

Segmentación en palabras de textos escritos por alumnos de primer curso de Educación Primaria: un estudio en el contexto español¹

Rafaela Gutiérrez Cáceres
Universidad de Almería (España)  

<https://dx.doi.org/10.5209/rced.86118>

Recibido: Enero 2023 • Evaluado: Abril 2023 • Aceptado: Mayo 2023

ES Resumen: INTRODUCCIÓN. Uno de los aspectos básicos para la adquisición de una adecuada competencia comunicativa escrita en Educación Primaria es el conocimiento y uso de la ortografía y, en especial, la segmentación de palabras. Así, en el marco del proyecto de investigación Scolinter realizado en colaboración con la Universidad Grenoble Alpes y la Universidad Milán Bicocca, se ha desarrollado este trabajo en la Universidad de Almería cuyo objetivo es analizar la segmentación de textos escritos por alumnos del primer curso de Educación Primaria procedentes del contexto español. MÉTODO. Este estudio se ha realizado según el siguiente procedimiento metodológico: recogida de textos redactados a partir de la visualización de cuatro imágenes de una secuencia narrativa; diseño del corpus Scolinter constituido por textos escritos a partir de la recopilación, transcripción, anotación y tratamiento automático del lenguaje llevado a cabo a través de la herramienta informática Aliscol; descripción y análisis de los textos en el contexto español sobre la segmentación en palabras. RESULTADOS. Entre otros resultados más relevantes se ha encontrado que el 74.7% de los textos escritos presentan casos de segmentación no convencional. Según la tipología de estos, existe un predominio de casos de hiposegmentación en comparación con la hipersegmentación, siendo los más comunes aquellos en los que el primer formante involucrado es gramatical y el segundo un formante lexical. DISCUSIÓN. Se constata en concordancia con investigaciones previas que la hiposegmentación representa el tipo de segmentación no convencional más frecuente en los textos escritos por alumnos de primer curso de Educación Primaria. Este estudio ha contribuido no solo al avance de la investigación educativa desde una perspectiva psicogenética, sino también a la reflexión y mejora sobre la enseñanza de la escritura desde un enfoque orientado a la comprensión de las razones por las que el alumnado realiza la segmentación no convencional.

Palabras clave: Corpus lingüístico; expresión escrita; ortografía; segmentación de las palabras; educación primaria.

ENG Word segmentation in texts written by first year primary school children: a study in the Spanish context

Abstract: INTRODUCTION. One of the basic aspects for the acquisition of adequate written communicative competence in Primary Education is the knowledge and use of spelling and, in particular, word segmentation. Thus, within the framework of the Scolinter research project carried out in collaboration with the University of Grenoble Alpes and the University of Milan Bicocca, this study has been developed at the University of Almeria in order to analyse the segmentation of texts written by first-year students of Primary Education from the Spanish context. METHOD. This study has been carried out according to the following methodological procedure: collection of texts written from the visualization of four images of a narrative sequence; design of the Scolinter corpus made up of written texts from the collection, transcription, annotation and automatic treatment of the language using the computer tool Aliscol; description and analysis of texts in the Spanish context on word segmentation. RESULTS. Among the most relevant results, it was found that 74.7% of the written texts present cases of unconventional segmentation. According to the typology of these, there is a predominance of cases of hyposegmentation compared to hypersegmentation, the most common being those in which the first formant involved is grammatical and the second is a lexical formant. DISCUSSION. In agreement with previous research, we found hyposegmentation represents the most frequent type of

¹ Fuente de financiación: Universidad Grenoble Alpes. Campaña de estancias de científicos extranjeros, Grenoble INP- Año 2021 (Grenoble, Francia).

unconventional segmentation in texts written by first-year students of Primary Education. This study has contributed not only to the advancement of educational research from a psychogenetic perspective, but also to reflection and improvement on the teaching of writing from an approach aimed at understanding the reasons why students perform unconventional segmentation.

Keywords: Linguistic corpus; written expression; spelling; word segmentation; primary education.

Sumario: 1. Introducción. 2. Metodología. 3. Análisis y discusión de datos. 4. Conclusiones. 5. Referencias bibliográficas.

Cómo citar: Gutiérrez Cáceres, R. (2024). Segmentación en palabras de textos escritos por alumnos de primer curso de Educación Primaria: un estudio en el contexto español. *Revista Complutense de Educación*, 35(4), 685-695.

1. Introducción

Una de las finalidades principales en la etapa de Educación Primaria es desarrollar la competencia en comunicación lingüística y, en este caso, la capacidad de escribir textos (Ley Orgánica 3/2020, de 29 de diciembre). El dominio de la expresión escrita es clave en cuanto que desempeña una función epistémica y comunicativa en una sociedad donde se multiplican los intercambios a través del lenguaje escrito (Salvador Mata, 2008; Zayas, 2022). Uno de los aspectos básicos para la adquisición de una adecuada competencia comunicativa escrita en Educación Primaria es el conocimiento y uso de la ortografía y, en este caso, la segmentación de palabras. Por ello, con el fin de favorecer el aprendizaje de la ortografía es necesario que el alumno participe activamente en el proceso de construcción social de textos con propósitos claros y destinatarios reales (Nemirosky, 1999; Teberoksy y Tolchinsky, 1995; Teruggi, 2019; Tolchinsky, 2008; Zuccalá y Sandbank, 2018).

En la actualidad, la inclusión en el currículo sobre el desarrollo de las competencias ortográficas está adquiriendo cada vez más relevancia, dadas las dificultades que encuentran los alumnos al escribir textos, desde los inicios de su escolarización y que se extiende a lo largo de las siguientes etapas educativas (Fernández-Rufete, 2015). Por otra parte, Echauri (2000, p.1) señala que los problemas que plantea el dominio ortográfico “son una realidad omnipresente frente a la que el profesorado se siente a menudo impotente y sin saber muy bien cómo actuar”. Así pues, la demanda de una adecuada atención a las dificultades ortográficas en el alumnado de Educación Primaria, contrasta con la insuficiente formación del profesorado en esta temática (Camacho et al., 2021; Camps, 2006; Teruggi, 2019). De ahí, el anhelo de mejorar la formación docente y, para ello, se requiere llevar a cabo estudios sobre las competencias ortográficas del alumnado en la expresión escrita que redunden en el avance del conocimiento científico tanto a nivel teórico como práctico.

La investigación sobre las competencias ortográficas del alumnado de Educación Primaria se ha desarrollado desde una pluralidad de enfoques y líneas de investigación de diferentes disciplinas científicas. Entre otras, se encuentra la perspectiva psicológica clínica, desde la cual los estudios se orientan al análisis de dichas competencias tras la aplicación de pruebas estandarizadas como el dictado de palabras o de textos (Angelelli et al., 2008; Marinelli et al., 2016). Este tipo de estudios se focalizan en la identificación, descripción y clasificación de los errores ortográficos, con el fin de explorar posibles déficits a nivel fonológico o léxico (Joyce et al., 2022; Zadunaisky-Ehrlich et al., 2021).

