



UNIVERSIDADE ESTADUAL PAULISTA  
“JÚLIO DE MESQUITA FILHO”  
Campus de São José do Rio Preto

Roberta Pereira Fiel

ESTUDO LONGITUDINAL DE HIPOSEGMENTAÇÕES EM TEXTOS  
DO ENSINO FUNDAMENTAL II

São José do Rio Preto  
2018

Roberta Pereira Fiel

ESTUDO LONGITUDINAL DE HIPOSEGMENTAÇÕES EM TEXTOS  
DO ENSINO FUNDAMENTAL II

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Estudos Linguísticos, junto ao Programa de Pós-Graduação em Estudos Linguísticos, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de São José do Rio Preto

Financiadora: Fapesp (Processo N° 2015/26763-6)

Orientadora: Profa. Dra. Luciani Ester Tenani

São José do Rio Preto  
2018

Fiel, Roberta Pereira.

Estudo longitudinal de hipossegmentações em textos do ensino fundamental II / Roberta Pereira Fiel. -- São José do Rio Preto, 2018  
125 f. : il.

Orientador: Luciani Ester Tenani

Dissertação (mestrado) – Universidade Estadual Paulista  
“Júlio de Mesquita Filho”, Instituto de Biociências, Letras e Ciências  
Exatas

1. Linguística. 2. Ensino fundamental. 3. Escrita. 4. Palavra  
(Linguística) 5. Hipossegmentações. I. Universidade Estadual Paulista  
"Júlio de Mesquita Filho". Instituto de Biociências, Letras e Ciências  
Exatas. II. Título.

CDU – 41

## ERRATA

FIEL, Roberta Pereira. **Estudo longitudinal de hipossegmentações em textos do Ensino Fundamental II**. 2018.125 f. Dissertação (Mestrado em Estudos Linguísticos) - Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista "Júlio de Mesquita Filho", São José do Rio Preto, 2018.

<b>Folha</b>	<b>Linha</b>	<b>Onde se lê</b>	<b>Leia-se</b>
3	3	Por fim, agradeço à FAPESP pelo financiamento desta pesquisa	Por fim, agradeço à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) pelo financiamento desta pesquisa (processo nº 2015/26763-6)

Roberta Pereira Fiel

ESTUDO LONGITUDINAL DE HIPOSEGMENTAÇÕES EM TEXTOS DO  
ENSINO FUNDAMENTAL II

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Estudos Linguísticos, junto ao Programa de Pós-Graduação em Estudos Linguísticos, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Campus de São José do Rio Preto

Financiadora: Fapesp (Processo N° 2015/26763-6)

Orientadora: Profa. Dra. Luciani Ester Tenani

Comissão Examinadora

Profa. Dra. Luciani Ester Tenani

Profa. Dra. Cristiane Carneiro Capristano

Prof. Dr. Lourenço Chacon Jurado Filho

São José do Rio Preto  
22 de março de 2018

À Sophia, para que um dia também seja contagiada pelo desejo de fazer ciência!

## **AGRADECIMENTOS**

A realização e concretização deste trabalho somente foi possível graças a duas grandes mulheres: minha querida orientadora, Luciani Ester Tenani, e minha amada mãe, Cidinha Fiel. Digo isso, pois é nelas que encontro a força e o exemplo de mulher que quero ser. À Luciani, agradeço pela dedicação, paciência e ensinamentos; à minha mãe, agradeço pelo cuidado, amor e compreensão.

Tenho muito a agradecer também a meu amigo e companheiro de vida, Renan Rosa Silva, com quem dividi as dificuldades e as alegrias do processo de desenvolvimento desta pesquisa.

Agradeço a meu pai, Roberto Fiel, e à minha irmã, Soraya Fiel, pelo incentivo e apoio. Agradeço ainda à Eliane e ao Antônio Carlos pelo carinho e acolhimento.

Manifesto também o meu agradecimento à professora Cristiane Carneiro Capristano e ao professor Lourenço Chacon pelas leituras atenciosas e pelas inestimáveis contribuições na qualificação e na defesa. Igualmente, agradeço à professora Ana Ruth Moresco Miranda por ter lido e debatido este trabalho no IX Seminário de Estudos Linguísticos da Unesp e pelas sugestões de melhoria.

Da mesma forma, sou grata à professora Fabiana Cristina Komesu pela amizade e por ter aceitado ser membro suplente de minha banca de qualificação e de defesa; e à professora Idalice Rillo que, com muita paciência, me auxiliou no desenvolvimento da análise estatística que apresento neste trabalho.

Não poderia deixar de agradecer aos membros da linha da pesquisa Oralidade e Letramento, em especial à Tainan Garcia, à Amanda Gatto, à Giovanna Alves e à Carina Maciel pelo companheirismo nos momentos de estudo e nos momentos de dificuldades.

Minha gratidão aos meus amigos, em especial, Jean Michel Pimentel Rocha, Lilian Maria da Silva, Gabriela de Oliveira Codinhoto pela disposição em sempre tirar minhas dúvidas e me ajudar.

Meu muito obrigada aos meus amigos de pós-graduação Helker Nhoato, Aline Gomes Garcia, Luísa Ferrari, Luana Bianchi, Ana Luiza Ferancini e Bárbara Fante pelo companheirismo ao longo desses dois anos.

Agradeço ainda aos demais amigos de graduação Natália Trigo, Dibo Mussi Neto, Anderson Ferris, Odair Júnior, Eloísa Moriel, Ingrid Zanata, Monelise Vilela, Mariana de

Fátima, Cíntia Braga e Bárbara Assis que, mesmo não tendo participado diretamente da elaboração deste trabalho, proporcionaram bons momentos de descontração.

Por fim, agradeço à FAPESP pelo financiamento desta pesquisa, possibilitando minha dedicação à produção deste trabalho.

