness and verbal ability) and nonverbal ability discriminated children who produced unconventional word segmentations.

Children's scores were standardized to allow for comparison across the different tasks. Given the differences in their nature and frequency of occurrence in children's writing we ran two separate Cluster Analyses, one for hyposegmentation and one for hypersegmentation. Table 3 presents the mean z-scores and standard deviations for each variable in the analysis.

Table 2 Means and standard deviations for children's linguistic and nonverbal skills by grade

	Grade					
	1		2		3	
	(n^{a} = 21)		(n^{a} = 28)		(n^{a} = 23)	
	M	SD	*M*	SD	*M*	SD
Nonverbal ability	79.88	18.98	84.18	10.96	82.25	17.38
Verbal ability	114.61	20.39	119.39	17.60	120.92	14.52
Morphology	4.90	2.28	6.96	3.21	8.83	2.96
Reading	37.48	23.46	58.64	7.18	62.22	4.53

[a] Number of children in each group who completed all tests

Table 3 Standardized scores for occurrences of unconventional word segmentation and children's linguistic and nonverbal skills

	Grade					
	1		2		3	
	(n = 20)		(n = 28)		(n = 24)	
	M	SD	*M*	SD	*M*	SD
Hyposegmentation	0.31	1.23	–0.02	1.17	–0.24	0.32
Hypersegmentation	0.60	1.73	–0.17	0.36	–0.30	0.11
Nonverbal ability	0.05	1.01	0.04	0.78	–0.09	1.23
Verbal ability	–0.08	1.07	–0.01	1.07	0.08	0.88
Morphology	–0.62	0.74	–0.04	1.00	0.53	0.92
Reading	–0.89	1.41	0.23	0.52	0.49	0.33

Cluster analysis for occurrences of hyposegmentations in children's texts

A two-step cluster analysis for hyposegmentation produced two clusters which accounted for 41% and 59% of children, respectively. Table 4 describes the Cluster Profiles, displaying standardized scores (mean z-scores and standard deviations) for all the aforementioned variables in the analysis. Occurrences of hyposegmentations differentiated the two clusters; children in Cluster 1 (high hyposegmentation) produced more occurrences of hyposegmentation in their writing whereas Cluster 2 (low hyposegmentation) included children with fewer occurrences of hyposegmentations in their writing (Kolmogorov–Smirnov test, $Z = 1.74$, $p < .01$, $r = .20$).

Inspection of means in Table 4 shows that children in Cluster 1 (high hyposegmentation) scored relatively lower than children in Cluster 2 (low hyposegmentation) for all the cognitive measures. Children in Cluster 2 (low hyposegmentation) performed statistically significantly better than Cluster 1 (high hyposegmentation) in terms of reading (Kolmogorov–Smirnov test, $Z = 2.91$, $p < .001$, $r = .35$), the morphology test. (Kolmogorov–Smirnov test, $Z = 2.39$, $p < .001$, $r = .29$) and verbal ability scores (Kolmogorov–Smirnov test, $Z = 2.51$, $p < .001$, $r = .30$). However, there were no significant differences between the two clusters in nonverbal ability (Kolmogorov–Smirnov test, $Z = .95$, $p = .32$, $r = .11$). We also tested the statistical significance of the difference between the two clusters in vocabulary, working memory and the proportion of hypersegmentations in children's writing. Children in Cluster 2 (low hyposegmentation) achieved significantly higher scores in vocabulary (Cluster 1 $M = -.53$, SD = 1.08; Cluster 2 $M = .29$, SD = .79, Kolmogorov–Smirnov test, $Z = 1.48$, $p = .01$, $r = .18$) and produced significantly fewer occurrences of hypersegmentations (Kolmogorov–Smirnov test, $Z = 1.35$, $p = .05$, $r = .16$). In contrast no significant difference between the two clusters was found in working memory (Cluster 1 $M = -.26$, SD = 1.03; Cluster 2 $M = .06$, SD = .82, Kolmogorov–Smirnov test, $Z = .44$, $p = .99$, $r = .05$).

