

ISSN: 2036-5330

DOI: 10.32076/RA14206

La segmentazione delle parole nei testi di alunni di classe prima: costruzione del corpus e analisi del fenomeno

Word segmentation in the texts of first grade pupils: corpus construction and analysis of the phenomenon

Elisa Farina¹

Sintesi

Come imparano i bambini a segmentare le parole nella scrittura? Quali caratteristiche presentano le segmentazioni indebite che producono? Rispondere a queste domande implica analizzare ampi corpora di testi scritti autonomamente da alunni all'inizio del loro percorso di scolarizzazione, difficilmente analizzabili attraverso il trattamento automatico della lingua (TAL). Il presente contributo vuole allora offrire un duplice apporto alle ricerche linguistiche e didattiche: da un lato si propone di descrivere un modello automatico di analisi di testi che presentano un'ortografia non convenzionale e, dall'altro, descrivere, analizzare e comprendere il fenomeno della segmentazione delle parole al fine di rendere maggiormente coerente il rapporto tra apprendimento-insegnamento.

Parole chiave: Corpus linguistico; Ortografia; Segmentazione delle parole; Iposegmentazione; Ipersegmentazione.

Abstract

How do children learn to segment words in their writing? What are the characteristics of the improper segmentations produced by children? Answering these questions implies analysing large corpora of texts written by first grade pupils, which are difficult to analyse using automatic language processing (TAL). This paper therefore intends to make a twofold contribution to linguistic and didactic research: on the one hand, it aims to describe an automatic model for analysing texts with unconventional spelling and, on the other hand, to describe, analyse and understand the phenomenon of word segmentation in order to make the learning-teaching relationship more coherent.

Keywords: Language corpus; Spelling; Word segmentation; Hyposegmentation; Hypersegmentation.

1. Università degli Studi di Milano-Bicocca, elisa.farina@unimib.it

1. Introduzione

Le ricerche sulle competenze ortografiche degli alunni di scuola primaria intercettano una pluralità di approcci e linee di indagine che afferiscono a discipline e modelli interpretativi differenti. Tale campo di indagine risulta particolarmente esplorato dagli studi di matrice psicologica o clinica volti alla valutazione delle abilità ortografiche degli alunni all'inizio del percorso scolastico, partendo dalla somministrazione di test quali dettati di parole o non parole, oppure dettati di testi. Disegni di ricerca come questi sono finalizzati alla rilevazione e alla classificazione degli errori di scrittura, nonché alla valutazione delle competenze ortografiche degli alunni anche in funzione dell'individuazione di possibili deficit ortografici (Bozzo *et al.*, 2000; Sartori *et al.*, 2007; Angelelli *et al.*, 2008; Tressoldi *et al.*, 2012; Marinelli *et al.*, 2016). Attraverso l'impiego di batterie standardizzate viene dunque valutata la correttezza o meno della scrittura sotto dettatura di parole aventi specifiche caratteristiche (omofone/non omofone, digrammi, trigrammi, non parole, ecc.) con l'obiettivo di identificare possibili deficit a carico della strategia fonologica o lessicale (Patterson, 1986; Barry, 1994).

Altri filoni di ricerca, che hanno avuto un impulso soprattutto a partire dagli anni Ottanta, indagano in modo evolutivo il processo di acquisizione della lingua scritta il cui inizio - è ormai appurato - è da intendersi antecedente rispetto all'ingresso nella scuola primaria. (Ferreiro & Teberosky, 1985; Frith, 1985; Pascucci *et al.*, 1986; Dyson, 1993; Ehri, 1999; Pinto *et al.*, 2017; Castoldi & Chicco, 2017).

In particolar modo le ricerche psicogenetiche (Ferreiro *et al.*, 1996; Ferreiro, 2003) hanno evidenziato come il processo di acquisizione della lingua scritta «segua una linea evolutiva regolare, nei diversi ambienti culturali, nelle diverse situazioni educative e nelle diverse lingue» (Ferreiro, 2003, p. 10), e come la scrittura venga appresa grazie alla formulazione di ipotesi che i bambini elaborano per comprenderne il funzionamento, indipendentemente dalla presenza di un'istruzione formalizzata. Lungo tale processo evolutivo, la stabilizzazione della corrispondenza grafema-fonema - che porterà all'acquisizione della scrittura alfabetica - non può essere considerata come l'ultima tappa del cammino poiché gli alunni, ben presto, dovranno confrontarsi con altri problemi che emergono durante la scrittura di testi: l'ortografia, la separazione delle parole, l'uso delle maiuscole, la punteggiatura, ecc. La convenzionalità ortografica, intesa come ciò che viene assunto quale riferimento per la correzione di un testo dal nostro sistema di codifica, deve essere considerata come «punto di arrivo rispetto all'acquisizione di una parte importante del sistema di scrittura» (Ferreiro *et al.*, 1996, p. 31). Tuttavia, se si vuole comprendere come gli alunni acquisiscono le regole del nostro sistema di scrittura è necessario superare la dicotomia «corretto/non corretto» e allargare lo sguardo alla comprensione delle ipotesi costruttive che stanno alla base degli «errori» ortografici (Gaulmyn *et al.*, 1996).

Ciononostante, lo studio di come gli alunni si appropriano di queste regole del sistema ortografico attraverso la produzione autonoma di testi porta con sé problemi metodologici di non semplice risoluzione (Borghi,

2011) soprattutto se i soggetti coinvolti sono alunni di classe prima della scuola primaria la cui scrittura non è ancora convenzionale (Wolfarth, 2019). I testi di questi bambini, infatti, oltre a presentare parole che ortograficamente non seguono la norma (mancanza di doppie, uso non corretto di accenti e apostrofo, inversione di fonemi, mancanza di grafemi, ecc.) possono essere caratterizzati anche da una separazione non convenzionale tra le parole (fusioni o separazioni indebite) difficilmente riconoscibile con il trattamento automatico della lingua (TAL)². Proprio a seguito di questi problemi, le ricerche - seppure tuttora esigue considerando il valore anche sociale riconosciuto all'ortografia - (Seriani & Benedetti, 2009; Colombo, 2011), mirano alla descrizione e all'analisi delle competenze ortografiche partendo dalla raccolta di ampi corpora di testi scritti da alunni che hanno già raggiunto la scrittura convenzionale (Colombo, 2011; Barbagli *et al.*, 2014; Cignetti *et al.*, 2016).

Il presente contributo vuole allora offrire un duplice apporto alle ricerche finalizzate alla descrizione e all'analisi delle competenze ortografiche degli alunni: da un lato presentare un ampio corpus di testi di alunni di classe prima e descrivere il modello utilizzato per il trattamento automatico della lingua, dall'altro descrivere il fenomeno della segmentazione delle parole, utilizzando una prospettiva psicogenetica volta a comprendere le ragioni che si celano dietro a "normali" errori ortografici. In particolar modo, rispetto alla segmentazione delle parole, la ricerca qui descritta ha voluto contribuire a una maggiore descrizio-

ne e comprensione del fenomeno cercando di rispondere alle seguenti domande: "Quali sono le tipologie di ipo- ipersegmentazione più frequenti? Quali caratteristiche possiedono? Tali caratteristiche sono le stesse sia per le ipo sia per le ipersegmentazioni? Inoltre, considerando le caratteristiche della lingua italiana, le iposegmentazioni sono strettamente legate all'uso dell'apostrofo? Per fornire una risposta a tali domande sono stati individuati i seguenti obiettivi:

- analizzare la frequenza e percentuale dei testi con segmentazione convenzionale/non convenzionale e le tipologie di segmentazioni (iposegmentazioni, ipersegmentazioni e ipoi-ipersegmentazioni);
- analizzare la frequenza e la percentuale, all'interno di ciascuna tipologia di segmentazione non convenzionale, dei sottotipi di segmentazione partendo dal costrutto di *formante*;
- individuare i casi più frequenti di ipo e ipersegmentazione costituiti da due formanti per comprendere quali sono gli elementi grammaticali coinvolti;
- individuare la percentuale di iposegmentazioni legate all'uso dell'apostrofo.