## RESUMO

Esta dissertação trata da caracterização longitudinal da escrita de alunos do EF II no que diz respeito às chamadas hipossegmentações de palavras escritas – como “puraqui” (“por aqui”), nas quais há a ausência não-convencional de fronteira gráfica. Nossos objetivos são: (i) identificar, por meio de análise quantitativa, se há correlação (ou não) entre número de hipossegmentações e tempo de escolarização; e (ii) descrever qualitativamente as hipossegmentações, quanto a aspectos prosódicos dos enunciados falados e aspectos gráficos relativos às informações da própria convenção ortográfica. Para alcançar esses objetivos, nos baseamos, por um lado, em aparato teórico da fonologia prosódica, modelo *relation-based*, que concebe a existência de sete constituintes prosódicos que estruturam os enunciados das línguas do mundo; por outro lado, em abordagem da escrita como constituída de modo heterogêneo. Dos resultados obtidos na análise quantitativa, destacamos que há correlação entre aumento dos anos de escolarização e diminuição de ocorrência de hipossegmentação. No que se refere aos resultados quantitativos das estruturas envolvidas, destacamos: (i) a junção entre clítico e palavra prosódica é a característica do maior conjunto de dados; (ii) a junção entre dois clíticos é a segunda estrutura mais recorrente, predominando a hipossegmentação “oque”; (iii) a junção entre duas palavras prosódicas, a terceira mais recorrente no material analisado, decorre da mobilização de várias características linguísticas, como a hipossegmentação de estruturas perifrásticas que constituem exemplos de mudança linguística em curso; (iv) a junção de palavra prosódica e clítico é a estrutura menos recorrente, sendo a maioria dos dados decorrente da combinação de palavras com a ausência do hífen, que levou à formação de possíveis palavras prosódicas; (v) a junção envolvendo mais de uma palavra prosódica e/ou clítico ocorreu apenas em três dados, que abrangem estruturas como a frase entoacional e o enunciado fonológico. No que se refere aos resultados qualitativos, a partir de análise de cunho linguístico-textual, os casos em que há a flutuação entre convencional e não-convencional: (i) se distinguem entre si pela configuração prosódica, gramatical e linguística-textual; (ii) são indícios mais explícitos da inserção dos alunos em práticas orais/faladas e letradas/escritas; (iii) são marcas do complexo processo que envolve o Outro como instância representativa da linguagem (e da escrita em particular), a escrita na complexidade de seu funcionamento (heterogeneamente constituída) e o aluno enquanto sujeito escrevente. A principal contribuição desta dissertação está em: (i) fazer análise quantitativa e qualitativa de hipossegmentações no EF II; e (ii) evidenciar a complexidade que subjaz às relações entre prosódia e escrita por meio da segmentação não-convencional de palavras.

Palavras-chave: Palavra. Segmentação não-convencional. Hipossegmentação. Prosódia. Práticas orais/faladas. Práticas letradas/escritas, Ensino Fundamental II.

## **ABSTRACT**

*This work deals with the longitudinal characterization of the writing by students from Junior High School (EF II in Brazil) with respect to the hyposegmentations of written words in which there is the unconventional absence of graphic frontier (e.g. "puraqui" - "por aqui" in Portuguese - "around here" in English). To reach these objectives, we are based on a theoretical apparatus of prosodic phonology, on the one hand, a relation-based model, which conceives the existence of seven prosodic constituents that structure the utterances of the world's languages; and, on the other hand, in the approach of writing as constituted in a heterogeneous way. From the results obtained in the quantitative analysis, we highlight that is a correlation between the increase in the years of schooling and a decrease in the occurrence of hyposegmentations. Regarding the quantitative results of the structures, we highlight: (i) the junction between clitic and prosodic word is the characteristic of the largest data set; (ii) the junction between two clitics is the second most recurrent structure; (iii) the junction between two prosodic words, the third most recurrent in the material analyzed, derives from the mobilization of several linguistic characteristics, such as the hyposegmentation of periphrastic structures that are examples of linguistic change in progress; (iv) the prosodic and clitic word junction is the least recurrent structure, most of which results from the combination of words with the absence of the hyphen, which led to the formation of possible prosodic words; (v) the junction involving more than one prosodic and / or clitic word occurred only in three data, covering structures such as the intonational phrase and phonological utterance. Regarding the qualitative results, from a linguistic-textual analysis, we highlight that the cases in which there is a fluctuation between conventional and unconventional are: (i) distinguished by their prosodic, grammatical and linguistic-textual configuration ; (ii) more explicit indications of students' insertion into oral / spoken practices and literacy / written practices; (ii) marks of the complex process involving the Other as an instance representative of language (and writing in particular), writing in the complexity of its functioning (heterogeneously constituted) and the student as a writing subject. The main contribution of this work is: (i) to make a quantitative and qualitative analysis of hyposegmentations in by students from Junior High School (EF II) in Brazil Elementary School; and (ii) to show the complexity that underlies the relations between prosody and writing through unconventional segmentation of words.*

**Keywords:** *Word. Unconventional word segmentation. Hyposegmentation. Prosody. Oral/spoken practices. Literacy/written practices. Junior High School (EF II - Brazil).*

## Lista de Figuras

Figura 1. Exemplo do primeiro princípio de constituição da hierarquia prosódica _____	27
Figura 2. Exemplo do segundo princípio da constituição da hierarquia prosódica _____	27
Figura 3. Estruturas n-árias e binárias _____	28
Figura 4. Exemplo do quarto princípio de construção da hierarquia prosódica _____	28
Figura 5. Hierarquia prosódica segundo Nespore Vogel (1986, 2007) _____	29
Figura 6. O grupo clítico de Nespore Vogel (2007) _____	31
Figura 7. Exemplo de critério de identificação de dados de segmentação não-convencional	48
Figura 8. Exemplo de critério de identificação de dados de segmentação não-convencional	49
Figura 9. Exemplo de critério de identificação de dados de segmentação não-convencional	49
Figura 10. Distribuição dos dados de hipossegmentação no decorrer do EF II _____	56
Figura 11. Média de hipossegmentações ao longo do EF II _____	57
Figura 12. Coeficiente de hipossegmentação de 115 alunos _____	59
Figura 13. Coeficiente de hipossegmentação de 92 alunos _____	60
Figura 14. Distribuição de outliers nos quatro anos do EF II _____	62
Figura 15. Texto em que ocorreu o dado “seila” _____	79
Figura 16. Texto em que ocorreu o dado “enportapara” _____	81
Figura 17. Texto em que ocorreu o dado “fruipra” _____	82
Figura 18. Enunciado da proposta - Cordel (6º ano - 2008) _____	92
Figura 19. Texto "O rei do Hip-Hop" _____	93
Figura 20. Texto “A minha vida” _____	94
Figura 21. Flutuação de segmentação de “o que” _____	98
Figura 22. Enunciado da proposta de redação - Internet e o uso do MSN (6º ano - 2008) ____	99
Figura 23. Texto com flutuação de segmentação da estrutura “o que” _____	100
Figura 24. Enunciado da proposta - Massacre em Realengo (9º ano - 2011) _____	102

## Lista de quadros

Quadro 1. Hipossegmentações com processo de alçamento .....	70
Quadro 2. Hipossegmentações emprego da letra <m> por <n> .....	71
Quadro 3. Hipossegmentações com processo de degeminação.....	73
Quadro 4. Hipossegmentações sem processo de degeminação .....	73
Quadro 5. Hipossegmentações das sequências "e o" e "e a" .....	75
Quadro 6. Hipossegmentações com redução do verbo "estar" .....	77
Quadro 7. Hipossegmentações da sequência "para mim" .....	83
Quadro 8. Hipossegmentações com redução de palavra .....	84
Quadro 9. Hipossegmentações com choque de acento.....	84
Quadro 10. Hipossegmentações de estruturas perifrásticas.....	85
Quadro 11. Hipossegmentações de verbo + pronome .....	86
Quadro 12. Hipossegmentações de verbo + preposição ou artigo.....	87
Quadro 13. Hipossegmentações de advérbio + conjunção, preposição ou pronome .....	89
Quadro 14. Hipossegmentações de mais de uma palavra e/ou clíticos .....	90