Table 4 Cluster profiles for hyposegmentation occurrences (standardized scores)

	Cluster			
	1		2	
	High Hyposegmentation		Low Hyposegmentation	
	(n = 28)		(n = 41)	
	M	SD	M	SD
Hyposegmentation	0.46	1.48	–0.29	0.25
Nonverbal ability	–0.20	0.89	0.06	1.06
Verbal ability	–0.75	0.84	0.44	0.76
Morphology	–0.72	0.66	0.49	0.89
Reading	–0.82	1.15	0.53	0.30

Table 5 presents the distribution of children in the two clusters by school grade. There was a close association between grade and occurrences of hyposegmentation in children's writing (χ^2 (2) = 19.03, p < 001, r = .54). The majority of children in Year 1 were in Cluster 1 (high hyposegmentation), children in Year 2 were distributed evenly between the two clusters and children in Year 3 were predominantly in Cluster 2 (low hyposegmentation).

Cluster analysis for occurrences of hypersegmentations in children's texts

Two clusters, including 83% and 17% of the children, respectively, were derived from the Two-step Cluster Analysis carried out on the proportion of hypersegmentations in children's texts. Table 6 displays the variables included in the analysis along with their means and standard deviations for each cluster. The first cluster accounted for relatively low frequencies of hypersegmentations per text. Children in Cluster 1 (low hypersegmentation) had better scores on the reading test (Kolmogorov–Smirnov test, Z = 2.82, p < .001, r = .34) and the morphology test (Kolmogorov–Smirnov test, Z = 2.17, p < .001, r = .26) than in Cluster 2 (high hypersegmentation). However, no statistical significant difference was found between the clusters in verbal

Table 5 Distribution of children according to the occurrences of unconventional word segmentations in their text according to grade (%)

	Grade		
	1	2	3
Hyposegmentation			
Cluster 1: High hyposegmentation	78	43	9
Cluster 2: Low hyposegmentation	22	57	91
Hypersegmentation			
Cluster 1: Low Hypersegmentation	50	89	100
Cluster 2: High Hypersegmentation	50	11	0

Table 6 Cluster profiles for hypersegmentation occurrences (standardized scores)

	Clusters			
	1		2	
	Low hypersegmentation		High hypersegmentation	
	(n = 57)		(n = 12)	
	M	SD	*M*	SD
Hypersegmentation	–0.27	0.18	1.38	1.95
Nonverbal Ability	–0.02	1.03	–0.17	0.88
Verbal ability	0.05	0.97	–0.48	0.98
Morphology	0.24	0.90	–1.14	0.57
Reading	0.33	0.52	–1.64	1.21

ability (Kolmogorov–Smirnov test, $Z = .93$, $p = .36$, $r = .11$) and nonverbal ability scores (Kolmogorov–Smirnov test, $Z = .80$, $p = .54$, $r = .09$). We also tested the statistical significance of the difference between the two clusters in vocabulary, working memory and the proportion of hyposegmentations in children's writing. There were no statistical significant difference between Cluster 1 (low hypersegmentation) and Cluster 2 (high hypersegmentation) in vocabulary (Clusters 1 $M = -.03$, SD = 1.02; Cluster 2 (M = –.09, SD = .91; Kolmogorov–Smirnov test, $Z = .43$, $p = .50$, $r = .05$), in working memory (Cluster 1 M = –.04, SD = .92; Cluster 2 M = –.20, SD = .97, Kolmogorov–Smirnov test, $Z = .41$, $p = .99$, $r = .05$) or in the occurrences of hyposegmentations (Kolmogorov–Smirnov test, $Z = 1.09$, $p = .18$, $r = .13$).

The distribution of children in both clusters by grade was examined (Table 5). As the table shows there is a close association between grade and occurrences of hypergmentation in children's writing ($\chi^2\,(2) = 19.03, p < .001$, $r = .52$). While children in Year 1 were evenly distributed between the two clusters, the majority of children in Year 2 and all children in Year 3 were in Cluster 1 (low hypersegmentation).