2. Che cosa è una parola?

La capacità di produrre testi scritti è un obiettivo centrale nell'insegnamento della lingua scritta verso cui si concentra, fin dall'inizio del percorso di scolarizzazione, gran parte degli sforzi dei docenti. Tuttavia, nelle più

2. Per Trattamento Automatico della Lingua si intende, in linea generale, l'elaborazione della lingua parlata o scritta attraverso l'uso della tecnologia.

comuni pratiche didattiche, (superando la dicotomia tra i metodi: Cisotto, 2019; Castoldi & Chicco, 2017) l'insegnamento della lingua scritta procede lungo una linea progressiva che possiamo definire "dalla parola al testo", ossia, si procede molto più frequentemente (e forse anche facilmente e "tranquillamente") con la scrittura di parole e l'acquisizione delle principali convenzioni ortografiche (digrammi, trigrammi, uso dell'h, ecc.) secondo un iter sempre più spesso definito dai libri di testo. All'interno di questo percorso viene dato per scontato il concetto di parola con scarsa consapevolezza rispetto al fatto che, su un piano linguistico, storico e psicolinguistico, tale costruito è tutt'altro che naturale.

Dal punto di vista linguistico la parola ha una realtà intuitiva ma è priva di uno status teorico: «[...] word, thatmost elusive of all-nitsencountered in the segmentation of utterances» (Markiel, 1970, in Ferreiro *et al.*, 1996, p. 40). La nozione di parola sembra dunque essere incerta, intuitiva tanto che, a seconda dei modelli interpretativi utilizzati, è stata definita come unità sintattica minima, unità sintattico-referenziale connessa a un determinato uso grammaticale, unità intonativa o, ancora, come unità grafica (Ferreiro *et al.*, 1996). È ormai appurato che il concetto di parola scritta non è sovrapponibile a quello di parola nell'enunciato, dove i confini seguono regole differenti (si pensi all'enunciato "C'era una volta" che, dal punto di vista grafico, presenta una separazione in quattro unità, inesistenti nell'enunciazione). Sembra dunque essere la scrittura a marcare, attraverso l'utilizzo di spazi bianchi, la separazione di ciò che, nel parlato, corrisponde a un continuum sonoro. Tale affermazione è facilmente "verificabile"

osservando la modalità di dettatura di testi da parte degli insegnanti di scuola primaria i quali, per far sì che i bambini separino le parole, le dettano molto frequentemente una a una e, in alcuni casi, dicendo esplicitamente "spazio" per marcare la separazione (Farina, 2014). Di fatto, come sostiene Eynard, se l'alunno durante la dettatura trasformasse correttamente per iscritto ciò che sente, commetterebbe una grande quantità di errori (Eynard, 1992). Scrivendo, dunque, separiamo le parole ma sarebbe più corretto affermare che è la scrittura a definire il concetto di parola dato che: «la scrittura ci offre la migliore definizione *pratica* (e non teorica) di parole: insieme di lettere separate da uno spazio bianco» (Ferreiro *et al.*, 1996, p. 42). Se, linguisticamente, il concetto di parola risulta quindi incerto e difficilmente definibile, anche dal punto di vista storico la pratica della separazione delle parole segue un andamento altalenante con una preferenza della *scriptio continua* dal II secolo fino al Rinascimento; la separazione sistematica delle parole da parte dell'autore del testo (e non del lettore) risulta infatti essere un'acquisizione medioevale (Günter, 1997). Separare le parole, dunque, nella cultura occidentale non è stato un fatto scontato, immediato e precoce e, inoltre, ciò che viene separato dagli spazi bianchi dipende dalla lingua; nell'Inuit (lingua parlata dal popolo Inuit in Canada) ad esempio, solo i nomi propri sono segmentati in una frase, mentre in ebraico le preposizioni all'interno di una frase si presentano attaccate alle parole del contenuto, come morfemi legati (Correa & Dockrell, 2007; Tolchinsky, 2006).

Date queste premesse, le incertezze nel separare correttamente le parole, incontrate

dai bambini all'inizio del percorso di scolarizzazione, non devono stupire soprattutto se, come abbiamo visto, la parola è una partitura del testo e si costruisce attraverso la scrittura (e non viceversa); inoltre, la pratica della segmentazione delle parole può risultare più complessa se all'inizio del processo di insegnamento della lingua scritta non ci si sofferma a riflettere con gli alunni su che cosa sia una parola. Interrogativo, questo, a cui hanno rivolto l'attenzione gli studi psicolinguistici i quali confermano che per i bambini prealfabetizzati il concetto di parola risulta essere molto complesso (prediligendo il termine "nome") (Ferreiro & Vernon, 1992) e che la segmentazione tra le parole si fa sistematica verso il secondo-terzo anno di scolarizzazione.

Attraverso l'osservazione e l'analisi delle pratiche didattiche, laddove gli insegnanti sono consapevoli della problematicità che il concetto di parola porta con sé, è possibile riscontrare quanto affermato dagli studi psicolinguistici. Dal segmento di trascrizione, riportato qui di seguito, relativo a una conversazione avvenuta in classe prima, attorno alle parole che costituiscono il titolo del libro "Giulio coniglio e il lupo", scritte su dei cartellini, si può evincere quanto il concetto di parola non possa essere dato per scontato³:

- *Quelle che vedete sui cartellini sono tutte parole?*
- A.: lettere.
- R.: ma le lettere insieme sono parole. (Legge «coniglio»), anche questa è una parola, (legge «e»), sì, è una parola, (legge «il»), è una parola, anche

«lupo» è una parola!

- *Siete tutti d'accordo con R.?*
- A.: la «e» non è una parola perché non ci sono tante lettere. Anche nella mia (si riferisce a quanto scritto sul cartellino che ha in mano, cioè «il») ce ne sono poche. Per me «il» non è una parola.
- M.: è vero. La «e» è solo una lettera.
- R.: però «e il lupo» è una parola tutta insieme.
- L.: non sono parole quelle due (indica «e» e «il») perché sono senza nome.

Da questo breve esempio emerge in modo evidente l'ipotesi di quantità minima (A.: "ce ne sono poche"), (Ferreiro & Teberosky, 1985) riscontrabile in diverse lingue e la difficoltà, messa in luce dagli studi psicolinguistici sulla segmentazione delle parole, ad accettare che alcune categorie grammaticali come articoli, preposizioni o congiunzioni possano essere considerate parole.

2.1. Studi psicolinguistici sulla segmentazione delle parole nei testi di alunni di scuola primaria

La segmentazione delle parole all'interno di testi autonomi scritti da bambini all'inizio della scolarizzazione risulta un campo ancora poco esplorato, soprattutto in ambito nazionale, anche per i problemi metodologici che una simile indagine comporta. A partire dagli anni Novanta del secolo scorso, tuttavia, in ambito internazionale il fenomeno è stato

3. Protocollo raccolto dall'insegnante Rosalba Negri durante una conversazione in classe prima nella Scuola Primaria di Cossato Centro (BI). Il corsivo indica gli interventi dell'insegnante.

studiato anche su corpora ampi e attraverso disegni di ricerca comparativi e/o longitudinali per comprendere se vi fosse una differenza all'interno delle scritture alfabetiche e quale potesse essere l'evoluzione del fenomeno con il crescere dell'età e delle competenze di scrittura (Gombert, 1990; Ferreiro *et al.*, 1996; Correa & Dockrell, 2007; Piacente & Querejeta, 2012, Querejeta *et al.*, 2013; Ugarte & Díaz Argüero, 2017). I principali apporti che tali ricerche hanno messo in luce riguardano differenti aspetti.