## Lista de Tabelas

Tabela 1. IDESP da escola analisada e do estado _____	46
Tabela 2. Número de textos selecionados _____	47
Tabela 3. Quantidade de dados de palavras homônimas _____	51
Tabela 4. Resultado do teste post hoc Tukey para 115 alunos. _____	58
Tabela 5. Coeficiente de hipossegmentação de 92 alunos _____	60
Tabela 6. Resultado do teste post hoc Tukey para 92 alunos. _____	61
Tabela 7. Relação de sujeitos considerados outliers _____	63
Tabela 8. Tipos combinação de hipossegmentações _____	65
Tabela 9. Coeficiente dos dados de hipossegmentação _____	65
Tabela 10. Coeficiente de dados do tipo 1 _____	67
Tabela 11. Possibilidades de combinação de dados do tipo 1 _____	68
Tabela 12. Quantidade de hipossegmentações de "de" e "a" _____	69
Tabela 13. Coeficiente de dados do tipo 2 _____	74
Tabela 14. Número de ocorrência da hipossegmentação oque _____	74
Tabela 15. Coeficiente de dados do tipo 3 _____	76
Tabela 16. Coeficiente de dados do tipo 4 _____	86
Tabela 17. Coeficiente de dados do tipo 5 _____	90

## **Lista de Abreviaturas**

EF: Ensino Fundamental

EF I: Ensino Fundamental Ciclo I

EF II: Ensino Fundamental Ciclo II

PB: Português Brasileiro

PE: Português Europeu

## Lista de Símbolos

$\sigma$	Sílaba
$\Sigma$	Pé métrico
$\omega$	Palavra prosódica
CG	Grupo clítico
$\phi$	Frase fonológica
I	Frase ou sintagma entoacional
U	Enunciado fonológico

## SUMÁRIO

<b>APRESENTAÇÃO</b>	<b>12</b>
<b>CAPÍTULO 1</b>	<b>16</b>
<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>16</b>
1.1 HIPOSEGMENTAÇÃO: POSSÍVEIS INTERPRETAÇÕES	16
1.2 PERSPECTIVA TEÓRICA SOBRE A ESCRITA	22
1.3 PROSÓDIA E HIPOSEGMENTAÇÃO	25
1.4 HIPOSEGMENTAÇÕES NO EF I E NO EF II	33
<b>CAPÍTULO 2</b>	<b>44</b>
<b>MATERIAL E MÉTODOS</b>	<b>44</b>
2.1 MATERIAL ANALISADO	44
2.2 CONSTITUIÇÃO DO OBJETO DE ANÁLISE	47
2.3 FORMAS DE ANÁLISE	52
<b>CAPÍTULO 3</b>	<b>55</b>
<b>ANÁLISE DOS DADOS</b>	<b>55</b>
3.1 HIPOSEGMENTAÇÃO: TENDÊNCIAS QUANTITATIVAS DE DADOS LONGITUDINAIS	55
3.2 HIPOSEGMENTAÇÃO: CARACTERÍSTICAS PROSÓDICAS	64
3.2.1 Tipo 1: Clítico + palavra prosódica	66
3.2.2 Tipo 2: Clítico + clítico	74
3.2.3 Tipo 3: Palavra prosódica + palavra prosódica	76
3.2.4 Tipo 4: Palavra prosódica + clítico	85
3.2.5 Tipo 5: Outras combinações de palavras	90
3.3 CASOS ESPECIAIS DE HIPOSEGMENTAÇÃO	91
3.4 FLUTUAÇÕES ENTRE O CONVENCIONAL E NÃO-CONVENCIONAL	97
<b>CAPÍTULO 4</b>	<b>106</b>
<b>CONSIDERAÇÕES FINAIS</b>	<b>106</b>
<b>REFERÊNCIAS</b>	<b>111</b>
<b>ANEXOS</b>	<b>118</b>
ANEXO 1. Proposta – Internacionalização da Amazônia (9º ano - 2011)	118
ANEXO 2. Proposta – Massacre em Realengo (9º ano – 2011)	121

---

## APRESENTAÇÃO

---

Definir o que é palavra não é uma tarefa simples, uma vez que seus limites podem variar a depender do critério considerado, como demonstram Veloso (2016), Rosa (2000), pois uma palavra pode ser delimitada por critérios gráficos, fonológicos, sintáticos, lexicais, entre outros.

Segundo Rosa (2000), aprender a escrever é, dentre outras coisas, aprender a identificar e a assinalar os limites da palavra gráfica. Esses limites, de acordo com a autora, são caracterizados por espaços em branco, hifens, quebras de linha, letras com traçados diferenciados e sinais de pontuação. No PB, uma das funções do sinal de hífen, por exemplo, é a de ligar elementos de palavras compostas. Assim, a sequência “grande chuva” é formada por duas palavras lexicais, já a sequência “guarda-chuva”, por sua vez, é formada por somente uma palavra composta. Essa mesma sequência, de um ponto de vista morfológico funciona como uma unidade lexical com um único significado e, portanto, configura-se como uma única palavra morfológica. Entretanto, de um ponto de vista rítmico, funciona como duas palavras prosódicas<sup>1</sup>, também denominadas por palavras fonológicas<sup>2</sup> (BISOL, 2004,1996), já que cada uma apresenta acento primário, destacados a seguir: “guarda-chuva”. Essa configuração leva à identificação de duas palavras prosódicas, menores do que a palavra morfológica.

A configuração de a palavra prosódica ser menor do que a palavra morfológica também se verifica em palavras derivadas de afixos acentuados, como “simplesmente”, “belamente”, “pré-estreia”. Para esses casos, Quadros e Schwindt (2008) mostram que o acento da base se mantém, evidenciando a existência de duas palavras prosódicas que correspondem a uma palavra morfológica. Já, na sequência “grande chuva”, a palavra prosódica coincide em tamanho com a palavra morfológica.

No PB, a palavra prosódica também pode ser maior que a palavra morfológica, como mostram Quadros e Schwindt (2008). Essa configuração encontra-se em sequências de palavras e pronome átono, como em “cale-se”, na qual há somente uma palavra prosódica, já que a partícula “se” não apresenta um acento primário. Na hierarquia prosódica, palavra prosódica e

---

<sup>1</sup> Segundo Nespor e Vogel (1986, 2007), a palavra prosódica é o nível da hierarquia prosódica que representa a interação entre os componentes morfológicos e fonológicos de uma gramática e a sua principal característica é portar um único acento primário. Explicações mais detalhadas sobre esse nível são dadas na seção 1.3 deste capítulo.