A partir de la década de los 80 surgen otros enfoques de investigación centrados en el análisis de la evolución del proceso de adquisición de la lengua escrita (Ferreiro y Teberosky, 1981; Frith, 1985; Gosse et al., 2021; Kemp y Treiman, 2023; Pinto et al., 2017; Salas y Caravolas, 2019). Concretamente, la investigación desde la perspectiva psicogenética (Ferreiro y Teberosky, 1985; Ferreiro, 2013, 2019a) ha puesto de manifiesto cómo el proceso de adquisición de la lengua escrita se concibe como un proceso interactivo de construcción del conocimiento que tiene lugar antes del inicio de la escolarización y, además, cómo el niño aprende a escribir gracias a la elaboración de hipótesis propias que se plantea para comprender el funcionamiento del sistema de escritura. Estas hipótesis siguen un proceso de conceptualización que culmina con el acceso a los principios del sistema alfabetico. Sin embargo, la correspondencia grafema - fonema que conduce a la adquisición de la escritura alfabetica, no puede considerarse como la última etapa de este proceso evolutivo pues los alumnos se enfrentan a otras cuestiones relacionadas con la convencionalidad ortográfica: la ortografía, la separación de palabras, el uso de mayúsculas, la puntuación, etc. En este sentido, si se pretende conocer cómo los alumnos adquieren las reglas de nuestro sistema de escritura, es necesario superar la dicotomía “correcto/incorrecto” y ampliar la mirada hacia la comprensión de las hipótesis constructivas que subyacen en los “errores” ortográficos (De Gaulmyn et al., 1996).

Indagar cómo los alumnos se apropián de las reglas del sistema ortográfico a través del análisis de datos obtenidos tras la producción espontánea de textos, trae consigo problemas metodológicos que no son fáciles de abordar (Borghi, 2014), especialmente si los sujetos participantes son alumnos de primer grado de Primaria que se encuentran en proceso de aprendizaje del sistema de escritura convencional (Wolfarth, 2019). De hecho, en el análisis de los textos escritos se han encontrado no solo palabras incorrectas desde el punto de vista ortográfico (uso incorrecto de acentos, inversión de fonemas, omisión de grafemas, etc.), sino también casos de segmentación no convencional (unión o separación indebida de palabras), aspectos que dificultan el análisis a través del tratamiento automático del lenguaje (“Natural Language Processing”).

Precisamente debido a estos problemas, las investigaciones, aun considerando el valor social de la ortografía (Colombo, 2011; Serianni y Benedetti, 2009), se enfocan a la descripción y análisis de competencias ortográficas tras la recopilación de textos escritos por alumnos que ya han alcanzado la convencionalidad de la escritura (Barbagli et. al., 2014; Cignetti et.al., 2016).

1.1. Segmentación de palabras en textos escritos por alumnos de Educación Primaria

Según los modelos teóricos existentes, el término “palabra” entendido como unidad mínima de la sintaxis queda reducido a la simple adquisición de las convenciones ortográficas. No obstante, es preciso distinguir entre la noción de palabra escrita y el concepto de palabra en el enunciado. Por ejemplo, en la expresión “Érase una vez” en su dimensión escrita se encuentran tres unidades gráficas claramente delimitadas, mientras que en su enunciación oral dicha separación de palabras no existe. En efecto, la escritura delimita la separación de palabras mediante el uso de espacios en blanco, cuando en el habla se refiere a un continuo sonoro. Por tanto, la escritura en cuanto actividad sociocultural, según la perspectiva psicogenética, es la que define el concepto de palabra como un conjunto de letras separadas por un espacio en blanco (Ferreiro et al., 1996).

Así pues, dada la problemática que trae consigo el concepto de palabra, es posible comprender las dudas que tienen los niños al inicio de su escolarización a la hora de separar las palabras (Ferreiro y Teberosky, 1985). Sobre todo, si se considera la palabra como una parte del texto que se construye a través de la escritura; además, la práctica de la segmentación de palabras puede resultar más compleja si durante el proceso de enseñanza de la lengua escrita los alumnos no reflexionan sobre qué es una palabra. Ferreiro y Vernon (1992) señalan cómo el concepto de palabra les resulta muy complicado para los niños prealfabetizados (prefiriendo el término “nombre”). En concreto, varios estudios realizados sobre la segmentación de palabras en diversas lenguas han puesto de manifiesto la dificultad para comprender que algunas categorías gramaticales como artículos, preposiciones o conjunciones pueden ser consideradas palabras (Ferreiro et al., 1996).

Es a partir de la década de los 90 cuando tiene lugar el impulso de estudios psicolingüísticos realizados, sobre todo, en el ámbito internacional con distintos corpus y a través de diseños de investigación comparativos y/o longitudinales con objeto de analizar las diferencias existentes en cuanto a la segmentación de las palabras en las diferentes escrituras alfábéticas, así como conocer la evolución de dicho fenómeno en función de la edad y la competencia lingüística del alumno (Alegria Ugarte y Díaz Argüero, 2017; Báez, 1999; Böhm y Mehlem, 2020; Correa y Dockrell, 2007; Fabbretti y Pontecorvo, 2005; Farina, 2022; Ferreiro y Pontecorvo, 1996; Piacente y Querejeta, 2012; Querejeta et al., 2013; Sparrow, 2014; Viramontes et al., 2021). A continuación, se muestran las principales contribuciones acerca de la segmentación en palabras de textos escritos en alumnado de Educación Primaria que se han encontrado en algunas investigaciones más relevantes.

De una parte, se ha realizado el análisis de las características y tipología de la segmentación no convencional en textos escritos, siendo la clasificación la siguiente (Báez, 1999; Correa y Dockrell, 2007): a) hiposegmentación: se refiere a la unión no convencional de palabras, como por ejemplo, “seva” (se va); b) hipersegmentación: consiste en la separación indebida de segmentos de una palabra utilizando el espacio en blanco, por ejemplo “en tonces” (entonces). Recientemente, en las últimas investigaciones (Querejeta et al., 2013) se ha explorado otro tipo de segmentación no convencional de carácter mixto, denominado hipohipersegmentación, donde tiene lugar tanto la hiposegmentación como la hipersegmentación como, por ejemplo, “lagati tamadre” (la gatita madre).