Children's spelling errors

The nature of the children's spelling errors was examined across the unconventional segmentation clusters. Overall the proportion of misspellings in the children's writing decreased according to grade. As expected children in Year 1 ($M = .48$, SD = .35) produced more spelling errors than did children in Year 2 ($M = .22$, SD = .19; Kolmogorov–Smirnov test, $Z = 1.76$, $p = .004$, $r = .21$). Similarly, children in Year 2 produced more spelling errors in their texts than children in Year 3 ($M = .08$; SD = .07; Kolmogorov–Smirnov test, $Z = 2.16$, $p < 0.01$, $r = .25$).

We examined whether children with more occurrences of unconventional word segmentation in their writing produced a distinctive pattern of spelling errors. Children's spelling errors in each category were expressed in terms of proportions according to the total number of misspellings produced per child. Table 7 presents the mean proportion of the spelling errors in each category according to schooling.

Phonologically acceptable errors were the predominant type of spelling error across all grades; these errors increased between Year 1 and Year 2 and showed a slow decrease between Year 2 and Year 3. This reflects the increase in children's use of more complex syllabic pattern than the CVCV pattern in Year 2. The use of illegal letters or the writing of illegal letter strings occurred mainly in Year 1 and decreased with schooling. By contrast, the inverse process occurred with the bound morpheme omission. Older children wrote more complex sentences which required more complex grammatical features such as agreement rules and thus provided greater scope for errors. In addition some morpheme omission errors reflected the difference between spoken and written language, where bound morphemes related to past and infinitive verbs which are not pronounced orally but required in writing. The relative

Table 7 Mean proportion of word spelling errors per words written by grade

	Grade					
	1		2		3	
	M	SD	*M*	SD	*M*	SD
Letter string	0.13	0.24	0.01	0.02	0.00	0.00
Phonologically acceptable	0.42	0.31	0.62	0.21	0.59	0.29
Illegal letter	0.19	0.22	0.15	0.15	0.11	0.12
Letter omission	0.13	0.16	0.08	0.09	0.09	0.11
Illegal letter order	0.04	0.10	0.01	0.04	0.00	0.00
Morpheme omission	0.08	0.11	0.11	0.12	0.14	0.18
Morpheme substitution	0.01	0.03	0.03	0.06	0.02	0.04
Morpheme addition	0.00	0.01	0.00	0.00	0.05	0.20

frequency of letter omission, however, did not change across the three year groups and reflected children's attempts to spell words with more complex syllabic patterns (digraphs, nasal vowels, consonants clusters). Bound morpheme addition and substitution errors or producing the word letters in a wrong order rarely occurred in the children's texts.

The extent to which hyposegmentation and hypersegmentation were related to spelling errors was examined by contrasting patterns of errors across the clusters previously identified. These patterns are presented in Table 8.

Children who produced more hyposegmentation in their texts produced fewer occurrences of phonologically acceptable errors (Kolmogorov–Smirnov test, $Z = 1.47$, $p = .03$, $r = .18$), but a greater proportion of string of letters errors (Kolmogorov–Smirnov test, $Z = 1.36$, $p = .05$, $r = .16$) and illegal letter representation (Kolmogorov–Smirnov test, $Z = 1.72$, $p = .005$, $r = .21$). No statistically significant difference was observed between the two clusters in terms of letter omission (Kolmogorov–Smirnov test, $Z = 1.10$, $p = .18$, $r = .13$), illegal letter order errors (Kolmogorov–Smirnov test, $Z = .87$, $p = .43$, $r = .10$), the frequency of substitutions of morphemes (Kolmogorov–Smirnov test, $Z = .35$, $p = 1.00$, $r = .04$), addition of morphemes (Kolmogorov–Smirnov test, $Z = .30$, $p = 1.00$, $r = .04$), or omission of morphemes (Kolmogorov–Smirnov test, $Z = .91$, $p = .38$, $r = .11$).