<sup>2</sup> Bisol (1996), com base em Mattoso Câmara Jr. (1969), usa o termo “palavra fonológica”. Em trabalhos sobre o português Europeu, Vigário (2001) emprega “palavra prosódica” como versão em português de *phonological word*. Neste trabalho, optamos por “palavra prosódica”, pois é a mais difundida em trabalhos que abordam constituintes prosódicos (cf. TENANI, 2016; GUZZO, 2015; FERNANDES-SVARTMAN, 2012; BOOP, 2010).

partícula átona configuram-se como grupo clítico<sup>3</sup>. Nessa perspectiva, não há correspondência entre palavra prosódica e palavra morfológica.

Segundo Rosa (2000), de uma perspectiva gramatical, as palavras também podem ser classificadas de acordo com o seu tipo de significado: lexical ou gramatical. Palavras com significado lexical são aquelas que podem ser interpretadas semanticamente mesmo quando isoladas. Nesse grupo de palavras, encontram-se substantivos, adjetivos, verbos e advérbios. As palavras gramaticais, por outro lado, são aquelas interpretadas a partir da relação sintática que mantêm com as palavras lexicais do enunciado. Nesse grupo, encontram-se artigos, conjunções, pronomes relativos, pronomes reflexivos e preposições. Rosa (2000) explica que, de modo geral, as palavras lexicais formam classes abertas, isto é, classes em que, a princípio, podem ser acrescentadas novas criações, empréstimos linguísticos. As palavras gramaticais, por sua vez, formam classes fechadas, pois as chances de se adicionar um novo artigo ou conjunção a uma língua são muito remotas.

Devido à complexidade morfossintática das palavras gramaticais, os aprendizes, muitas vezes, encontram dificuldade em segmentá-las, como mostram Tenani (2016) e Paranhos (2014), para alunos do EF II. Esses trabalhos que tratam de segmentação de palavras escritas no EF II definem três tipos de ocorrências desse fenômeno:

- i. Hipersegmentação: decorre da presença não-convencional de uma fronteira gráfica no interior da palavra, como na ocorrência “co migo”, para a qual a convenção prevê “comigo”;
- ii. Hipossegmentação: ocorre quando há a ausência não-convencional de fronteira entre as palavras, como em “concerteza”, em que, convencionalmente espera-se “com certeza”;
- iii. Mescla<sup>4</sup>: ocorrência simultânea de hipossegmentação e hipersegmentação na mesma sequência, como, por exemplo, no dado “pora qui”, no qual há, em um primeiro momento, a hipossegmentação da palavra “por” com a primeira sílaba da palavra “aqui” e, posteriormente, a hipersegmentação da palavra “aqui”.

Neste trabalho, serão abordadas hipossegmentações com o objetivo central de descrever e analisar quantitativamente dados de hipossegmentações e apresentar um panorama

---

<sup>3</sup> A definição de grupo clítico, bem como a sua complexidade serão tratados na seção 1.3 desta dissertação.

<sup>4</sup> Em trabalhos sobre a escrita infantil, esse tipo de segmentação não-convencional pode ser chamado de “mescla” (CHACON, 2013) ou “híbrido” (CUNHA, 2010). Neste trabalho, optamos por utilizar a nomenclatura “mescla”, uma vez que nossa perspectiva teórica se aproxima daquela defendida por Chacon (2013).

longitudinal das tendências dessas grafias em textos escritos por alunos que cursaram os quatro anos do Ensino Fundamental II (doravante, EF II) em uma escola pública da cidade de São José do Rio Preto, interior do estado de São Paulo. Vislumbramos observar como aspectos prosódicos dos enunciados falados e aspectos gráficos relativos às informações da própria convenção ortográfica podem ser relevantes para identificar motivações das decisões quanto à segmentação gráfica em palavras escritas.

Ligados ao objetivo geral, nossos objetivos específicos são: i) identificar ocorrências e classificá-las em tipos de estruturas prosódicas (por exemplo, “meajude” = clítico + palavra prosódica) em que ocorrem as hipossegmentações e encontrar tendências; ii) identificar ocorrências e classificá-las em tipos de estruturas morfossintáticas que envolvem hipossegmentações (por exemplo, “meajude” = pronome + verbo) e estabelecer tendências; (iii) identificar se há correlação (estatisticamente significativa ou não) entre número e tipo de hipossegmentações em relação aos anos escolares.

A escolha desse tipo de segmentação se deu por dois motivos: (i) pelos poucos trabalhos que tratem de hipossegmentações em textos do EF II<sup>5</sup> de uma perspectiva longitudinal e (ii) pelo fato de que o corpúsculo, que será apresentado em breve, não ter sido ainda analisado em sua totalidade no que se refere aos dados de hipossegmentação.

Destacamos que o caráter longitudinal desta pesquisa se mostra relevante por permitir uma análise que contemple o percurso dos alunos escreventes quanto à noção de palavra no decorrer dos quatro últimos anos do EF. Abaurre (1991, 1992), ao tratar de dados de escrita infantil, sinalizou que estudos com esse caráter temporal possibilitam a observação do contato do aluno com a escrita convencional, sobretudo com aquela ensinada na escola, mostrando o modo como as crianças propõem segmentações para seus enunciados escritos e os critérios que norteariam essas diferentes proposições. Neste trabalho, o material analisado permitiu desenvolver um estudo longitudinal de hipossegmentações no EF II, permitindo ampliar o olhar sobre como a noção de palavra ortográfica se mostra ao longo do EF, contribuindo, assim, para os estudos desenvolvidos sobre essa temática no âmbito do grupo de pesquisa “Estudo sobre a linguagem”.

No que diz respeito ao modo de organização deste trabalho, no capítulo 1, dedicamos atenção aos subsídios teóricos. Apresentamos possíveis interpretações dadas às segmentações não-convencionais de palavra e explicitamos a perspectiva de interpretação assumida. Adiantamos que, nesta dissertação, interpretamos os dados de segmentação-convencional como

---

<sup>5</sup> Dentre os poucos trabalhos sobre hipossegmentações no EF II, podemos citar Tenani e Fiel (2016); Tenani e Paranhos (2014) e Tenani (2011; 2010, 2008).

evidências de estruturas prosódicas que subjazem às escolhas não aleatórias de representação gráfica de fronteira de palavras, pistas do trabalho do sujeito com a linguagem que revelam a complexidade da segmentação de enunciados escritos (TENANI, 2016; CAPRISTANO, 2007, 2004; CHACON, 2006; 2005, 2004).