Entre otros resultados se puede destacar que, con independencia de la lengua, del tipo de letra utilizada y del método de enseñanza de la lengua escrita adoptado, en los textos escritos por alumnos de primer curso de Educación Primaria existe un claro predominio de los casos de hiposegmentación en comparación con la hipersegmentación (Fabbretti y Pontecorvo, 2005; Farina, 2022; Sparrow, 2014; Viramontes et al., 2021). Además, se ha encontrado que la hiposegmentación sigue siendo la categoría predominante entre los tipos de segmentación no convencional incluso en los niveles superiores de la escolarización (Alegria Ugarte y Díaz Argüero, 2017; Báez; 1999; Correa y Dockrell, 2007; Ferreiro y Pontecorvo, 1996).

De otra parte, se han llevado a cabo estudios cualitativos cuyo objetivo es indagar cuáles son los “segmentos” que se encuentran involucrados en los distintos casos de segmentación no convencional. Así, con el fin de conocer el tipo de segmentos que integran la segmentación no convencional, investigaciones recientes han adoptado el concepto de formante morfológico (Di Tullio, 2014), estableciendo dos tipologías: a) formante lexical: se refiere a palabras léxicas que poseen un significado propio (sustantivos, adjetivos, verbos, adverbios); b) formante gramatical: se trata de palabras gramaticales o funcionales carentes de significado (artículos, preposiciones, pronombres, conjunciones).

La mayoría de este tipo de estudios (Alegria Ugarte y Díaz Argüero, 2017; Piacente y Querejeta, 2012; Querejeta et al., 2013) señalan que, entre los tipos de hiposegmentación según la categoría de los formantes involucrados, predomina aquella referida a “formante gramatical + formante lexical” (por ejemplo, “derepen-te” (de repente), “sefue” (se fue)), así como la relativa a “formante gramatical + formante gramatical” (por ejemplo, “enla” (en la), “alos” (a los)). Los resultados muestran que la distribución de las categorías de los formantes difiere según el nivel de conocimiento sobre la segmentación convencional de los niños. Así, en los textos donde la segmentación convencional está más consolidada se confirma que los formantes implicados más frecuentes son los descritos anteriormente, mientras que en los textos donde predomina la segmentación no convencional, se encuentran otros tipos de hiposegmentación menos habituales como, por ejemplo, la tipología “formante lexical + formante lexical” y un formante lexical seguido de un formante gramatical. Estos resultados nos llevan a afirmar que “en niveles iniciales del conocimiento los niños exploran diversos criterios y posibilidades de separación de las palabras, variedad de criterios que irán abandonando

para quedarse con solo unos pocos a medida que aumenta el conocimiento de la escritura" (Alegria Ugarte y Díaz Argüero, 2017, p. 22). Por tanto, considerando que el fenómeno de la segmentación no convencional se ha encontrado en distintos sistemas alfabeticos, es posible señalar que la segmentación no convencional, más que un error ortográfico, es el resultado de una actividad cognitiva que realizan los alumnos al intentar comprender el funcionamiento de nuestro sistema de escritura.

Si bien existen estudios psicolingüísticos realizados, es preciso señalar que los textos escritos por alumnos de Educación Primaria se encuentran muy poco representados en los grandes corpus constituidos, los cuales son además muy limitados y poco accesibles (Elalouf, 2005; Salvador Mata 1999-2000). El estudio de textos a partir de grandes corpus permite indagar las competencias del alumnado y su evolución a lo largo del proceso de aprendizaje de la escritura, para lo cual es necesario el uso de las herramientas digitales propias del tratamiento automático del lenguaje. No obstante, existe escasez de estudios donde se utilice este tipo de herramientas que permita realizar el análisis automático de textos a partir de grandes corpus digitales (Wolfarth, 2019).

Así, en base a la revisión de la literatura previamente expuesta y en el marco del proyecto Scolinter, realizado en colaboración con la Universidad Grenoble Alpes (Francia) y la Universidad Milán Bicocca (Italia), se ha desarrollado este estudio en la Universidad de Almería cuyo objetivo es conocer la segmentación realizada por alumnos del primer curso de Educación Primaria procedentes del contexto español, a partir del análisis del corpus Scolinter y a través del uso de la herramienta digital Aliscol que integra el tratamiento automático del lenguaje, así como también comprender las razones de la segmentación no convencional desde un punto de vista psicogenético.

2. Metodología

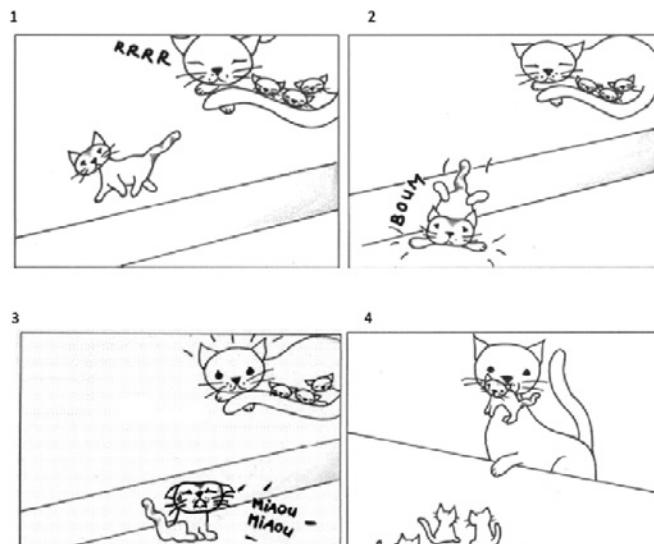
Este trabajo se ha desarrollado dentro del proyecto de investigación Scolinter realizado en la Universidad de Almería en colaboración con la Universidad Grenoble Alpes (Grenoble, Francia) y la Universidad Milán Bicocca (Milán, Italia). El objetivo general de este proyecto es la de describir y analizar las competencias ortográficas y narrativas de alumnos de 6 a 11 años de edad a partir de la construcción de un corpus digitalizado de carácter trilingüe (francés, italiano y español).

Para ello, este proyecto se ha llevado a cabo siguiendo un protocolo de investigación (Wolfarth et al., 2018) que ha permitido recopilar los textos escritos en diferentes cursos y contextos y analizarlos en sus diferentes dimensiones. Así, este estudio se ha realizado según el siguiente procedimiento metodológico: recogida de textos escritos; construcción y diseño del corpus Scolinter desde la recopilación, transcripción, anotación y tratamiento automático del lenguaje a través de la herramienta informática Aliscol; descripción y análisis de textos escritos.

A fin de desarrollar una descripción de forma detallada, a continuación se presenta la construcción del corpus de textos escritos por alumnos de primer curso de Educación Primaria procedentes del contexto español.