In primo luogo, sono state identificate e classificate le segmentazioni indebite prodotte dagli alunni andando così a definire il concetto di iposegmentazione (tendenza a lasciare meno separazione rispetto a quella convenzionale, come, per esempio: "seneva" = se ne va), di ipersegmentazione (tendenza a lasciare uno spazio bianco dove non è necessario come, per esempio: "in tanto" = intanto) e, nelle ultime ricerche (Querejeta *et al.*, 2013) (comprovate anche nello studio qui presentato) quello di segmentazioni miste denominate "ipoipersegmentazioni" (come, per esempio, in "a limprovviso = all'improvviso, in cui compare sia una iposegmentazione sia una ipersegmentazione). Tutti questi studi sembrano confermare, indipendentemente dalla lingua, dal tipo di carattere utilizzato (script, maiuscolo, corsivo) e dal metodo di insegnamento della lingua scritta adottato, una netta prevalenza, in termini quantitativi, delle iposegmentazioni (nel primo anno di scuola) rispetto alle ipersegmentazioni. Inoltre, gli studi di Ferreiro (Ferreiro *et al.*, 1996), Báez (1999), Ugarte e Díaz Argüero (2017) evidenziano come le iposegmentazioni continuino ad essere la categoria predominante

tra le segmentazioni non convenzionali anche nei gradi superiori di scolarizzazione.

In secondo luogo, le analisi qualitative condotte su questi corpora hanno consentito di indagare quali sono gli elementi maggiormente implicati nei fenomeni di segmentazione non convenzionale, cercando di comprenderne le ragioni. Per questo scopo, al fine di poter includere il maggior numero di elementi coinvolti nei fenomeni di ipo e ipersegmentazione, le ricerche più recenti utilizzano il concetto di *formante morfologico* (Di Tullio, 1997) distinguendo i formanti lessicali (sostantivi, aggettivi, verbi, avverbi) dai formanti grammaticali (articoli, preposizioni, pronomi, congiunzioni); i primi comprendono parole lessicali che possiedono un significato proprio, mentre i secondi sono costituiti da parole grammaticali (o funzionali) prive di significato. Se la maggior parte degli studi concordano nel ritenere che nelle segmentazioni non convenzionali predomina la tipologia (formante grammaticale + formante lessicale) o l'unione di due formanti grammaticali, le ricerche di Ugarte e Díaz Argüero hanno evidenziato una diversa distribuzione in base alle competenze di scrittura dei bambini: nei testi in cui la segmentazione convenzionale è più stabile (in alunni di classe quarta o comunque in testi con poche segmentazioni non convenzionali), le categorie maggiormente implicate si confermano quelle precedentemente descritte. Nei testi con il maggior numero di segmentazioni non convenzionali si trovano anche abbinamenti meno usuali come, per esempio, formante lessicale + formante lessicale (il gattova = il gatto va), oppure formante lessicale seguito da formante grammaticale (il gattoe = il gatto e). Tali risultati portano a so-

stenero che «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 aumenta el conocimiento de la escritura»⁴ (Ugarte & Díaz Argüero, 2017, p. 22).

Superando la dicotomia tra corretto/non corretto e considerando che tale fenomeno è stato riscontrato in diverse lingue alfabetiche, è possibile affermare che le segmentazioni non convenzionali, più che errori ortografici, sono il risultato di un esercizio razionale che gli alunni compiono nel cercare di comprendere il funzionamento del nostro sistema di scrittura.

3. La ricerca

La ricerca qui presentata è parte integrante del progetto Scolinter realizzato in collaborazione con l'Università di Grenoble Alpes (Francia) e di Almería (Spagna), finalizzato alla costruzione di un corpus digitalizzato, longitudinale e trilingue (italiano, francese e spagnolo), volto a descrivere le competenze ortografiche e narrative di alunni dai 6 agli 11 anni. La ricerca è stata realizzata grazie alla condivisione di un protocollo di ricerca (Wolfarth *et al.*, 2018) che ha consentito, nei diversi anni e contesti, di raccogliere i testi seguendo la medesima metodologia. Al fine di elaborare un'analisi e descrizione dettagliata, qui di seguito verranno presi in considerazione solamente i dati italiani riferiti ai testi prodotti dagli alunni di classe prima della scuola

primaria. Nell'anno scolastico 2018/2019, mese di maggio, sono stati raccolti 689 testi coinvolgendo 39 scuole del territorio lombardo, appartenenti a contesti socioculturali differenti e selezionate a partire dalla disponibilità data dai dirigenti scolastici contattati attraverso una lettera. I testi sono stati raccolti dai ricercatori fornendo la seguente consegna, adottata nella ricerca Lire-Écriteau CP (Goigoux 2016) e definita nel protocollo: "Oggi ognuno di voi scriverà la storia di un piccolo gatto. Guardate bene i disegni. Scriverete la storia sul foglio che vi consegno, come siete capaci. Avrete 15 minuti di tempo ma dovrete lavorare da soli; nessuno potrà aiutarvi nello scrivere il testo".

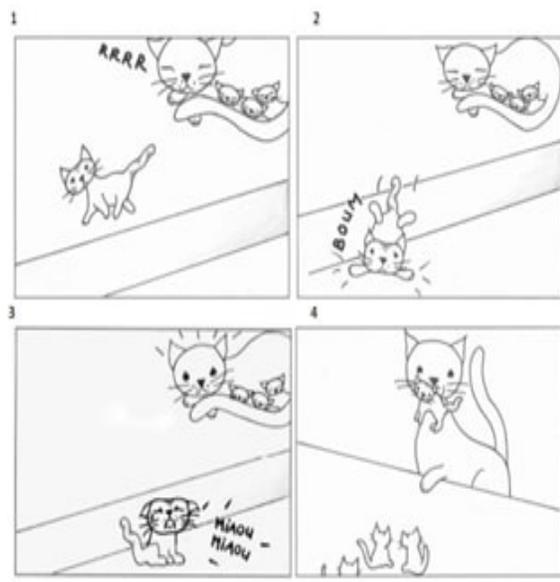


Fig. 1 - Immagini utilizzate per la produzione narrativa.

Tutti i testi raccolti sono stati successivamente digitalizzati al fine di costruire il corpus Scolinter e consentire il trattamento

4. Ai livelli iniziali di conoscenza, i bambini esplorano una varietà di criteri e possibilità per separare le parole; una varietà di criteri che abbandoneranno gradualmente per mantenerne solo alcuni man mano che la loro conoscenza della scrittura cresce.

automatico della lingua. Le fasi che hanno portato alla digitalizzazione del corpus, oltre alla scansione di tutti i testi, possono essere descritte in termini di: trascrizione, normalizzazione e allineamento.

3.1. *Trascrizione, normalizzazione e allineamento dei testi*

Seguendo il protocollo elaborato da Wolfarth *et al.* (2018) tutti i testi sono stati trascritti rispettando, più fedelmente possibile, la produzione dell'alunno: è stato dunque riscritto il testo mantenendo tutte le scelte linguistiche adottate dall'alunno, compresi gli errori ortografici, sintattici, di segmentazione delle parole, la presenza di segni tra una parola e l'altra e i criteri utilizzati per andare a capo al termine della riga o di un paragrafo⁵. Confrontando le Figg. 2 e 3 è possibile comprendere il lavoro di trascrizione svolto dai ricercatori.

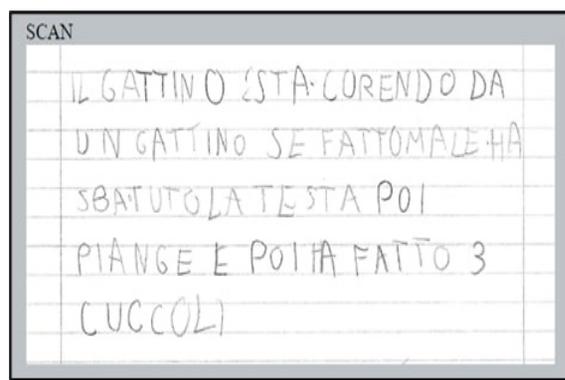


Fig. 2 - Testo dell'alunno I00000711.