Spelling error patterns for the two hypersegmentation clusters are presented in Table 9. Children with more hypersegmentations, that is children in Cluster 2, produced more string of letters (Kolmogorov–Smirnov test, $Z = 1.35$, $p = .05$, $r = .16$) and illegal letter representation (Kolmogorov–Smirnov test, $Z = 1.51$, $p = .02$, $r = .18$). In contrast, they presented relatively fewer phonologically acceptable errors (Kolmogorov–Smirnov test, $Z = 1.80$, $p = .003$, $r = .22$). Children in the two clusters did not differ in the production of illegal letter order errors (Kolmogorov–Smirnov test, $Z = 1.26$, $p = .08$, $r = .15$), in the occurrences of letter omission (Kolmogorov–Smirnov test, $Z = 1.08$, $p = .20$, $r = .13$), morpheme substitutions (Kolmogorov–Smirnov

Table 8 Mean proportion of spelling error by clusters for hyposegmentation occurrences

	Clusters			
	Cluster 1		Cluster 2	
	High hyposegmentation		Low hyposegmentation	
	(*n* = 28)		(*n* = 41)	
	M	SD	*M*	SD
Letter string	0.06	0.12	0.02	0.16
Phonologically acceptable	0.43	0.25	0.62	0.27
Illegal letter	0.23	0.19	0.10	0.12
Letter omission	0.13	0.14	0.08	0.10
Illegal letter order	0.03	0.09	0.00	0.00
Morpheme omission	0.09	0.10	0.13	0.17
Morpheme substitution	0.02	0.06	0.02	0.04
Morpheme addition	0.00	0.00	0.03	0.16

Table 9 Mean proportion of spelling errors by clusters for hypersegmentation occurrences

	Clusters			
	Cluster 1		Cluster 2	
	Low hypersegmentation		High hypersegmentation	
	(*n* = 57)		(*n* = 12)	
	M	SD	*M*	SD
Letter string	0.02	0.13	0.12	0.16
Phonologically acceptable	0.60	0.26	0.27	0.20
Illegal letter	0.13	0.15	0.29	0.18
Letter omission	0.09	0.10	0.17	0.18
Illegal letter order	0.00	0.02	0.06	0.12
Morpheme omission	0.13	0.15	0.05	0.08
Morpheme substitution	0.02	0.04	0.03	0.09
Morpheme addition	0.02	0.13	0.00	0.00

test, $Z = .26$, $p = 1.00$, $r = .03$), morpheme additions (Kolmogorov–Smirnov test, $Z = .17$, $p = 1.00$, $r = .02$) or omission of morphemes (Kolmogorov–Smirnov test, $Z = .86$, $p = .46$, $r = .10$).

Discussion

This study examined the developmental changes in the establishment of word boundaries in written text in the early phases of formal schooling. Specifically we examined whether children who presented more unconventional word segmentations in their writing are characterised by different cognitive and linguistic profiles and patterns of spelling errors.

Decreases in hyposegmentation and hypersegmentation occurred with age; however, these changes were not related to improved performance on measures of non-verbal ability or working memory. Thus, in Portuguese the establishment of word boundaries in written text is not explained by visuospatial skills or limitations in processing resources as measured by working memory. In contrast highly significant differences were evident on a range of linguistic and literacy measures. Moreover, occurrences of hyposegmentation and hypersegementation presented different patterns of associations. Although both hyposegmentation and hypersegmentation were related to lower levels of reading and morphological awareness occurrences of hypersegmentation were limited to the initial stages of education. Higher occurrences of hyposegmentation patterns were also associated with lower levels of vocabulary and verbal ability. These results suggest that advances across literacy and language contribute to an understanding of language specific word segmentation patterns. Although both hyposegmentation and hypersegmentation errors reflect conceptual dimensions of the language system, hypersegmentation appears to result from literacy experience while hyposegmentation reflects a broader language based dimension.

Our findings are also in line with those of other studies showing a link between reading and spelling (Ehri, 1997; Ellis, 1997). Reading printed words accurately and efficiently was related to children's orthographic representations. As reading skills developed so did the ability to segment unstressed morphemes in speech. In addition the knowledge of spelling-to-sound correspondences used for reading also underpinned children's strategies in spelling. Good spellers outnumbered bad spellers in the relative frequency of phonological misspellings (Waters et al., 1985). Similar patterns applied to children's ability to identify word boundaries in print. Occurrences of hyposegmentations as well as hypersegmentations were related to spelling errors which reflected children's poor phonological skills such as illegal letter use or letter omissions.