Ainda no capítulo 1, versamos sobre a concepção de escrita que possibilita a interpretação dessa complexidade inerente aos dados de segmentação. Logo em seguida, apresentamos os pressupostos do modelo de Fonologia Prosódica de Nespor e Vogel (1986, 2007), na qual nos baseamos para realizar as análises prosódicas. Nesse capítulo, resenhamos trabalhos que se dedicaram à análise de dados de hipossegmentação em textos escritos por alunos do EFI e do EF II.

No capítulo 2, descrevemos os aspectos metodológicos adotados nesta pesquisa. Tecemos considerações sobre o material de investigação, dando ênfase às características do banco de dados de textos do Ensino Fundamental II. Após a descrição do material, discorremos sobre os critérios empregados para estabelecer um recorte do material que privilegiasse nossos objetivos.

No capítulo 3, realizamos a análise dos dados. Inicialmente, apresentamos, com base em um tratamento estatístico, a tendência de ocorrência longitudinal dos dados de hipossegmentação. Em seguida, descrevemos as principais características prosódicas e morfossintáticas que os dados de hipossegmentação mobilizam. Também realizamos uma análise de cunho linguístico-textual de dados selecionados, a fim de evidenciar quão complexo dados de segmentação de escrita podem se revelar.

No capítulo 4, sistematizamos as discussões desenvolvidas, dando destaque àquelas que consideramos as mais relevantes. Para concluir a dissertação, tecemos considerações sobre questões que ficaram em aberto e sobre possíveis encaminhamentos futuros.

---

## CAPÍTULO 1

---

### FUNDAMENTAÇÃO TEÓRICA

Como adiantamos, neste capítulo, apresentamos os subsídios teóricos nos quais nos baseamos para descrever e interpretar os dados de hipossegmentação.

Na seção 1.1, dedicamos a nossa atenção às diferentes perspectivas pelas quais os dados de segmentação de palavras escritas podem ser interpretados e, dentre as abordagens encontradas, explicitamos a que adotamos. A partir da perspectiva de análise adotada, explicitamos, na seção 1.2, a concepção de escrita na qual nos baseamos para analisar os dados de segmentação de palavras.

Para análise de características linguísticas das hipossegmentações, valemo-nos dos pressupostos teóricos da Fonologia prosódica de Nespor e Vogel (1986, 2007), explicitados na seção 1.3. Explicamos, com base nessas autoras, o funcionamento da chamada hierarquia de domínios prosódicos e, dentre esses domínios, aqueles relevantes para identificação de regularidades das hipossegmentações.

Finalizamos o capítulo com resenha de trabalhos que se ocupam da análise de dados de hipossegmentação sob um viés linguístico. Nosso intuito, a partir dessa resenha, é delinear os pontos de partidas para realização deste trabalho.

#### **1.1 Hipossegmentação: possíveis interpretações**

Em uma perspectiva presente em ambiente escolar, apresentar muitos casos de segmentação não-convencional de palavras escritas pode ser considerado um problema de alfabetização ou até mesmo um distúrbio de escrita, como na perspectiva de Zorzi (2006). Segundo esse autor, todas as crianças cometem “erros” durante a aprendizagem da escrita; entretanto, esses erros não podem ser muito frequentes; eles devem diminuir e se tornarem ocasionais, até que haja um total domínio do sistema ortográfico.

Zorzi (2006) afirma ainda que as crianças que apresentam trajetória diferente dessa, exibindo grande diversidade e quantidade de segmentações não-convencionais por um período

longo, podem ser diagnosticadas como portadoras de transtornos de aprendizagem, uma vez que esse tipo de ocorrência é considerado sintoma de déficits no processo de aprendizagem. Quando esses sintomas são percebidos, os professores tendem a encaminhar os alunos para aulas de reforço, com intuito de fazê-los aprender aquilo que não conseguiram em sala de aula, e até para acompanhamento fonoaudiológico, para sanar possíveis problemas de distúrbios de aprendizagem.

Esse tipo de interpretação está ligado a uma concepção de escrita que enxerga interferências da fala na escrita. Essa visão pode ser constatada no trecho seguinte, no qual Zorzi (2006) argumenta sobre os problemas identificados na escrita de alunos:

A oralidade exerce forte influência quando as crianças formulam hipóteses a respeito de como as palavras devem ser escritas, gerando uma tendência muito grande no sentido de que os padrões de fala determinem o modo de escrita. Tudo ocorre como se o sistema de representação gráfica do português fosse fonético, e não alfabético. Isto quer dizer que a criança toma como referência, para determinar que letra deve usar, aqueles sons que ela é capaz de identificar em seu padrão de pronúncia da palavra procedendo então a uma correspondência um a um: som falado / letra que escreve (ZORZI, 2006, p. 8).

Visão dicotômica semelhante a essa adotada por Zorzi (2006) também é mobilizada por discentes de um curso de Licenciatura em Letras chamados a analisar erros e descompassos em produções escritas de alunos do sexto ano do EF II, como mostram Capristano e Angelo (2015). Nesse trabalho, os autores argumentam que é uma visão que vê interferências da fala na escrita que guia o olhar desses futuros professores quando analisam produções escritas de alunos do EF II. Fala e escrita são concebidas como modalidades linguísticas dicotômicas.

Sob certo ponto de vista de abordagem clínica, os chamados erros de separação e junção de palavras podem ser interpretados como indicativos de um transtorno da escrita chamado de *disortografia*. Esse transtorno, como afirmam Fernández et al (2010), está diretamente relacionado aos quadros de transtorno específico da aprendizagem, como, por exemplo, o distúrbio de aprendizagem. Ainda segundo esses autores, os sintomas típicos desse transtorno são “os erros de segmentação da cadeia de fala que ocasionam segmentação indevida de palavras e união incorreta de palavras [...]” (FERNANDÉZ et al, 2010, p. 4).

Essa perspectiva, similarmente à concepção adotada por Zorzi (2006), se apoia em uma suposta interferência da fala na escrita semelhante àquela predominante em ambiente escolar, pois se entende que os casos de segmentação não-convencionais de palavras escritas ocorrem pelo fato de o aluno se basear na oralidade para segmentar as porções do enunciado escrito. Fernández et alii (2010, p. 2) argumentam que essa suposta interferência inicial é esperada e

deve se extinguir quando os escreventes “compreenderem o que é o princípio alfabético, pois, dessa forma, eles poderão fazer uso de maneira mais eficaz da informação obtida pela decodificação letra-som”. Entretanto, se o aluno continua produzindo textos com segmentações não-convencionais nos anos finais do EF, pode ser diagnosticado com disortografia.