En el mes de julio del curso académico 2018/2019 se ha llevado a cabo en el aula la actividad referida a la producción individual por escrito de una historia a partir de la visualización de cuatro imágenes (véase Figura 1). Para la realización de dicha actividad se proporcionaba al alumnado la siguiente consigna, tomada del protocolo de la investigación "Lire–Écrire au CP" (Goigoux, 2016): "Hoy cada uno de vosotros va a escribir una historia, la historia de un pequeño gato. Os mostraré estas imágenes. Observad bien los dibujos. Vais a escribir la historia en la hoja de papel que os entrego. Tenéis 15 minutos y tenéis que trabajar solos; nadie podrá ayudaros a escribir el texto".

Figura 1. Imágenes utilizadas en la actividad de escritura



Nota. Tomado de Goigoux (2016)

La recogida de textos realizada por parte de los propios investigadores ha tenido lugar en 8 centros educativos situados en la provincia de Almería, pertenecientes a diferentes contextos socioculturales y seleccionados en base al interés y la cooperación por parte de los directores y profesores, previo contacto y presentación de información acerca del proyecto, bajo el compromiso y responsabilidad respecto al anonimato y la confidencialidad de los datos obtenidos.

Del total de 382 textos recopilados, 304 han sido seleccionados para el análisis presentado en este trabajo, quedando eliminados aquellos textos que no alcanzaban al menos el nivel de escritura alfabetica inicial. Tras la recogida de los textos, estos han sido digitalizados e incorporados al corpus trilingüe Scolinter para proceder sucesivamente con las fases de transcripción, normalización y alineación según el protocolo elaborado por Wolfarth et al. (2018).

La transcripción de los textos se ha realizado respetando la producción escrita de cada alumno, incluidos los errores ortográficos, sintácticos, de segmentación de palabras y la puntuación. Posteriormente, se ha llevado a cabo la fase de normalización, que consiste en elaborar la versión convencional de cada texto escrito siguiendo un meticuloso procedimiento de anotación manual previamente elaborado por el equipo de investigadores (Ponton et al., 2021), a través del cual se insertan unos códigos referidos al uso de la puntuación y del discurso directo, se marcan las partes ilegibles escritas por el alumno, se agregan palabras que faltan y se corrigen los errores ortográficos (véase Figura 2).

Figura 2. Texto escaneado, transcrita y normalizado (nº 99)



Nota. Tomado de <http://scoledit.org/scolinter/>

Finalmente, se ha desarrollado la fase referida a la alineación entre la transcripción y la normalización de cada texto, la cual se genera a través de una herramienta específica denominada Aliscol, diseñada por Wolfarth (2019) y adaptada al español. Una vez realizada la alineación entre el texto transcrita y el texto normalizado, se obtienen datos específicos relativos a distintos aspectos, tales como la categoría gramatical, el lema, el tipo de error ortográfico y la segmentación de las palabras (véase Figura 3).

Figura 3. Muestra de alineación correspondiente al texto nº 99

1	SegTrans	SegNorm	Catégorie	Lemme	StatutErreur	StatutSegm
2	El	El	ADE	el	01-Norme	01-Norme
3	gato	gato	SUS	gato	01-Norme	01-Norme
4	ba	va	VPE	ir	02-Phono	01-Norme
5	paseando	paseando	VNP	pasear	01-Norme	01-Norme
6	y	y	CCO	y	01-Norme	01-Norme
7	se	se	PPE	se	01-Norme	01-Norme
8	cae	cae	VPE	caer	01-Norme	01-Norme
9	esta	está	VPE	estar	03-Approxphono	01-Norme
10	llorando	llorando	VNP	llorar	01-Norme	01-Norme
11	mama	mamá	SUS	mamá	03-Approxphono	01-Norme
12	gata	gata	SUS	gato	01-Norme	01-Norme
13	selevanta	lo	PPE	lo	06-Sintaxis	03-HypoSeg
14	selevanta	levantar	VPE	levantar	01-Norme	03-HypoSeg
15	y	y	CCO	y	01-Norme	01-Norme
16	es	está	VPE	estar	03-Approxphono	02-HyperSeg
17	ta	está	VPE	estar	03-Approxphono	02-HyperSeg
18	feliz	feliz	ADJ	feliz	01-Norme	01-Norme

Tras la construcción y el diseño del corpus de textos escritos, se ha procedido a la fase de análisis enfocado al estudio de las competencias ortográficas y narrativas. Como se ha mostrado con anterioridad, el análisis presentado en este trabajo se ha centrado en la segmentación en palabras de textos escritos por alumnos de primer curso de Educación Primaria procedentes del contexto español, en respuesta a las siguientes preguntas: ¿Cuáles son los tipos más frecuentes de segmentación no convencional? ¿Qué características poseen la hiposegmentación y la hipersegmentación? Así, los objetivos concretos de este estudio son:

- Analizar los datos descriptivos sobre textos con segmentación convencional y textos con segmentación no convencional.
- Identificar los tipos de segmentación no convencional: hiposegmentación, hipersegmentación e hipohipersegmentación.
- Analizar la frecuencia y porcentaje de los sub-tipos de segmentación no convencional, según el número y tipo de formantes.
- Conocer los casos más frecuentes de hiposegmentación e hipersegmentación de dos formantes respecto al tipo de formantes que están involucrados.

3. Análisis y discusión de resultados

Tras el análisis de la segmentación de palabras en cada texto y en relación con el primer objetivo específico, se ha encontrado que el 25.33% de los textos presentan segmentación convencional. Esto significa que solo uno de cada cuatro niños, al final del primer curso de Primaria, ha logrado escribir un texto sin hacer uniones de palabras o separaciones indebidas entre segmentos de palabras. Al igual que en investigaciones anteriores (Fabbretti y Pontecorvo, 2005; Ferreiro y Pontecorvo, 1996; Sparrow, 2014), la segmentación de palabras resulta ser un problema incluso para aquellos alumnos que ya han entendido el principio alfabético de nuestro sistema de escritura.

Centrándonos en el uso de cada tipo de segmentación no convencional por parte del alumnado, se han distribuido los textos en tres categorías (véase Tabla 1), siendo la más frecuente aquella que incluye textos que presentan tanto casos de hiposegmentación como de hipersegmentación. Solo el 11.18% de los alumnos utilizan en sus textos únicamente la hipersegmentación.

Tabla 1. Distribución de textos según el uso de cada tipo de segmentación no convencional

	Frecuencia	Porcentaje
Textos que presentan únicamente hiposegmentación	77	25.33%
Textos que presentan únicamente hipersegmentación	34	11.18%
Textos que presentan ambos tipos de segmentación no convencional	116	38.16%
Total textos con segmentación no convencional	227	74.67%

Dentro del 74.67% de textos que presentan segmentación no convencional, como se puede observar en la tabla 2, el mayor porcentaje se sitúa en aquellos textos en los que el alumno utiliza con más frecuencia la hiposegmentación en comparación con la hipersegmentación. Como ya se ha encontrado en varios estudios (Correa y Dockrell, 2007; Farina, 2022; Ferreiro y Pontecorvo, 1996; Viramontes et al., 2021), la hiposegmentación representa el tipo de segmentación no convencional que más predomina en los textos escritos.