Come emerge dalla Fig. 2, l'alunno utilizza lo stampato maiuscolo, va a capo al termine della riga senza spezzare le parole (marcato con il segno / nella trascrizione), omette i segni di punteggiatura e compie delle segmentazioni non convenzionali tra le parole (non sempre facili da disambiguare, soprattutto nello stampato maiuscolo, come nel caso di "sba.tutola testa") inserendo anche, in alcuni casi, un punto di separazione tra una parola e l'altra: tali scelte sono state riportate fedelmente nella trascrizione, così come gli errori ortografici e sintattici (Fig. 3).

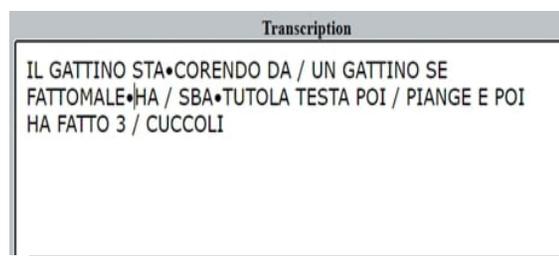


Fig. 3 - Trascrizione del testo I00000711.

Osservando il testo che, per ortografia, sintassi, segmentazione delle parole e punteggiatura non rispetta la convenzionalità, è possibile comprendere la difficoltà di giungere a un trattamento automatico della lingua. A tale scopo si è resa necessaria la fase di standardizzazione, definita "normalizzazione", consistente nel produrre un'altra versione del testo aggiungendo, attraverso l'uso della piattaforma Scolinter⁶, alcuni *tag* che rendessero visibili gli interventi di punteggiatura, aggiunta di parole mancanti, presenza del discorso diretto o la marcatura di parti illeggibili prodotte dagli alunni; l'ortografia è stata invece normalizzata senza l'aggiunta di

5. Tutti i segni non corrispondenti a grafemi o non identificabili come segni di interpunzione sono stati codificati come "disegni". Se, come nel caso della Fig. 2, sono visibili dei segni che l'alunno ha però provato a cancellare (come quello presente tra il "gattino" e "sta"), questi non sono stati considerati. I puntini inseriti tra una parola o l'altra (come tra "sta" e "correndo") sono stati riportati nella trascrizione.

6. <http://scoledit.org/scolinter/index.php>

etichette, come mostrato nella Fig. 4.

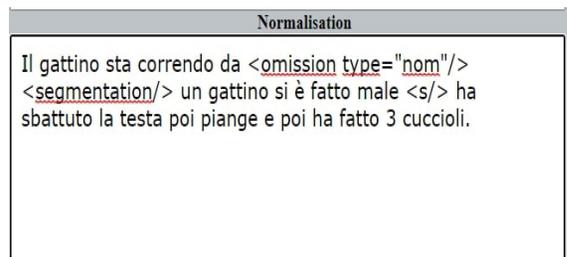


Fig.4 - Normalizzazione del testo I00000711.

La fase di normalizzazione è consistita dunque in una annotazione manuale che, nel caso del testo I00000711, ha riguardato l'aggiunta di un tag in corrispondenza del segmento "sta correndo da" poiché l'alunno non inserisce il complemento, di un altro tag per indicare la mancanza di un punto <segmentation/>, l'inserimento di una virgola <s> e l'eliminazione degli errori ortografici. I testi così normalizzati, rispetto a quelli trascritti, presentano delle caratteristiche leggibili dal TAL; ciò ha consentito di procedere con la fase successiva definita di "allineamento" attraverso uno strumento specifico denominato Aliscol, ideato per il francese da Wolfarth (2019) e adeguato all'italiano. Tale strumento produce un allineamento forma-forma tra la trascrizione e la corrispondente normalizzazione: la differenza fra la produzione iniziale e la normalizzazione permette di descrivere in modo dettagliato diversi aspetti della lingua fornendo informazioni, come si evince dalla Fig. 5, rispetto alla categoria grammaticale, al lemma, al tipo di errore ortografico e, non da ultimo, alla segmentazione delle parole.

IdEleva	Seg Trans	Seg Norm	Categorie	Lemma	Status Erreur	Status Segm
I00000711	IL	ll	ADE	ll	01-Normè	01-Normè
I00000711	GATTINO	gattino	SUS	gattino	01-Normè	01-Normè
I00000711	STA	sta	VPE	sta	01-Normè	01-Normè
I00000711	CORRENDO	correndo	VNP	correndo	06-ApproxPhono	01-Normè
I00000711	DA	da	PRE	da	01-Normè	01-Normè
I00000711	UN	un	AIN	un	01-Normè	01-Normè
I00000711	GATTINO	gattino	SUS	gattino	01-Normè	01-Normè
I00000711	SE	si	PPE	si	02-Phono	03-HypoSeg
I00000711	SE	è	VC1	è	02-Phono	03-HypoSeg
I00000711	FATTOMALE	fatto	VC2	fatto	01-Normè	03-HypoSeg
I00000711	FATTOMALE	male	SUS	male	01-Normè	03-HypoSeg

Fig. 5 - Esempio semplificato di allineamento del testo I00000711

Grazie alla fase di allineamento è stato dunque possibile procedere con una pluralità di analisi a seconda del focus di ricerca: l'indagine qui descritta, come definito in precedenza, si è focalizzata sulla segmentazione delle parole con l'obiettivo di descrivere il fenomeno all'interno dei testi di classe prima.

4. Analisi e discussione dei dati

All'interno del corpus è stata svolta una prima analisi quantitativa che ha permesso di calcolare la frequenza e la percentuale dei testi che presentano una segmentazione non convenzionale. Dal numero totale di testi raccolti (689) sono stati eliminati quelli caratterizzati da una scrittura non ancora alfabetica⁷ poiché non significativi rispetto al fenomeno della segmentazione delle parole. Come evidenziato anche dalle ricerche precedenti, la segmentazione delle parole risulta infatti es-

7. Un esempio di testo eliminato è rappresentato dal numero I000001011 (ALHZLFOMLEUZRRI) che presenta una scrittura presillabica (senza alcuna corrispondenza tra grafema e fonema) e scritto in *scriptio continua*. Un altro esempio è rappresentato dal numero I00027121 (IL GTNO SAVADNO) che presenta una scrittura sillabica. In entrambi i casi i bambini non sono ancora giunti a una scrittura alfabetica stabile per cui non risulta significativo analizzare la segmentazione delle parole.

sera un problema anche per quegli alunni che hanno già compreso il principio alfabetico del nostro sistema di scrittura. A conferma di ciò, come illustrato nella Tab. 1, i testi che presentano una segmentazione convenzionale sono solamente un quarto rispetto al totale dei testi analizzati; ciò significa che, solo un

Tipologia di testi	Frequenza	Percentuale
Segmentazione convenzionale	167	25%
Segmentazione non convenzionale	501	75%
Totale	668	100%

Tab.1 - *Distribuzione dei testi in base alla segmentazione.*

bambino su quattro del campione analizzato, al termine della classe prima è riuscito a scrivere un testo senza compiere unioni o separazioni indebite tra le parole⁸.

All'interno del 75% dei testi con segmentazione non convenzionale, sono stati

Tipologia di testi	Frequenza	Percentuale
Testi con solo iposegmentazioni	336	50,3%
Testi con solo ipersegmentazioni	29	4,3%
Testi con ipo-ipersegmentazioni	136	20,4%
Totale	668	75%

Tab. 2 - *Distribuzione dei testi in base alla tipologia di segmentazione non convenzionale.*

analizzati quelli che presentano unicamente iposegmentazioni, quelli in cui sono state

identificate solamente ipersegmentazioni e quelli che includevano entrambe le tipologie.

Il fenomeno della iposegmentazione, come già evidenziato dagli studi di Ferreiro *et al.*, (1996), Correa e Dockrell (2007) risulta essere quello predominante tanto che, in unione anche alla ipersegmentazione, è presente in quasi tutti i testi non convenzionali; solo un 4,3% dei testi presenta unicamente casi di ipersegmentazione.