Em resumo, essa abordagem interpreta casos de segmentações não-convencionais como:

- i. decorrentes de uma suposta interferência da fala na escrita;
- ii. indícios de que o aluno não seguiu trajeto “adequado” de desenvolvimento da linguagem escrita e por isso não consegue usá-la de forma “eficaz”;
- iii. sintomas de distúrbios de aprendizagem.

Entendemos que compreender casos de segmentação não-convencional de palavras como problemas de alfabetização ou como indícios de patologia é desconsiderar a relação heterogênea que todo nós temos com a língua. Assim, nesta pesquisa, fazemos nossa afirmação de Abaurre, Fiad e Mayrink-Sabinson (1997):

“erros” cometidos pelos aprendizes de escrita/leitura são, de fato, “[...] preciosos indícios de um processo em curso de aquisição da representação escrita da linguagem, registros dos momentos em que a criança torna evidente a manipulação que faz da própria linguagem, história da relação que com ela (re)constrói ao começar a escrever/ler” (ABAURRE, FIAD, MAYRINK-SABINSON, 1997 p. 16-17).

Desse modo, assumimos uma abordagem linguística que vê no chamado erro ortográfico um dado epilinguístico, de trabalho do sujeito da linguagem, seja em seu modo de enunciação falado, seja em seu modo de enunciação escrito.

No trabalho do sujeito com (sua) escrita, a organização prosódica dos enunciados falados seria uma de suas ancoragens, ao mesmo tempo em que estaria ancorado no que supõe ser as convenções da escrita. Ao escrever, o sujeito buscaria plasmar em sua escrita o que ele imagina ser o registro convencional dos enunciados, baseado em suas práticas orais/faladas e letradas/escritas (TENANI, 2013). Essa perspectiva é adotada por autores como Capristano (2007 a, b), Chacon (2013) e Tenani (2016, 2004) para análise de dados de segmentações não-convencional tanto do EF I quanto do EF II.

Nesses trabalhos, a análise de características fonológicas presentes nos dados de segmentação não-convencional é feita com base no modelo de Fonologia Prosódica de Nespor

e Vogel (1986, 2007), segundo o qual os enunciados das línguas do mundo organizados hierarquicamente em sete constituintes prosódicos<sup>6</sup>.

Capristano (2007a) propõe a análise de produções textuais de crianças da primeira série EF I vislumbrando observar marcas na segmentação não-convencional de palavras que permitissem identificar pistas de um imaginário infantil sobre a escrita. Baseada nos resultados de sua pesquisa<sup>7</sup>, argumenta que dois fatores pareceram atravessar fortemente as estruturas das segmentações não-convencionais: (i) fator ligado aos aspectos prosódicos dos enunciados; (ii) fator ligado àquilo que o escrevente imagina ser o código escrito institucionalizado (CORRÊA, 2004, 1997).

Em trabalho posterior, Capristano (2007b) investiga em que medida a escrita de duas crianças do EF I se modifica em direção ao que preveem as convenções ortográficas que determinam como devem ser feitas as delimitações de palavras. De modo específico, a autora propõe uma interpretação para a trajetória das crianças em direção à palavra escrita convencional. A partir dos resultados desse trabalho, a autora argumenta que as segmentações de palavras escritas se modificam ao longo do tempo em função da mudança que se opera na relação sujeito/linguagem. Essa relação, de acordo com Capristano (2007b), envolve três componentes essenciais: o *Outro* como instância de representação do funcionamento da linguagem em seu modo de enunciação escrito; a *escrita*, constituída de modo heterogêneo; e o *sujeito*, concebido sócio-historicamente, permeado por ideologias e responsável por movimentar diversas posições-sujeito no seu enunciado.

Dessa maneira, a autora mostra que, em um mesmo texto, é possível contemplar movimentos de subjetivação do sujeito submetidos a diversos critérios: (i) ora submetidos à escrita do outro – momentos em que há contraste entre uma escrita não passível de interpretação e uma escrita em pleno acordo com as convenções ortográficas; (ii) ora submetida ao funcionamento da linguagem em seu modo de enunciação escrito – momentos em que contrastam “acertos” e “erros”. Nesses momentos, como afirma a pesquisadora, é possível observar que os parâmetros que atuam para a constituição da escrita da criança são múltiplos, já que encontram-se segmentações convencionais e não-convencionais que “parecem ter sido determinadas pela atuação de diferentes limites prosódicos e diferentes informações ligadas à inserção das crianças em práticas letradas” (CAPRISTANO, 2007b, p. 164); (iii) ora submetida

---

<sup>6</sup> A Fonologia prosódica de Nespor e Vogel (1986, 2007), bem como o funcionamento dos setes constituintes da hierarquia prosódica proposta por essas autoras encontram-se na seção 1.4 desse capítulo.

<sup>7</sup> Os resultados de pesquisa de Capristano (2007a) estão resumidos na seção 1.4 deste trabalho.

à observação do próprio sujeito escrevente; momentos em que se observa marcas de correção, rasuras.

Chacon (2013) analisa segmentações não-convencionais de palavras extraídas de textos infantis e aponta, em perspectiva como a adotada por De Lemos (2002), que a flutuação de segmentação de palavras escritas pode dar pistas de mudança de posição da criança no decorrer de sua constituição como sujeito escrevente. Chacon (2013) argumenta, que em ocorrências de flutuação de segmentação como “e la” / “ela” (“ela”); “foipa sua” / “foipasia” (“foi passear”); “alende mão tanha” / “a len das mãotanha” (“além das montanhas”), é possível observar fragmentos de enunciados escritos submetidos ao movimento da língua. Para Chacon (2013), esse movimento é caracterizado pelos desajustes em relação às convenções ortográficas que dão pistas do não-reconhecimento, por parte da criança, da diferença entre a sua escrita e a escrita do outro e do que caracteriza a sua escrita em relação à escrita do outro. O pesquisador considera que umas das evidências desse não-reconhecimento é o fato de haver, em um mesmo enunciado, uma segmentação convencional (“ela”) e uma não-convencional (“e la”). Entretanto, de acordo com o autor, a presença de uma ocorrência segmentada de forma convencional não quer dizer necessariamente que a criança tenha conhecimento das convenções ortográficas. Outra evidência desse não-reconhecimento, segundo o pesquisador, é a ausência de rasuras na escrita desses dados, visto que essa rasura seria um indício do distanciamento da criança em relação a sua escrita.

Chacon (2013) explica que, diferente do que acontece em dados como “inpe” / “en (rasura) pé” (“em pé”), “ouvil” / “ou (rasura) vil” (“ouviu”), “ele fante” / “e (rasura) elefante” (“elefante”), é possível detectar marcas de rasura em pelo menos um dos elementos dos pares. A presença dessas marcas de correção indicia uma possível busca de ancoragem da criança em “diferentes instâncias de sua constituição como sujeito escrevente, instâncias que decorrem de uma escrita/fala e de uma leitura/audição que são recuperadas no próprio ato de escrita e deixam pistas em seu produto final” (CHACON, 2013, p. 381).