Tabla 2. Distribución de textos según el predominio de cada tipo de segmentación no convencional

	Frecuencia	Porcentaje
Textos con predominio de la hiposegmentación	140	46.05%
Textos con predominio de la hipersegmentación	55	18.09%
Textos que presentan tanto hipo como hiper	32	10.53%
Total textos con segmentación no convencional	227	74.67%

En referencia al segundo objetivo específico, para identificar los tipos de segmentación no convencional se ha procedido al análisis de la frecuencia y porcentaje de los mismos en relación con la longitud del texto, en cuanto a número total de palabras escritas (véase Tabla 3).

Tabla 3. Tipos de segmentación no convencional en relación con la longitud de los textos

Tipos de segmentación no convencional	Frecuencia	Porcentaje
Hiposegmentación	552	6.66%
Hipersegmentación	214	2.58%
Hipohipersegmentación	51	0.62%
Frecuencia de palabras en textos con segmentación no convencional	8292	

Como se puede observar en la tabla 3, la frecuencia y porcentaje de cada tipo de segmentación no convencional en relación con la longitud de los textos es baja, con un marcado predominio de la hiposegmentación (6.66%). En consecuencia, estos resultados nos llevan a indicar que los casos de segmentación no convencional no involucran de forma arbitraria a todas las palabras escritas, siendo necesario realizar un análisis cualitativo de los casos más frecuentes, como se muestra a continuación.

Respecto al tercer objetivo específico y en relación con el número de formantes implicados en los casos de hiposegmentación e hipersegmentación, los resultados encontrados se exponen en la tabla 4.

Tabla 4. Tipos de hiposegmentación e hipersegmentación según el número de formantes

Número de formantes	Hiposegmentación		Hipersegmentación	
	Frecuencia	Porcentaje	Frecuencia	Porcentaje
2 formantes	462	83.70%	203	94.86%
3 formantes	79	14.31%	11	5.14%
4 formantes	8	1.45%	-	-
5 formantes	1	0.18%	-	-
6 formantes	2	0.36%	-	-
Total casos	552		214	

Como se puede observar, la mayoría de los casos de segmentación no convencional están constituidos por dos formantes, siendo de un 94.86% los relativos a la hipersegmentación y de 83.70% a la hiposegmentación. Esto indica que al realizar uniones y separaciones indebidas, los alumnos siguen un criterio definido para ambos tipos de segmentación no convencional; el criterio resulta ser aún más restrictivo en el caso de la hipersegmentación dado que el alumnado no separa indebidamente más de tres segmentos de una palabra.

Profundizando en el análisis del tipo de formantes involucrados en los casos de hiposegmentación e indagando las características que estos presentan, es posible comprender las razones por las cuales los alumnos tienden a realizarlos. En este sentido, como se muestra en la tabla 5, prevalecen con un 71.01% aquellos casos de hiposegmentación cuyo primer formante es gramatical y considerando el segundo formante, resultan más frecuentes los relativos a la tipología lexical, como por ejemplo “ifueron” (y fueron). Estos datos que confirman los resultados de estudios previos (Alegria Ugarte y Díaz Argüero, 2017; Querejeta et al., 2013), indican las dificultades de los niños en entender que son palabras los formantes gramaticales que están involucrados en el fenómeno de la hiposegmentación, los cuales se encuentran desprovistos de significado.

Tabla 5. Tipos de hiposegmentación según el tipo de formantes

Tipo de formantes	Frecuencia	Porcentaje
Gramatical + Lexical	321	58.15%
Gramatical + Gramatical	71	12.86%
Lexical + Gramatical	25	4.53%
Lexical + Lexical	45	8.15%
Total casos hiposegmentación	462	

Con objeto de comprender las características y los motivos por los que los alumnos unen indebidamente las palabras y en respuesta al cuarto objetivo específico, es especialmente relevante identificar las categorías gramaticales más frecuentes involucradas en el primer formante de los casos de hiposegmentación. Así, en lo que respecta a la tipología relativa a los formantes “gramatical + lexical”, se ha encontrado que el

mayor porcentaje de casos, con un 33.76%, hace referencia a la unión de un pronombre personal y de un verbo en forma personal. Concretamente, se trata del caso más frecuente de hiposegmentación cuyo primer formante es el pronombre personal “se” (24.24%), como por ejemplo, “secallo” (se cayó), “sedesperto” (se despertó), “se fue” (se fue), “setropezo” (se tropezó). Además, en esta categoría se ha observado que el 8.44% de los casos de hiposegmentación involucran el pronombre personal “le/lo/la” como primer formante seguido del verbo, como por ejemplo, “ollebo” (lo llevó), “locogio” (lo cogió). Otro de los formantes gramaticales frecuentes involucrados en esta tipología es la preposición seguida de un formante lexical (10.82%), siendo el caso más representativo cuando este es un verbo en forma no personal, como por ejemplo “alloar” (a llorar), “adar” (a dar). Resulta interesante señalar, asimismo, que el 5.62% de los casos aluden a una hiposegmentación constituida por un determinante posesivo y un sustantivo (“sumadre” (su madre)) y el 4.97% involucran un artículo determinado seguido de un sustantivo (“elgato” (elgato)).

En el conjunto de los casos de hiposegmentación formados por dos formantes gramaticales, es preciso destacar que predominan aquellos con un 5.62% en los que el primer formante es una preposición y el segundo un artículo determinado (“alos” (a los), “porla” (por la)). Estos datos, desde el punto de vista psicogenético, pueden explicarse gracias a la noción de hipótesis de cantidad mínima (Ferreiro y Teberosky, 1985) por la que los alumnos se centran en aceptar palabras compuestas por pocas letras y, en consecuencia, tienden a no aislarlos sino a unirlos a la palabra que sigue.

En lo que respecta al análisis de los casos de hipersegmentación, no es posible considerar el concepto de formante, dado que generalmente no es posible identificar los tipos de segmentos que los integran (por ejemplo, “colora do” (colorado)). En consecuencia, el análisis se centra en la identificación de cuáles son los segmentos mayormente separados y qué características presentan. En la tabla 6 se exponen los casos más frecuentes con sus ejemplos más representativos, así como los porcentajes correspondientes en relación con el total de casos de hipersegmentación.