Se in questa prima analisi l'unità considerata è stata il testo nella sua totalità, successivamente sono state calcolate la frequenza e la percentuale dei casi di segmentazione non convenzionale in relazione alla quantità di parole scritte (occorrenze lessicali).

Casi di segmentazione non convenzionale	Frequenza	Percentuale
Iposegmentazione	1066	3,99%
Ipersegmentazione	161	0,6%
Ipoipersegmentazione	40	0,14%
Frequenza di parole nei testi	26712	

Tab. 3 - *Distribuzione dei casi di segmentazione non convenzionale in relazione alle occorrenze lessicali.*

I dati della Tab. 3 appaiono particolarmente significativi se confrontati con quelli della Tab. 1: nonostante il fenomeno delle segmentazioni non convenzionali sia molto frequente tra gli alunni al termine della classe prima (75% dei testi hanno una segmentazione non convenzionale), la percentuale dei casi è piuttosto bassa, con una prevalenza marcata per le iposegmentazioni. Tale risul-

8. La lunghezza dei testi non si è rivelata una variabile statisticamente significativa rispetto alla percentuale di segmentazioni non convenzionali: testi più corti non sono risultati, di fatto, più corretti dal punto di vista delle segmentazioni.

tato, di conseguenza, ci induce ad analizzare in modo più approfondito i singoli casi di ipoiper e ipoipersegmentazione poiché, la loro bassa frequenza rispetto al numero totale di occorrenze lessicali, ci fa ipotizzare che i casi individuati presentino delle caratteristiche specifiche e non coinvolgano in modo arbitrario tutte le parole scritte.

Focalizzando dunque l'attenzione in modo specifico sui casi di segmentazione non convenzionale è stato possibile individuare alcuni sottotipi partendo dal costrutto di *formante morfologico* (Di Tullio, 1997); nella Tab. 4 sono presentati i casi individuati all'interno del corpus.

Come si evince dai dati riportati nella Tab. 4, la quasi totalità dei casi è costituita da due formanti; ciò significa che gli alunni, pur compiendo unioni e separazioni indebite, seguono un criterio preciso sia per le ipoipersegmentazioni sia per le ipersegmentazioni;

in quest'ultimo caso il criterio appare ancora più restrittivo poiché non compaiono mai casi in cui gli alunni ipersegmentano più di tre formanti. Nel caso delle iposegmentazioni, il costrutto di formante lessicale e grammaticale consente di scoprire le caratteristiche che tali unioni non convenzionali presentano. Relativamente alle ipersegmentazioni non è invece possibile prendere in considerazione il costrutto di formante poiché l'ipersegmentazione spezza la parola in due parti (es: in sieme; in vece; a veva) di cui la seconda ("sieme", "vece", "veva") non è classificabile.

Focalizzando dunque l'attenzione sulle iposegmentazioni tra due formanti, all'interno del corpus predominano le segmentazioni tra un formante grammaticale e uno lessicale (per esempio: sisveglia, siera, aterra, unaltro).

Se il predominio della categoria "grammaticale + lessicale" conferma i risultati degli studi precedenti (Querejeta *et al.*, 2013;

Numero di formanti	Iposegmentazione	Ipersegmentazione	Esempi
2 formanti	969 (90,90%)	158 (98,14%)	aprendere; in sieme
3 formanti	86 (8,07%)	3 (1,86%)	seneva; di e tro
4 formanti	10 (0,94%)	0	sièfattomale
5 formanti	1 (0,09%)	0	edorainpoi
Totale casi	1066	161 (100%)	

Tab. 4 - Distribuzione dei casi di ipo e ipersegmentazione in base al numero di formanti.

Tipo di formante	Frequenza	%	Esempi
gramm+lex.	444	41,64%	aprendere
gramm+gramm.	143	13,41%	sene
lex+gramm.	35	3,28%	gattoe
lex+lex.	347	32,55%	cera
Totale	969	90,89%	

Tab. 5 - Distribuzione delle tipologie di formanti nei casi di iposegmentazione.

Ugarte & Díaz Argüero, 2017), il 32,55% della categoria “lessicale + lessicale” risulta una novità che ha indotto i ricercatori a svolgere un’analisi qualitativa per comprenderne le ragioni. Andando ad analizzare solamente quest’ultima categoria, emerge un dato particolarmente interessante che giustifica la frequenza inaspettata dei casi: 264 su 347 si riferiscono all’unione indebita di “cera” all’interno dell’incipit (c’era una volta). In questo caso la “C”, pur essendo solamente una lettera, costituisce di fatto un avverbio (C’era = era in quel luogo/tempo) e, pertanto, rientra nella categoria “formante lessicale” anche se è facile comprendere come, per alunni di classe prima, una singola lettera non possieda una funzione referenziale autonoma ma presenti le stesse caratteristiche degli altri formanti grammaticali. Inoltre, l’unione di “cera”, dal punto di vista psicogenetico, è spiegabile grazie alla nozione di *ipotesi di quantità minima* (Ferreiro & Teberosky, 1985) per cui gli alunni faticano ad accettare parole costituite da poche lettere e, conseguentemente, tendono a non isolarle ma a unirle alla parola che segue. Infine, è plausibile che la norma ortografica che richiede l’uso dell’apostrofo non sia ancora automatizzata al termine della classe prima. L’analisi condotta sul caso “C’era” ci porta ad affermare che, per comprendere e descrivere le caratteristiche delle ipo e ipersegmentazioni, anche la nozione di formante - utile per una prima analisi - consente solo parzialmente di capire le ragioni per cui gli alunni tendono a compiere separazioni non convenzionali. Risulta dunque necessario procedere con un’analisi capace di mettere in luce ulteriori caratteristiche di tali formanti.

4.1. Le iposegmentazioni: analisi dei casi più frequenti

Dopo aver disambiguato il caso della categoria lessicale + lessicale, le due categorie che presentano un numero maggiore di iposegmentazioni, come illustrato nella Tab. 5, appartengono all’unione di formante grammaticale + lessicale e di formante grammaticale + grammaticale. L’analisi effettuata sulle occorrenze lessicali maggiormente implicate nei casi di iposegmentazione ha messo tuttavia in luce che, per comprendere le caratteristiche e le ragioni per cui gli alunni iposegmentano, non è particolarmente rilevante osservare il secondo formante (lessicale o grammaticale) quanto, piuttosto, il primo formante (grammaticale) e andare a considerarne il valore sintattico.

Nel caso specifico delle occorrenze “se/si/”, per esempio, appare poco significativo notare se queste siano seguite da un formante lessicale (sevai) o da uno grammaticale (sene), quanto piuttosto individuare la funzione sintattica che il primo formante svolge all’interno della frase. Le particelle “se/si” rappresentano il caso più frequente (183 casi) di iposegmentazione tra due formanti e, in tutti i casi, le particelle svolgono funzione di pronomi riflessivo (siè, sifa, sene); non vi sono esempi di iposegmentazione nel momento in cui la particella “se” svolge funzione di congiunzione.

Anche nel caso della “e” - che presenta una frequenza di 112 casi - non si trovano differenze significative osservando la natura del secondo formante (“ela”; “eallora”; “esveglia”).

Focalizzando l’attenzione sui casi di iposegmentazione che coinvolgono “lo/la/gli/le/l’” (205 casi) è possibile rilevare che la frequenza

maggiore (153 casi) si presenta nel momento in cui tali occorrenze svolgono la funzione di pronomi clitici complemento (“lamette”; “lo-prende”, “glidice”). Le forme che coinvolgono l’uso degli articoli (52 casi) risultano invece meno frequenti (“lamamma; “laltro”, “gliamici”).