Compartilhando da mesma concepção teórica sobre escrita de Capristano (2007a, b) e Chacon (2013), Tenani (2004) propõe análise do funcionamento de segmentações não-convencionais de palavras a partir da relação entre textos escritos por crianças e textos escritos por adultos, ambos os grupos em início de aquisição da escrita. O objetivo da autora nesse trabalho é mostrar possibilidades de “sentidos associados às diferentes segmentações da cadeia fônica” (TENANI, 2004, p. 236).

Tenani (2004) argumenta que é dessas possibilidades de sentidos que muitos textos chistosos se valem para produzir humor, pois, para entender a piada, é necessário perceber

possíveis “similaridades na cadeia fônica que estão ‘estrategicamente’ articuladas com as dessemelhanças que se tornam visíveis por meio da consideração das possibilidades de localização das fronteiras de constituintes prosódicos e dos elementos proeminentes do enunciado” (TENANI, 2004, p. 237). A autora ainda pontua que é por essa razão que, muitas vezes, é possível encontrar em um mesmo texto diferentes tipos de segmentação para uma mesma palavra. Para finalizar, Tenani (2004) afirma que dados não-convencionais de escrita são relevantes por permitirem que se mostrem “‘as fendas e rachaduras’ das estruturas linguísticas” (TENANI, 2004, p. 242).

Em trabalho mais recente, Tenani (2016) analisa o funcionamento de segmentações não-convencionais de palavras escritas identificadas em uma amostra longitudinal de textos do EF II. Nesse trabalho, a autora propõe que os dados de segmentação não-convencional sejam definidos como palavras escritas cujas fronteiras gráficas não seguem as convenções ortográficas do português. A pesquisadora expõe que, nessa definição, estão subjacentes dois posicionamentos teóricos importantes para a análise de dados de segmentação de palavras. O primeiro posicionamento refere-se à noção de fronteira. Segundo a autora, essa noção é mais abrangente, uma vez que a presença ou ausência não-convencional do branco são tipos de representações de fronteiras de palavra na escrita. O segundo posicionamento teórico diz respeito à noção de fronteira a partir de um ponto de vista linguístico, pois, segundo Tenani (2016), tratar de fronteira de palavra implica problematizar o fato de não haver, necessariamente, “sobreposições entre fronteiras de palavra prosódica e morfológica, por exemplo, o que põe em evidência a complexidade da noção de palavra, unidade linguística que se constitui de características fonológicas, morfológicas, sintáticas e semânticas” (TENANI, 2016, p. 174 - 175).

Nesse trabalho, além da análise linguística, Tenani (2016) propõe uma análise de cunho linguístico-discursivo de dados de hipersegmentação. Nessa análise, a autora considera a organização prosódica desses dados e as produções textuais em que eles ocorrem. As reflexões desenvolvidas a partir dessa análise permitiram sugerir que a relação entre prosódia e escrita se atualiza em cada produção textual e que esta é constituída de “tensões entre possibilidades de representação do verbal e do verbo-visual da língua(gem)” (TENANI, 2016, p. 15).

Em síntese, para esses autores:

- i. a inserção da criança/adolescente em práticas orais/faladas e letradas/escritas aponta para o caráter heterogêneo da escrita (CORRÊA, 1997, 2004, 2006);
- ii. os pressupostos de teorias fonológicas (como o modelo de Nespor e Vogel (1986, 2007)) para descrever e analisar aspectos linguísticos observados em textos escritos

contribuem para um entendimento dos modos pelos quais a organização rítmico-entoacional dos enunciados constituem os textos escritos;

- iii. a dimensão textual-discursiva para a análise de dados de segmentação de palavra é importante, uma vez que esse tipo de perspectiva de análise aponta para a existência de um sujeito escrevente que se revela na constituição de diferentes discursos que se materializam no texto.

Do mesmo modo que Tenani (2016), Chacon (2013), Capristano (2007a, b), adotamos uma concepção de escrita constituída de modo heterogêneo. A assunção dessa concepção permite-nos analisar a produção escrita não como um produto neutro, mas como “[...] um modo de enunciação marcado pela diversidade linguístico-histórica dos escreventes” (CORRÊA, 2001, p.164). E considerar a diversidade linguístico-histórica dos escreventes permite-nos compreender que a escrita vai além dos aspectos gráficos e sonoros; é o lugar onde se produz sujeitos e sentidos (CORRÊA, 2013).

Na seção seguinte, trazemos aspectos dos trabalhos de Corrêa (2006, 2004, 1997) com o objetivo de explicar a relevância em se adotar uma concepção de escrita constituída de práticas orais/faladas e letradas/escrita para a análise que propomos fazer.

## 1.2 Perspectiva teórica sobre a escrita

A relação entre fala e escrita mobiliza vários tipos de reflexões em trabalhos que tratam da linguagem. Segundo Abaurre (1991), essas reflexões, muitas vezes, se dividem em duas posições “extremas”: (i) a tese da dependência radical da escrita à fala; (ii) a tese da autonomia da escrita com relação à fala. Neste trabalho, nos distanciamos desses dois tipos de visões e buscamos defender, juntamente com Corrêa (2004, 1997), que há entre fala e escrita uma relação de constituição e não de dependência ou dicotomia.

Em Corrêa (2004, 1997), há a proposta da noção de *heterogeneidade constitutiva da escrita*. Na concepção desse autor, a presença de marcas do oral no escrito não se dá por um processo de interferência entre as modalidades linguísticas da fala na escrita, pois a escrita não é homogênea, é, na verdade, heterogeneamente constituída pelos modos de enunciação falado e escrito.

Nessa proposta, fala e escrita são concebidas como modos de enunciação intrinsecamente ligados às práticas sociais orais/faladas e letradas/escritas, isto é, os fatos da língua (falado/escrito) só se realizam por serem modos de enunciação concretizados por meio de práticas de linguagem. Nesse sentido, o autor defende que a heterogeneidade é própria **da** escrita e não está presente **na** escrita, pois se trata da heterogeneidade como constitutiva da escrita e não como uma característica pontual e acessória desta.

O autor também defende o princípio bakhtiniano de dialogia de que toda produção de linguagem é sempre resultado de um já dito/ouvido e escrito/lido: “o encontro entre as práticas sociais do oral/falado e do letrado/escrito, considerada a dialogia com o já falado/escrito e ouvido/lido” (CORRÊA, 2004, p. 9).