Although children's decoding skills play an important role in helping children to establish word boundaries in their texts, the ability to segment words in writing goes beyond the understanding of the alphabetic principle. Children's sensitivity to grammar, especially to morphology, played an important role in the development of the ability to segment written words. Word segmentation in writing is visually expressed by the insertion of blank spaces, the criteria used to segment words in print are based on grammatical features, specifically on morphological classes. To segment words conventionally in writing, children were guided by their sensitivity to morphology either in separating a function word from the nearest content word in a sentence or by maintaining the integrity of a word which started with a syllable spelled like a function word.

Based on the evidence provided by this study that there are two ways in which morphological awareness contributes to the development of children's orthographic knowledge. The first is supported by the empirical evidence indicating children's sensitivity to morphology helps them to spell words in which the regularity is based on morphology (Nunes, Bryant, & Bindman, 1997). The second one is that children's ability to establish word boundaries in writing depends on their morphological knowledge. These findings have not been documented in the literature to date; typically studies of spelling development focus only on the production of single words. The data obtained to spelling in continous text in this study provide empirical evidence for this specific contribution of morphological awareness to spelling.

Our findings also showed that vocabulary made an important contribution to children's understanding of word boundaries in writing. Vocabulary knowledge provides both phonological and morphological information. Children with larger vocabularies have more detailed phonological representations of words (Dockrell & Messer, 2004). The phonological representations support spelling in two different ways: (a) by enhancing children's phonological awareness and (b) by specifically providing children with the phonological form of a word to be written.

Vocabulary also served to foster spelling development via children's morphological skills (Dockrell & Messer, 2004). Young writers find less difficulty in segmenting words such as nouns, verbs and adjectives because these words have a salient meaning (Ferreiro & Pontecorvo, 1996; Tolchinsky & Cintas, 2001). In contrast it is harder for young children to conventionally segment function words in a sentence (Ferreiro, 1999). Function words have mainly a grammatical meaning; as such children cannot use their semantic knowledge to help them to segment words, such as prepositions or clitics, from the nearest content word in a sentence.

The general pattern of our results has important implications for a model of children's conception of word boundaries in writing. It provided empirical evidence for the existence of a strong link between oral and written language as well as between reading and writing skills in the development of children's ability to establish word boundaries in writing. The acquisition of the concept of word in print is a complex process which requires an integration of different cognitive skills and knowledge related to the development of children's linguistic abilities. Despite the consistency of our findings, we are aware of the limits of our interpretation. As the current study was carried out with Brazilian children, it could be argued that our findings are language specific or that they might only be generalized to other languages with linguistic features very similar to Portuguese. Further cross-linguist research is needed in order to widen our understanding of the development of children's conception of word in writing as well as of those aspects that can impair such development.

Acknowledgements The authors are indebted to the children and their schools who welcomed us and the Educational Authorities in Rio de Janeiro (SME-RJ). Support for this study was provided by MCT/CNPq—Ministry of Science and Technology/The National Council for Scientific and Technological Development and CAPES.

References

Berninger, V. W., Vaughan, K., Abbott, R. D., Begay, K., Coleman, K. B., Curtin, G., Hawkins, J. M., & Graham, S. (2002). Teaching spelling and composition alone and together: Implications for the simple view of writing. *Journal of Educational Psychology, 94*, 291–304.

Bryant, P., & Nunes, T. (2004). Morphology and spelling. In T. Nunes & P. Bryant (Eds.), *Handbook of children's literacy* (pp. 91–118). Dordrecht: Kluwer Academic.

Bryant, P., Nunes, T., & Bindman, M. (2000). The relations between children's linguistic awareness and spelling: The case of the apostrophe. *Reading and Writing, 12*, 253–276.

Dockrell, J. E., & Messer, D. (2004). Lexical acquisition in the early school years. In R. A. Berman (Ed.), *Language development across childhood and adolescence* (pp. 35–52). Amsterdam: John Benjamins.

Ehri, L. C. (1997). Learning to read and to spell are one and the same, almost. In C. A. Perfetti, L. Rieben, & M. Fayol (Eds.), *Learning to spell: Research, theory and practice across languages* (pp. 237–269). Mahwah, NJ: Lawrence Erlbaum.