Un affondo particolare, date le caratteristiche della lingua italiana, deve essere fatto relativamente alle iposegmentazioni che riguardano l’uso dell’apostrofo. All’interno del corpus sono stati individuati 522 lemmi apostrofati di cui 159 scritti convenzionalmente e 363 iposegmentati; ciò significa che il 34,05% dei casi di iposegmentazione si presenta in corrispondenza dell’apostrofo. Tale percentuale piuttosto alta necessita - ancora una volta - di essere spiegata osservando attentamente le occorrenze lessicali coinvolte.

Apostrofo	Frequenza	Percentuale
c’	259	71,35%
l’(quell’, sull’, dall’, nell’)	90	24,80%
altro (es: d’)	14	3,85%
Totale	363	100%

Tab. 6 - Distribuzione dei casi di iposegmentazione legati all’apostrofo.

Come di evince dalla Tab. 6 il peso determinante nei casi di iposegmentazione legati all’apostrofo è dovuto al caso “cera”. Maggiormente significativo, tuttavia, per le implicazioni didattiche che questo comporta, risulta analizzare i casi di iposegmentazione in corrispondenza del lemma “l” nella sua forma semplice e composta (quell’/sull’/nell’/dall’). Esaminando le 90 iposegmentazioni presenti (Tab. 7) si evidenzia che i casi in

cui tali lemmi sono seguiti da un sostantivo (“luva, lequilibrio, lora) si presentano con una frequenza molto bassa (9) mentre i problemi maggiori si riscontrano quando, a seguire, vi è un verbo (laveva, lanno preso) oppure aggettivi e pronomi (quellaltro).

Lemma: l’, nell’, all’, sull’, dall’, quell’	Iposegmentazione
lemma+nome	9
lemma+verbo	53
lemma+agg/pronome	28
Totale	90

Tab. 7 - Distribuzione dei casi di iposegmentazione legati al lemma “l” apostrofato.

Similmente a quanto sopra descritto per i casi “lo/la/gli/le”, nel momento in cui il lemma “l” svolge funzione di pronomi clitico complemento, i casi di iposegmentazione sono più frequenti. Tale dato (che necessita di essere verificato attraverso altre indagini poiché riferito al lessico utilizzato in questi testi) porta a riflettere sulla didattica legata all’uso dell’apostrofo che, comunemente, viene affrontato presentando agli alunni, quasi esclusivamente, esempi che coinvolgono la lettera “l” come articolo seguito da un sostantivo (l’ape, l’erba, l’orso, l’imbuto, l’uva). Basta osservare un libro di testo della scuola primaria per verificare quanto questi casi siano frequenti se non unici. Il problema dell’uso corretto dell’apostrofo appare dunque, per gli alunni, più ampio e complesso rispetto a come, tradizionalmente, viene affrontato; la pratica di far elidere gli articoli “la/lo” seguiti da un sostantivo non sembra essere sufficiente per spiegare la complessità del fenomeno. Inoltre, il caso di “cera” precedentemente analizzato evidenzia come

l'uso dell'apostrofo debba essere appreso attraverso la pratica della scrittura di testi poiché gli esercizi specifici raramente propongono casi in cui l'elisione coinvolge la "c", così come i pronomi clitici complemento (la/lo), oppure situazioni in cui l'articolo è seguito da un aggettivo o da un pronome (l'altro giorno, l'unico, ecc.).

Il problema dell'apostrofo appena descritto porterebbe ad affermare che questo abbia un ruolo predominante nel determinare i casi di iposegmentazione; tuttavia, avendo a disposizione anche i dati riferiti alla segmentazione delle parole nel corpus francese e spagnolo (Ponton *et al.*, 2021) è possibile ipotizzare che l'apostrofo non sia la causa principale delle iposegmentazioni poiché nella lingua spagnola - in cui questa marca ortografica è assente - esse sono percentualmente superiori rispetto ai testi francesi e italiani dove la convenzione ortografica richiede l'uso dell'apostrofo.

4.2. Le ipersegmentazioni e le ipoipersegmentazioni: analisi dei casi più frequenti

Come affermato in precedenza, il costrutto di formante non risulta significativo per stu-

diare le ipersegmentazioni che, conseguentemente, sono state analizzate individuando quali sono i segmenti maggiormente separati e quali caratteristiche presentano. La Tab. 8 mostra i casi più frequenti con i relativi esempi.

Come già individuato dalla ricerca di Ferreiro (1996), in italiano il fenomeno della ipersegmentazione coinvolge principalmente la preposizione "in" che viene considerata come un'entità autonoma e dunque separata dalla parola di cui, convenzionalmente, fa parte. A seguire, anche attorno alla preposizione "a" si concentrano i casi più frequenti di ipersegmentazione; a differenza di quanto emerge dalla ricerca di Ferreiro *et al.*, (1996), invece, le lettere "e/i" non sono coinvolte frequentemente nei fenomeni di ipersegmentazione. Osservando la tabella, è facilmente rilevabile come, ad eccezione della congiunzione "e", le preposizioni appaiono come la categoria maggiormente implicata nelle ipersegmentazioni. Inoltre, analizzando gli esempi che coinvolgono la preposizione "da", si desume che, ciò che la norma ortografica ha unito (preposizione articolata) viene tuttavia separato dagli alunni (da + gli = dagli; da + i = dai; da + la = dalla). I rimanenti casi (56) non presentano una frequenza significativa

Casi di ipersegm.	Frequenza	Esempi
in	56	in sieme, in contro, in vece, in ciampò
a	20	a veva; a desso
da	9	da i; da gli; da la
e	7	e ra
di	5	di etro; di ceva
con	4	con tenti; con solò
al	4	al lora; all'ora
altro	56	fa to; de cise: mi se
Totale	161	

Tab. 8 - Distribuzione dei casi di ipersegmentazione.

ma riguardano comunque sequenze grafiche autonome (fa, de, mi). Rispetto alle iposegmentazioni, il margine di errore nel determinare i casi di ipersegmentazione è risultato maggiore a causa dell'uso nei testi del carattere stampato maiuscolo (77,99%) e stampato minuscolo (8,68%). Dal momento che, in questi caratteri, le lettere non sono legate, in alcuni testi è stato complicato valutare se si trattasse di una separazione tra lettere appartenenti a una stessa parola o tra due parole; le iposegmentazioni sono risultate invece più semplici da disambiguare poiché gli alunni, anche nello stampato maiuscolo, tendevano ad avvicinare in modo molto evidente le due parole.

Infine, per un'analisi completa del corpus, seppur la frequenza sia solamente di 40 occorrenze, è opportuno focalizzare l'attenzione anche sulle ipoipersegmentazioni. Nonostante sia complesso trovare un criterio che accomuni i casi di ipoipersegmentazione, quelli più frequenti (16) coinvolgono, ancora una volta, la preposizione articolata "dal/dai/dagli" (da gialtri, dai gialtri, da isuoi); gli alunni tendono, come nel caso delle ipersegmentazioni, a separare una sequenza grafica autonoma (da/dai/ai) e, successivamente, sembrano seguire maggiormente le caratteristiche foniche (es: in a gialtri; da gialtri, l'accento tonico cade sulla "a" della preposizione). Probabilmente, anche la ipoipersegmentazione che coinvolge i lemmi "l'aveva", che gli alunni scrivono come "la veva", può essere determinata da motivi fonici e dal fatto che "la" è riconosciuto maggiormente come unità grafica autonoma piuttosto che "l'". Tuttavia, a causa della bassa frequenza dei casi, risulta difficile definire le caratteristiche

di queste ipoipersegmentazioni e ipotizzare le ragioni per cui vengono compiute. Sarà necessario attendere i risultati delle classi successive per comprendere maggiormente il fenomeno e capire se si tratta, in classe prima, di un'esplorazione che gli alunni fanno rispetto ai criteri e alle possibilità di segmentazione delle parole, oppure se vi siano ragioni legate alla tipologia dei segmenti separati.

5. Conclusioni

Il contributo qui presentato ha voluto, da un lato, illustrare la costruzione di un ampio corpus digitalizzato di testi di alunni di classe prima e, dall'altro, analizzare la segmentazione delle parole in un'ottica psicogenetica volta a comprendere, attraverso la descrizione del fenomeno, le ragioni delle segmentazioni non convenzionali.