Tabla 6. Tipos de hipersegmentación según el tipo de segmento

Casos de hipersegmentación	Frecuencia	Porcentaje
Es + taba, Es + ta	30	14.77%
Era + se	24	11.82%
Des + pues	14	6.89%
Des + perto	7	3.44%
En + tonces	10	4.92%
En + contro	8	3.94%
A + el	3	1.47%
Ha + bia	3	1.47%
Por + que	4	1.97%

Se ha observado que los casos de hipersegmentación más frecuentes se refieren a aquellos en cuyo primer segmento se encuentran implicadas las formas verbales “es” y “era” (26.59%). A este tipo de casos les siguen aquellos que involucran a las preposiciones y, en especialmente, “en”, “de” y “a”. Al igual que en la investigación realizada por Ferreiro y Pontecorvo (1996), se puede destacar que el primer segmento mayormente presente en los casos de hipersegmentación se corresponde con secuencias gráficas autónomas y, por lo tanto, se encuentra separado del segundo segmento de la palabra al que convencionalmente debe unirse. Además, analizando el caso de la hipersegmentación “a el”, que involucran la preposición “a”, se puede señalar que lo que la norma ortográfica ha unido (preposición articulada) es, sin embargo, separado por el alumnado (a+el=al).

Respecto al análisis de la hipohipersegmentación, además de la baja frecuencia de casos, se puede observar que es difícil encontrar un criterio que identifique los distintos casos, de modo que nos permita conocer las características de este tipo de segmentación no convencional y comprender las razones por las que se utiliza. Así pues, es necesario analizar los resultados que se recopilen en los siguientes cursos de Educación Primaria, con el fin de indagar si se trata, en el primer curso, de una exploración que hacen los alumnos respecto a los criterios y posibilidades de segmentación de palabras o existen motivos específicos relacionados con la tipología de formantes implicados en la hipohipersegmentación.

4. Conclusiones

Pocos estudios se han cuestionado cómo se desarrolla el proceso de adquisición de la ortografía y por qué este se convierte en un escollo difícil de superar. Dar respuesta a estas preguntas requiere una serie de investigaciones aún no concluidas y que actualmente en nuestro país todavía siguen siendo insuficientes. A pesar de algunos estudios llevados a cabo, la adquisición de la ortografía sigue siendo un tema pendiente de resolver, acarreando problemas a largo plazo (Fayol y Jaffré, 2014; Viriot-Goeldel y Brissaud, 2019).

Igualmente, cabe señalar la escasez de estudios realizados a partir de grandes corpora y a través del uso de herramientas digitales propias del tratamiento automático del lenguaje. En este aspecto, es preciso hacer hincapié en la complejidad de realizar este tipo de análisis en textos escritos por alumnos de primer curso de Educación Primaria, teniendo en cuenta que dichos textos presentan un nivel de escritura que aún no es convencional (Wolfarth, 2019).

De ahí, la relevancia y originalidad de este trabajo de investigación que se ha desarrollado en la Universidad de Almería dentro del proyecto Scolinter en colaboración con la Universidad Grenoble Alpes y la Universidad Milán Bicocca. De hecho, este estudio se sitúa en un marco interdisciplinar con una gran proyección internacional gracias a la construcción de un corpus longitudinal y trilingüe y a la utilización de una nueva herramienta digital para el tratamiento de los textos ("Natural Language Processing") denominada Aliscol (Wolfarth, 2019). Para ello, se ha requerido de la coordinación e interdisciplinariedad del equipo de investigación integrado por especialistas en lingüística computacional, psicolingüistas, pedagogos y expertos en didáctica de la lengua (Ponton et al., 2021). Así, esta interconexión entre las diversas disciplinas ha favorecido la realización de un análisis comparativo del fenómeno de la segmentación en palabras que ha dado lugar a un estudio en profundidad de los textos en el contexto español.

Atendiendo a los resultados más relevantes del presente estudio, se puede confirmar que la segmentación supone una dificultad también para aquellos alumnos que, al finalizar el primer curso de educación primaria, ya han entendido el principio alfabetico de nuestro sistema de escritura: tres de cada cuatro alumnos, de hecho, ha realizado al menos un caso de segmentación no convencional en sus textos. El análisis de los casos de segmentación no convencional ha puesto de relieve cómo estos presentan características específicas, no son realizados "por casualidad" por los alumnos, ni tampoco involucran a todas las palabras. Así, donde la norma ortográfica sanciona como un "error" subyacen ideas concretas que los alumnos se plantean para comprender nuestro sistema de escritura.

Asimismo, siguiendo a investigaciones internacionales previas (Alegria Ugarte y Díaz Argüero, 2017; Correa y Dockrell, 2007; Farina, 2022; Ferreiro y Pontecorvo, 1996; Piacente y Querejeta, 2012; Querejeta et.al., 2013), el análisis de los datos obtenidos ha puesto de manifiesto cómo el concepto de formante more fológico es necesario, pero no suficiente para comprender las características de la segmentación no convencional.

En este sentido, en lo que se refiere a la hiposegmentación, que representa el tipo de segmentación no convencional que más predomina en nuestro corpus, se ha encontrado cómo la mayor parte se refiere a aquellos casos donde el primer formante es gramatical. Como se ha encontrado en otras investigaciones (Ferreiro y Pontecorvo, 1996; Gombert, 1990; Tolchinsky, 2003), estos datos revelan que en los casos de hiposegmentación se encuentran mayormente implicados los artículos, preposiciones y conjunciones, dado que este tipo de formantes no son fácilmente identificables como palabras semánticamente autónomas y no poseen un número adecuado de letras (hipótesis de cantidad mínima), mientras que los sustantivos, adjetivos y verbos, siendo semánticamente autónomos, son más fácilmente "aislados" como unidad-palabras.

En lo que respecta a la hipersegmentación, se ha observado que predominan aquellos casos cuyo primer segmento de la palabra se corresponde a una categoría lexical, como son las formas verbales "es" y "era", a los que siguen aquellos que involucran las preposiciones "en", "de" y "a". Así, al igual que los resultados encontrados en las investigaciones realizadas por Querejeta et al. (2013) y por Ferreiro et al. (1996) se puede afirmar que el mayor número de casos de hipersegmentación se sitúa en aquellos cuyo primer segmento implicado se corresponde con la identificación de secuencias gráficas autónomas con pocas letras.

Desde una mirada enfocada a las implicaciones prácticas de este trabajo, es preciso señalar que este estudio, ubicado en un escenario de reflexión y análisis sobre la adquisición de las competencias ortográficas, orienta el planteamiento de nuevas iniciativas en la formación de futuros profesionales de la educación de cara a la mejora de prácticas docentes que permitan el acceso a la lengua escrita. En este sentido, el análisis de los datos obtenidos conduce a reflexionar sobre las prácticas docentes en la enseñanza de la escritura: es evidente cómo un proceso de adquisición de la escritura llevado a cabo desde una perspectiva educativa centrada en la textualidad puede exponer directamente a los alumnos a problemas ortográficos de nuestro sistema de escritura, entre los que se encuentran los relativos a la segmentación entre palabras. De hecho, este enfoque didáctico posibilita en el alumnado de Educación Primaria la reflexión sobre el sistema de escritura a través de la elaboración de hipótesis propias, contribuyendo así a comprender tanto la función social de la escritura como los problemas que esta conlleva en relación con la segmentación y, en definitiva, a escribir textos de manera autónoma (Castoldi, 2019; Cisotto, 2019; Ferreiro, 2019b; Teruggi, 2019; Tolchinsky, 2008; Viramontes et al., 2021).