Ellis, N. (1997). Interactions in the development of reading, spelling: Stages, strategies, and exchange of knowledge. In C. A. Perfetti, L. Rieben, & M. Fayol (Eds.), *Learning to spell: Research, theory and practice across languages* (pp. 271–294). Mahwah, NJ: Lawrence Erlbaum.

Ferreiro, E. (1999). Oral and written words: Are they the same units? In T. Nunes (Ed.), *Learning to read: An integrated view from research and practice* (pp. 3–14). Dordrecht: Kluwer Academic.

Ferreiro, E., & Pontecorvo, C. (1996). Os limites entre as palavras. [Word boundaries]. In E. Ferreiro, C. Pontecorvo, N. R. Moreira, & I. G. Hidalgo (Eds.), *Chapeuzinho vermelho aprende a escrever* [Little red-hood learns how to write] (pp. 38–66). São Paulo: Ática.

Ferreiro, E., & Teberosky, A. (1982). *Literacy before schooling*. London: Heinemann Educational.

Günther, H. (1997). Aspects of a history of written language processing: examples from the roman world and the early Middle Ages. In C. Pontecorvo (Ed.), *Writing development: An interdisciplinary view* (pp. 129–149). Amsterdam: John Benjamins.

Marsh, G., Friedman, M., Welch, V., & Desberg, P. (1980). The development of strategies in spelling. In U. Frith (Ed.), *Cognitive processes in spelling* (pp. 339–353). London: Academic Press.

Nunes, T. (1999). Section 1: Introductory comments. In T. Nunes (Ed.) *Learning to read: An integrated view from research and practice* (pp. 3–14). Dordrecht: Kluwer Academic.

Nunes, T., Bryant, P., & Bindman, M. (1997). Morphological spelling strategies: Developmental stages and processes. *Developmental Psychology, 33*, 637–649.

Pellicer, A. (2004). Segmentation in the writing of Mayan language statements by indigenous children with primary schooling. In T. Nunes & P. Bryant (Eds.), *Handbook of children's literacy* (pp. 721–739). Dordrecht: Kluwer Academic.

Sandbank, A. (2001). On the interplay of genre and writing conventions in early text writing. In L. Tolchinsky (Ed.), *Developmental aspects in learning to write* (pp. 55–75). Dordrecht: Kluwer Academic.

Stein, L. (1994). *Teste de desempenho escolar [School achievement test]*. São Paulo: Casa do Psicólogo.

Strattman, K., & Hodson, B. W. (2005). Variables that influence decoding and spelling in beginning readers. *Child Language Teaching and Therapy, 21*, 165–190.

Swanson, H. L., & Berninger, V. (1994). Working memory as a source of individual difference in children's writing. In E. Butterfield (Ed.), *Children's writing: Toward a process theory of development of skilled writing* (pp. 31–56). Greenwich, CT: Jal Press.

Tolchinsky, L. (2003). *The cradle of culture and what children know about writing and numbers before being taught*. Mahwah, N J: Lawrence: Erlbaum.

Tolchinsky, L. (2006). Writing and written numbers as source of knowledge. In E. Teubal, J. E. Dockrell, & L. Tolchinsky (Eds.), Notational knowledge: Historical and developmental perspectives. Rotterdam: Sense Publishers.

Tolchinsky, L., & Cintas, C. (2001). The development of graphic words in written Spanish: What can be learnt from counterexamples? In L. Tolchinsky (Ed.), *Developmental aspects in learning to write* (pp. 77–95). Dordrecht: Kluwer Academic. .

Torrance, M., & Galbraith, D. (2005). The processing demands of writing. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 67–79). New York: Guildford Publications.

Treiman, R., & Cassar, M. (1996). Effects of morphology on children's spelling of final consonant clusters. *Journal of Experimental Child Psychology, 63*, 141–170.

Waters, G. S., Bruck, M., & Seidenberg, M. (1985). Do children use similar processes to read and spell words? *Journal of Experimental Child Psychology, 39*, 511–530.