Relativamente al primo obiettivo, l'indagine ha evidenziato la complessità di condurre ricerche su ampi corpora costituiti da testi prodotti da alunni la cui scrittura non è ancora convenzionale: il trattamento automatizzato della lingua ha richiesto una molteplicità di fasi lunghe ed elaborate, possibili grazie alla interdisciplinarietà dell'equipe di ricerca costituita da linguisti computazionali, psicolinguisti, pedagogisti ed esperti di didattica della lingua. La complessità insita nel disegno di ricerca non sarebbe stata pertanto affrontabile se non attraverso un disegno di ricerca capace di superare modelli interpretativi fondati su un sapere cristallizzato in discipline (Morin, 1994). L'interconnessione continua tra le diverse discipline ha reso maggiormente intelligibile il fenomeno grazie a una contaminazione feconda di linguaggi e sguardi che

ha consentito anche di limitare la soggettività dell'interpretazione del fenomeno stesso.

Rispetto al secondo obiettivo, l'indagine qui descritta ha confermato che la segmentazione delle parole costituisce un problema anche per quegli alunni che, al termine della classe prima, hanno già compreso il principio alfabetico del nostro sistema di scrittura: tre alunni su quattro, infatti, hanno prodotto almeno un caso di segmentazione non convenzionale all'interno del proprio testo. Ciò implica che tale fenomeno debba diventare oggetto di attenzione nel dibattito sull'acquisizione dell'ortografia poiché l'analisi delle segmentazioni indebite ha messo in luce come queste presentino delle caratteristiche specifiche e non vengano compiute "casualmente" dagli alunni; ciò che la norma ortografica sanziona come "errore" nasconde idee precise che gli alunni elaborano per comprendere il nostro sistema di scrittura.

L'analisi presentata, avvalendosi anche delle precedenti ricerche internazionali (Gombert, 1990; Ferreiro *et al.*, 1996; Correa & Dockrell, 2007; Piacente & Querejeta, 2012, Querejeta *et al.*, 2013; Ugarte, & Díaz Argüero, 2017) ha messo in luce come il costrutto di *formante morfologico* risulti necessario ma non sufficiente per comprendere le caratteristiche delle segmentazioni non convenzionali. In particolar modo, nel caso delle iposegmentazioni - che rappresentano anche nel nostro corpus il fenomeno più frequente - è stato possibile riscontrare come queste si concentrano laddove il primo formante è grammaticale, con scarsa rilevanza rispetto alla natura del secondo formante (lessicale o grammaticale). Integrando allora i risultati presentati nella ricerca di Querejeta *et al.*,

(2013) con quelli precedentemente elaborati da Ferreiro *et al.* (1996) è plausibile affermare che attorno ai formanti grammaticali si concentra il maggior numero di iposegmentazioni poiché questi sono costituiti da poche lettere e non hanno una funzione referenziale autonoma. Detto con le parole dei bambini, tali occorrenze lessicali vengono iposegmentate poiché "sono corte e non vogliono dire nulla". Quanto affermato porterebbe a pensare che, conseguentemente, le ipersegmentazioni presentino caratteristiche diverse rispetto alle iposegmentazioni; tuttavia, anche in questo caso, ciò che viene ipersegmentato sono principalmente le preposizioni (in particolar modo "in") che sono costituite da poche lettere e, anch'esse, non possiedono una funzione referenziale autonoma. Concordiamo allora con l'ipotesi di Ferreiro *et al.*, (1996) secondo la quale «la presenza esclusiva di qualche ipersegmentazione sia un segno positivo. Considerando che la maggior parte delle ipersegmentazioni corrisponde all'identificazione di sequenze grafiche autonome, possiamo concludere che i bambini che presentano solo ipersegmentazioni hanno già costituito un buon repertorio di dette sequenze grafiche» (p. 75). Questa ipotesi apre a ulteriori piste di ricerca che non è stato possibile, per ora, intraprendere: da un lato sarà necessario osservare se i 29 testi con solo ipersegmentazioni abbiano caratteristiche peculiari (per esempio, se la media degli errori ortografici fosse inferiore rispetto agli altri testi, significherebbe che questi alunni possiedono una maggior padronanza ortografica confermando che le ipersegmentazioni sono un segno positivo nel processo evolutivo di acquisizione del sistema or-

tografico) e, dall'altro, se in classe seconda aumenti la frequenza di ipersegmentazioni a scapito delle iposegmentazioni.

Inoltre, un'ulteriore analisi per comprendere in modo più approfondito il problema dovrà essere effettuata su quel 13,33% di testi scritti in corsivo al fine di verificare se il tipo di carattere possa incidere sulla segmentazione. Per ora, i testi francesi e spagnoli a nostra disposizione, scritti quasi esclusivamente in corsivo, non mostrano differenze significative.

Infine, lo studio delle segmentazioni non convenzionali conduce a delle riflessioni prettamente didattiche rispetto all'insegnamento della scrittura.

Da un lato è evidente come sia la scrittura a definire il concetto di parola intesa come insieme di lettere separate da uno spazio bianco e come il concetto di parola non possa essere dato per scontato. Dal punto di vista didattico questo porta a evidenziare come un percorso di acquisizione della scrittura che proceda dal testo alla parola (e non viceversa) possa esporre direttamente gli alunni ai problemi ortografici (tra cui la segmentazione tra le parole) del nostro sistema di scrittura. Partire dalla testualità significa dare agli alunni, fin dai primi giorni di scuola della classe prima, la possibilità di scrivere testi autonomamente, così da comprendere sia la funzione sociale della scrittura sia i problemi che questa comporta (Teruggi, 2019; Castoldi, 2019; Cisotto, 2019; Tolchinsky, 2008; Ferreiro, 2003; Orsolini & Pontecorvo, 1991), senza parcellizzare l'insegnamento in monodosi fondate su criteri stabiliti dall'adulto.

Dall'altro, un discorso specifico deve essere affrontato per l'apostrofo attorno a cui, in italiano, si concentrano frequenti iposeg-

mentazioni (anche se queste sono presenti in percentuale superiore in spagnolo dove non è previsto l'uso dell'apostrofo) (Ferreiro *et al.*, 1996; Ponton *et al.*, 2021). I dati a nostra disposizione evidenziano come il maggior numero di segmentazioni indebite legate all'apostrofo si presenta quando il primo formante svolge la funzione di pronomi personale complemento ("laveva, lincontrò) e non quando assume la funzione di articolo. Se tale dato fosse confermato da ulteriori investigazioni, sarebbe opportuno riflettere sulle comuni pratiche scolastiche basate esclusivamente su esempi che richiedono l'elisione della vocale dell'articolo quando questo si trova di fronte a sostantivi che iniziano per vocale.

Lo scopo ultimo, dunque, di questa investigazione e di tutti i disegni di ricerca volti a comprendere le modalità di acquisizione della lingua da parte degli alunni, consiste nel ricercare una coerenza sempre maggiore tra le modalità con cui gli alunni apprendono a scrivere e le pratiche di insegnamento messe in atto dai docenti. Un principio, questo, che garantisce quella *preservazione del senso dei saperi* (Chevallard, 1991) - fulcro della didattica delle discipline - non solo dal punto di vista dell'oggetto di insegnamento ma anche dalla prospettiva del soggetto che sta apprendendo.