Por último, cabe resaltar que el análisis y discusión de resultados ha abierto otras futuras líneas de investigación, las cuales son necesarias para el estudio en profundidad sobre la segmentación en textos escritos, tales como: explorar el proceso de adquisición del sistema ortográfico y, en concreto, la relación existente entre los errores ortográficos y la segmentación no convencional; indagar la evolución longitudinal de la segmentación en textos escritos a lo largo de la etapa de Educación primaria; investigar las prácticas de enseñanza-aprendizaje de la escritura y su incidencia en los procesos de adquisición de competencias ortográficas.

5. Referencias bibliográficas

- Alegria Ugarte, M. I. y Díaz Argüero, C. (2017). ¿Se escribe junto o separado?: Evolución en el uso de los espacios en blanco en textos de niños que inician su alfabetización. *Revista Signos*, 50 (93), 3-25. <https://doi:10.4067/S0718-09342017000100001>

- Angelelli, P., Marinelli, C., Iaia, M., Notarnicola, A., Costabile, D., Judica, A., Zoccolotti, P. y Luzzatti, C. (2008). *DDO: diagnosi dei disturbi ortografici in età evolutiva*. Erickson.
- Báez, M. (1999). La problemática de la segmentación gráfica del texto en palabras: indagación en niños de sectores rurales. *Lectura y Vida*, 20 (2), 49-66.
- Barbagli, A., Lucisano, P., Dell'Orletta, F., Monemagni, S. y Venturi, G. (2014). Tecnologie del linguaggio e monitoraggio dell'evoluzione delle abilità di scrittura nella scuola secondaria di primo grado. En R. Basili, A. Lenci, B. y Magnini (Eds.), *The First Italian Conference on Computational Linguistics CLiC-it 2014* (pp. 23-27). University Press.
- Böhm, M. y Mehlem, U. (2020). What is a Word?: Word segmentation in multilingual writers writing French and Moroccan Arabic. *Written Language & Literacy*, 23(2). <https://doi:10.1075/wll.00038.boh>
- Borghi, C. C. (2014). Valutazione e misurazione automatizzata della produzione scritta. *Italian Journal of Educational Research*, 7, 11-22. <https://ojs.pensamultimedia.it/index.php/sird/article/view/261>
- Camacho, A., Alves, R. A. y Boscolo, P. (2021). Writing Motivation in School: a Systematic Review of Empirical Research in the Early Twenty-First Century. *Educational Psychology Review*, 33 (1), 213-247. <https://doi.org/10.1007/s10648-020-09530-4>
- Camps, A. (2006). *La enseñanza de la ortografía*. Graó.
- Castoldi, M. (2019). Osservare e valutare i primi apprendimenti sulla lettura e scrittura: spunti da una ricerca. *RicercaZione*, 11 (1), 17-32. <https://doi.org/10.32076/RA11101>
- Cignetti, L., Demartini, S. y Fornara, S. (coord.) (2016). *Come TIscrivo? La scrittura a scuola tra teoria e didattica*. Aracne.
- Cisotto, L. (2019). Insegnare a leggere e a scrivere: strategie, pratiche, contesti. *RicercaZione*, 11 (1), 33-52.
- Colombo, A. (2011). «A me mi». *Dubbi, errori, correzioni nell'italiano scritto*. Franco Angeli.
- Correa, J. y Dockrell, J. (2007). Unconventional word segmentation in Brazilian children's early text production. *Reading and Writing*, 20, 815-831. <https://doi.org/10.1007/s11145-006-9049-3>
- De Gaulmyn, M. M., Gonnand, S. y Luis, M. H. (1996). «S'il te plaît, écris-moi l'histoire du Petit Chaperon Rouge...». Costruction de la norme orthographique et grammaticale du CP au CE2. *Repères*, 14, 121-140. <https://doi.org/10.3406/REPER.1996.2197>
- Di Tullio, A. (2014). *Manual de Gramática del Español*. Waldhuter.
- Echauri, J. M. (2000). ¿Reformar la ortografía o reformar la enseñanza de la ortografía? *Cuadernos de Cervantes*, 6 (30), 22-27.
- Elalouf, M. L. (Ed.) (2005). *Enseigner à écrire entre 10 et 14 ans: un corpus, des analyses, des repères pour l'enseignement*. Versailles, Scénén & CRDP de l'Académie de Versailles.
- Fabbretti, D. y Pontecorvo, C. (2005). What Little Red Riding Hood Tells Us about Italian Children's Writing. En D. Ravid y H. Shyldkrot (eds.), *Perspectives on Language and Language Development* (pp. 241-254). Kluwer Academic Publishers. https://doi.org/10.1007/1-4020-7911-7_18
- Farina, E. (2022). La segmentazione delle parole nei testi di alunni di classe prima: costruzione del corpus e analisi del fenomeno. *RicercaZione*, 14(2), 1-20. <https://hdl.handle.net/10281/405564>
- Fayol, M. y Jaffré, J. P. (2014). *L'orthographe. Que sais-je?* PUF.
- Fernández-Rufete, A. (2015). Enseñanza de la ortografía, tratamiento didáctico y consideraciones de los docentes de Educación Primaria de la provincia de Almería. *Investigaciones sobre Lectura*, 4, 7-24. <https://doi.org/10.24310/revistaisl.vi4.10964>
- Ferreiro, E. (2013). *El ingreso a la escritura y a las culturas de lo escrito*. Siglo XXI.
- Ferreiro, E. (2019a). Prospectivas en investigación sobre alfabetización inicial. Dicotomías metodológicas y epistemológicas. *Infancia y Aprendizaje*, 42 (1), 1-36. <https://doi.org/10.1080/02103702.2018.1550174>
- Ferreiro, E. (2019b). *Alfabetización: teoría y práctica*. Siglo XXI.
- Ferreiro, E. y Pontecorvo, C. (1996). La segmentación en palabras gráficas. En E. Ferreiro, C. Pontecorvo, N. Ríbero e I. García Hidalgo (1996). *Caperucita Roja aprende a escribir. Estudios psicolingüísticos comparativos en tres lenguas* (pp. 45-73). Gedisa.
- Ferreiro, E. y Teberosky, A. (1981). La comprensión del sistema de escritura: construcciones originales del niño e información específica de los adultos. *Lectura y Vida*, 2 (1), 1-10.
- Ferreiro, E. y Teberosky, A. (1985). *Los sistemas de escritura en el desarrollo del niño*. Siglo XXI.
- Ferreiro, E. y Vernon, S. (1992). La distinción palabra/nombre en niños de 4 y 5 años. *Infancia y Aprendizaje*, 58, 12-28.
- Ferreiro, E., Pontecorvo, C., Ríbero, N. y García Hidalgo, I. (1996). *Caperucita Roja aprende a escribir. Estudios psicolingüísticos comparativos en tres lenguas*. Gedisa.
- Frith, U. (1985). Beneath the surface of developmental dyslexia. En K.E. Patterson, J. Marshall y M. Coltheart (Eds.), *Surface dyslexia*, (pp. 301-330). Erlbaum.
- Goigoux, R. (Ed.) (2016). *Lire et écrire. Étude de l'influence des pratiques d'enseignement de la lecture et de l'écriture sur la qualité des premiers apprentissages*. Rapport de recherche, Institut Français de l'Éducation. <http://ife.ens-lyon.fr/ife/recherche/lire-ecrire/rapport/rapport-lire-et-ecrire>.
- Gombert, J. E. (1990). *Le développement métalinguistique*. PUF.
- Gosse C., Parmentier, M. y Van Reybroeck, M. (2021). How Do Spelling, Handwriting Speed, and Handwriting Quality Develop During Primary School? Cross-Classified Growth Curve Analysis of Children's Writing Development. *Frontiers in Psychology*, 12 (685681), 1-19. <https://doi.org/10.3389/fpsyg.2021.685681>
- Joyce, N., Broc, L., Marshall, C. R. y Dockrell, J. E. (2022). Spelling Errors in French Elementary School Students: A Linguistic Analysis. *Journal of Speech, Language, and Hearing Research*, 65 (9), 3456-3470. https://doi.org/10.1044/2022_JSLHR-21-00507