Bibliografia

- Angelelli, P., Notarnicola, A., Costabile, D., & Marinelli, C. V.** (2008). *DDO. Diagnosi dei disturbi ortografici in età evolutiva*. Trento: 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), pp. 49-66.
- Barbagli, A., Lucisano, P., Dell'Orletta, F., Montemagni, S., Venturi, G.** (2014). Tecnologie del linguaggio e monitoraggio dell'evoluzione delle abilità di scrittura nella scuola secondaria di primo grado. In R. Basili, A. Lenci & B. Magnini (Eds.), *The First Italian Conference on Computational Linguistics CLIC-it 2014* (pp. 23-27). Pisa: University Press.
- Barry, C.** (1994). Spelling routes (or roots or rutes). In G.D.A. Brown & N.C. Ellis (Eds.), *Handbook of spelling: Theory, process and intervention* (pp. 27-49). Chichester, UK: John Wiley.
- Borghini, C.** (2011). Valutazione e misurazione automatizzata della produzione scritta. *Giornale italiano della ricerca educativa*, 4 (7), pp. 11-22.
- Bozzo, M. T., Pesenti, E., Siri, S., & Usai, M. C.** (2000). *CEO - Classificazione degli errori ortografici*. Trento: Erickson.
- Castoldi, M.** (2019). Osservare e valutare i primi apprendimenti sulla lettura e scrittura: spunti da una ricerca. *RicercaAzione*, 11 (1), pp. 17-32. DOI: [10.32076/RA11101](https://doi.org/10.32076/RA11101) [Accessed 03.12.22].
- Castoldi, M., & Chicco, M.** (a cura di) (2017), *Imparare a leggere e a scrivere. Lo stato dell'arte*. Trento: Iprase.
- Chevallard, Y.** (1991). *La trasposizione didactique. De savoir savant au savoir enseigné*. Grenoble: La Pensée Sauvage.
- Cignetti, L., Demartini, S. & Fornara, S.** (a cura di) (2016). *Come Tiscrivo? La scrittura a scuola tra teoria e didattica*. Roma: Aracne.
- Cisotto, L.** (2019). Insegnare a leggere e a scrivere: strategie, pratiche, contesti. *RicercaAzione*, 11 (1), pp. 33-52. DOI: [10.32076/RA11102](https://doi.org/10.32076/RA11102) [Accessed 03.12.22].
- Colombo, A.** (2011). «A me mi». *Dubbi, errori, correzioni nell'italiano scritto*. Milano: Franco Angeli.
- Correa, J., & Dockrell, J. E.** (2007). Unconventional word segmentation in Brazilian children's early text production. *Reading and Writing*, 20, pp. 815-831. DOI: [10.1007/s11145-006-9049-3](https://doi.org/10.1007/s11145-006-9049-3) [Accessed 03.12.22].
- Di Tullio, A.** (1997). *Manual de gramática del español*, Buenos Aires: Edicial.
- Dyson, A. H.** (1993). *Social Worlds of Children Learning to Write*. New York: Teachers College Press.
- Ehri, L. C.** (1999). Phases of Development in Learning to Read Words. In J. Oakill & R. Beard (Eds.), *Reading Development and Teaching of Reading: A psychological perspective* (pp. 79-108). Oxford: Blackwell.
- Eynard, R.** (1992). *La scrittura nella scuola dell'obbligo*. Teramo: Lisciani & Giunti.
- Farina, E.** (2014). *Il dettato nella scuola primaria. Analisi di una pratica di insegnamento*. Milano: Franco Angeli.
- Ferreiro, E., & Teberosky, A.** (1985). *La costruzione della lingua scritta nel bambino*. Firenze: Giunti. Ed. Or. (1979).
- Ferreiro, E., & Vernon, S.** (1992). La distincion palabra/nombre en niños de 4 y 5 años. *Infancia y Aprendizaje*, 58, pp. 12-28.
-

- Ferreiro, E.** (a cura di) (1996). *Cappuccetto Rosso impara a scrivere: studi psicolinguistici in tre lingue romanze*. Firenze: La Nuova Italia.
- Ferreiro, E.** (2003). *Alfabetizzazione. Teoria e pratica*. Milano: Raffaello Cortina.
- Frith, U.** (1985). Beneath the surface of developmental dyslexia. In K.E. Petterson, J. Marshall, M. Coltheart (Eds.), *Surface dyslexia*, (pp. 301-330). Hillsdale, NJ: Erlbaum.
- Gaulmyn, M. M., Gonnand, S., & Luis, M. H.** (1996). «S'il te plait, écris-moi l'histoire du Petit Chaperon Rouge... ». Costruzione de la norme orthographique et grammaticale du CP au CE2. *Repères*, 14, pp. 121-140.
- Goigoux, R.** (éd.) (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*. Paris: Puf.
- 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.
- Orsolini, M., & Pontecorvo, C.** (a cura di) (1991). *La costruzione del testo scritto nei bambini*. Firenze: La Nuova Italia.
- Marinelli, C. V., Judica, A., Cucciaioni, C., Verni, F., Deidda, C., Notarnicola, A., Iaia, M., & 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, pp. 424-449.
- Morin, E.** (1994). *Sur l'interdisciplinarité*. Centre international de recherches et études transdisciplinaires, in: <https://ciret-transdisciplinarity.org/bulletin/b2c2.php>. [Accessed 10/06/2022]
- Pascucci, M., Pontecorvo, C., & Zuccheromaglio C.** (1986). *Guida alla lingua scritta: per insegnanti della scuola elementare e dell'infanzia*. Roma: Editori Riuniti.
- Patterson, K. E.** (1986). Lexical but nonsemantic spelling. *Cognitive Neuropsychology*, 3, pp. 341-367.
- Piacente, T., & Querejeta, M.** (2012). La separación entre palabras en la escritura infantil, *Revista Neuropsicología Latinoamericana*, 4 (1), pp. 1-17. DOI:10.5579/rnl.2012.0098 [Accessed 03.12.22].
- Pinto, G., Accorti Gamannossi, B., & Camilloni, M.** (2017). Percorsi di alfabetizzazione emergente. La competenza notazionale. In M. Castoldi, & M. Chicco (a cura di), *Imparare a leggere e a scrivere. Lo stato dell'arte* (pp. 29-46). Trento: Iprase.
- Ponton, C., Gutiérrez-Cáceres, R., Teruggi, L., Farina, E., Brissaud, C., & Wolfarth C.** (2021). Scolinter: un corpus trilingue. L'exemple de la segmentation en mots. *Langue Française*, 3, (211), pp. 37-50. DOI: 10.3917/lf.211.0037 [Accessed 03.12.22].
- Querejeta, M., Piacente, T., Guerrero Ortiz-Hernán, B., & Alva Canto, E. A.** (2013). La separación entre palabras en la escritura de niños que inician la escolaridad primaria. *Interdisciplinaria*, 30, (1), pp. 45-64.
- Sartori, G., Job, R., & Tressoldi, P. E.** (2007). *Batteria per la valutazione della dislessia e della disortografia evolutiva*. Firenze: Giunti OS.
- Serianni, L., & Benedetti, G.** (2009). *Scritti sui banchi. L'italiano a scuola tra alunni e insegnanti*. Roma: Carocci.
- Teruggi, L.** (2019). *Leggere e scrivere a scuola: dalla ricerca alla didattica*. Roma: Carocci.
- 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* (pp. 135-158). Rotterdam: Sense Publishers.

- Tolchinsky, L.** (2008). Usar la lengua en la escuela. *Revista iberoamericana de educación*, 46, pp. 37-54.
- Tressoldi, P.E., Cornoldi, C., & Re, A. M.** (2012). *Batteria per la Valutazione della Scrittura e della Competenza Ortografica - 2*. Firenze: Giunti OS.
- Ugarte, A., & 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. *RevistaSignos*, 50 (93), pp. 3-25. DOI: [10.4067/S0718-09342017000100001](https://doi.org/10.4067/S0718-09342017000100001) [Accessed 03.12.22].
- Wolfarth C., Brissaud, C., & Ponton, C.** (2018). Transcrire et normer un corpus scolaire: pour quelles analyses ? *Dyptique*, 36, pp. 121-145. DOI: [10.3917/lf.211.0037](https://doi.org/10.3917/lf.211.0037) [Accessed 03.12.22].
- Wolfarth, C.** (2019). *Apport du TAL à l'exploitation linguistique d'un corpus scolaire longitudinal*, Thèse de doctorat en Sciences du langage de l'Université Grenoble Alpes.