- Kemp, N. y Treiman, R. (2023). Early Spelling Development. En S. Q. Cabell, S. B. Neuman y N. P. Terry (eds.), *Handbook on the Science of Early Literacy* (pp. 107-117). The Guilford Press.
- Ley Orgánica 3/2020, de 29 de diciembre, por la que se modifica la Ley Orgánica 2/2006, de 3 de mayo, de Educación. BOE, 30 de diciembre de 2020.
- Marinelli, C. V., Iaia, M. y Angelelli, P. (2016). Nonna Concetta: una prova di dettato di brano per la valutazione delle abilità ortografiche nella scuola primaria. *Psicologia Clinica dello Sviluppo*, 3, 424-449.
- Nemirosky, M. (1999). *Sobre la enseñanza del lenguaje escrito... y temas aledaños*. Paidós.
- Piacente, T. y Querejeta, M. (2012). La separación entre palabras en la escritura infantil. *Revista Neuropsicología Latinoamericana*, 4 (1), 1-17. <https://doi.org/10.5579/rnl.2012.0098>
- Pinto, G., Accorti Gamannossi, B. y Camilloni, M. (2017). Percorsi di alfabetizzazione emergente. La competenza notazionale. En M. Castoldi y M. Chicco (coord.), *Imparare a leggere e a scrivere. Lo stato dell'arte* (pp. 29-46). Iprase.
- Ponton, C., Gutiérrez Cáceres, R., Teruggi, L., Farina, E., Brissaud, C. y Wolfarth, C. (2021). Scolinter: un corpus trilingüe. L'exemple de la segmentation en mots. *Langue Française*, 3 (211), 37-50.
- Querejeta, M., Piacente, T., Guerrero, B. y Alva, E. A. (2013). La separación entre palabras en la escritura de niños que inician la escolaridad primaria. *Interdisciplinaria*, 30 (1), 45-64. <https://doi.org/10.16888/interd.2013.30.1.3>
- Salas, N. y Caravolas, M. (2019). Dimensionality of early writing in English and Spanish. *Journal of Literacy Research*, 51 (3), 272-292. <https://doi:10.1177/1086296X19858146>
- Salvador Mata, F. (2008). *Psicopedagogía de la Lengua Escrita (Vol. II: Escritura)*. EOS.
- Salvador Mata, F. (1999-2000). Habilidades narrativas de alumnos de Educación Primaria en la producción de textos escritos. *Enseñanza: anuario interuniversitario de didáctica*, 17-18, 145-163.
- Serianni, L. y Benedetti, G. (2009). *Scritti sui banchi. L'italiano a scuola tra alunni e insegnanti*. Carocci.
- Sparrow, W. (2014). Unconventional Word Segmentation in Emerging Bilingual Students' Writing: A Longitudinal Analysis. *Applied Linguistics*, 35 (3), 263-282. <https://doi:10.1093/applin/amt012>
- Teberosky, A. y Tolchinsky, L. (1995). *Más allá de la alfabetización*. Santillana.
- Teruggi, L. (2019). *Leggere e scrivere a scuola: dalla ricerca alla didattica*. Carocci.
- Tolchinsky, L. (2003). *The Cradle of Culture and what Children Know about Writing and Numbers before Being Taught*. Lawrence Erlbaum Associates.
- Tolchinsky, L. (2008). Usar la lengua en la escuela. *Revista Iberoamericana de Educación*, 46, 37-54. <https://doi.org/10.35362/rie460715>
- Viramontes, E., Quiñones, J. y Burrola, L. M. (2021). La segmentación del texto en la alfabetización inicial. *Educando para educar*, 22 (41), 47-64.
- Viriot-Goeldel, C. y Brissaud, C. (2019). Enseigner et apprendre l'orthographe aujourd'hui. *Repères*, 60, 15-22. <https://doi.org/10.4000/reperes.2271>
- Wolfarth C., Brissaud C. y Ponton C. (2018). Transcrire et normer un corpus scolaire: pour quelles analyses ? *Dyptique*, 36, 121-145.
- Wolfarth, C. (2019). *Apport du TAL à l'exploitation linguistique d'un corpus scolaire longitudinal*. Tesis Doctoral en Ciencias del Lenguaje. Universidad Grenoble Alpes.
- Zaduansky-Ehrlich, S., Seroussi, B. y Stavans, A. (2021). Predictors of spelling errors in expository texts written by Hebrew-speaking elementary school children. *Journal for the Study of Education and Development*, 44 (1), 117-149. <https://doi.org/10.1080/02103702.2020.1848089>
- Zayas, F. (2022). *Darle a la lengua. Claves para la enseñanza de la lectura y de la escritura*. COedCO.
- Zuccalá, G. y Sandbank, A. (2018). Producir textos en los inicios de la alfabetización: el impacto de diferentes enfoques pedagógicos en la expresión evaluativa. *Bellaterra Journal of Teaching & Learning Language & Literature*, 11, 54-73. <https://doi.org/10.5565/rev/jtl3.771>