Corrêa (2004) sinaliza que a heterogeneidade da escrita se mostra por intermédio da relação do sujeito com a linguagem, uma vez que é por meio dessa relação que se faz possível recuperar a presença das práticas sociais nas produções discursivas. Para ele, quando o sujeito escrevente enuncia, circula por práticas sociais orais/faladas e letradas/escritas e deixa indícios na materialidade linguística desse processo de trânsito por práticas sociais, revelando como se dá a elaboração de (sua) escrita.

Ademais, em seus trabalhos, Corrêa (2004, 1997) defende que dissertações produzidas por vestibulandos (Unicamp de 1992) captam a circulação que o escrevente faz pelo imaginário sobre a constituição da escrita, evidenciando, por meio de marcas linguísticas, sua enunciação dividida tanto no que se refere ao modo de emergência da escrita como no que se refere ao modo de constituição de seu interlocutor e de sua própria constituição como escrevente. Assim, o sujeito escrevente circula por três eixos de constituição da escrita:

- i. o eixo da gênese da escrita;
- ii. o eixo do código escrito institucionalizado;
- iii. o eixo da relação com o falado/escrito e ouvido/lido.

Segundo Corrêa (2004, 1997), o primeiro eixo, relativo à escrita em sua suposta gênese, refere-se aos momentos em que, ao apropriar-se da escrita, o escrevente tende a tomá-la como representação termo a termo da oralidade, situação em que tende a igualar esses dois modos de realização da linguagem verbal. Em outras palavras, com o que imagina ser a representação termo a termo da fala pela escrita: “tentativa, por parte do escrevente, de representação do planejamento conversacional e do jogo argumentativo prosodicamente marcado ou, finalmente,

sua tentativa de registrar graficamente a materialidade fônico-pragmática do oral/falado” (CORRÊA, 1997, p. 185).

O eixo do imaginário do código escrito institucionalizado refere-se à representação que o sujeito escrevente faz do escrito institucionalizado para a sua escrita. Nas palavras de Corrêa (1997, p. 271), “o escrevente lida, basicamente, com o que supõe ser – a partir não só do que aprendeu na escola, mas, em grande parte, do que assimilou fora dela – a visão escolarizada de código institucionalmente reconhecido”. Nessa visão escolarizada, a escrita é considerada “autônoma, de tal forma que as marcas do oral/falado que eventualmente nela se pudessem reconhecer seriam vistas como desvios do instituído e, nesse sentido, deveriam ser tomadas como lhe sendo totalmente exteriores” (CORRÊA, 1997, p. 170 - 171)

Desse modo, nesse eixo é possível observar marcas linguísticas que indiciam a tentativa do escrevente de “adequar” a (sua) escrita ao que ele imagina ser a escrita escolar. Essa adequação, como mostra o autor, é estabelecida por meio dos entrelaçamentos do oral/falado e do letrado/escrito que, neste eixo, “se evidenciam sempre que o escrevente leva a extremos uma tal imagem sobre o código escrito. Ou seja, nesse eixo da circulação dialógica, esses encontros se mostram pelo excesso” (CORRÊA, 1997, p. 271- 272).

O terceiro eixo de constituição da escrita diz respeito à relação que o texto do escrevente mantém com o já falado e com o já ouvido, bem como com o já escrito e com o já lido. Em outras palavras, esse eixo refere-se aquilo que o escrevente imagina ser a relação apropriada com a exterioridade (outros textos, a própria língua, outros registros, outros enunciadores, o leitor) que constitui o seu texto. Esse eixo se caracteriza na “ideia de que a escrita é um tipo particular de enunciação em que relações com o mundo e com o falado se dão no próprio processo de escrever, caracterizando não a representação escrita do mundo e do falado, mas a representação que o escrevente faz da relação escrita/mundo e escrita/falado” (CORRÊA, 1997, p.341). Corrêa (1997) ainda argumenta que é possível observar essas marcas a partir da “generalização, colagem, réplicas em tom crítico, réplicas simplesmente responsivas, aproveitamento sem citação ou paráfrase, bem como pressuposição de que o leitor conhece os textos da coletânea” (CORRÊA, 1997, p. 343- 344).

Nesse ponto, é importante esclarecer que, no âmbito dessa dissertação, não é nosso intuito sinalizar o funcionamento desses eixos em cada ocorrência de hipossegmentação, uma vez que, ao realizar essa tarefa, fugiríamos dos objetivos aqui estabelecidos.

Outra importante contribuição dada por Corrêa (2006), e da qual nos valem, é a definição dos três modos de reconhecimento da heterogeneidade da escrita. Segundo o autor, a heterogeneidade da escrita pode ser reconhecida:

- i. por meio de aspectos da representação gráfica, já que, no processo da escrita, o escrevente oscila entre a tentativa de representação de características fonético-fonológicas (segmentais e/ou prosódicas) detectadas em sua variedade linguística falada e a convenção ortográfica institucionalizada, esta última baseada na variedade submetida à normalização;
- ii. pela referência à heterogeneidade da língua, pois, se a admitimos como estruturalmente marcada no sistema – contribuição trazida pela sociolinguística – não há razão para não a admitir na escrita, com a condição de que esta última seja vista como um modo de enunciação e não apenas como uma tecnologia;
- iii. pela circulação dialógica que o escrevente faz ao produzir o texto escrito, uma vez que o escrevente circula pelos três eixos de constituição da escrita, aqui já mencionados.

Para concluir esta seção, assumimos, ancorados nos pressupostos apresentados, uma abordagem que concebe a escrita pelo encontro entre os modos de enunciação falado e escrito e as práticas sociais orais e letradas para analisar as produções textuais que investigamos e para a análise das hipossegmentações extraídas desses textos.

Dessa maneira, interpretamos que os usos não-convencionais das fronteiras de palavras são registros da heterogeneidade da escrita. Em outras palavras, assumimos que as hipossegmentações são fruto da imagem que os escreventes têm do que seja a constituição da (sua) escrita, a partir de suas inserções em práticas orais/faladas e letradas/escritas de uso da linguagem.

Na próxima seção, apresentamos o modelo da fonologia prosódica que utilizamos para realizar a análise fonológica dos dados de hipossegmentação.

### **1.3 Prosódia e hipossegmentação**

Neste trabalho, adotamos o modelo chamado de *relation-based*, proposto por Nespor e Vogel (1986), o qual defende a existência de sete constituintes prosódicos que estruturam os enunciados das línguas do mundo. Do maior para o menor, tem-se, nesse modelo, sete domínios, a saber:

Enunciado fonológico (U)  
 Frase entoacional (I)  
 Frase fonológica ( $\phi$ )  
 Grupo clítico (CG)  
 Palavra prosódica ( $\omega$ )  
 Pé-métrico ( $\Sigma$ )  
 Sílabas ( $\sigma$ )

A organização prosódica desses constituintes é norteada por informações morfossintáticas e semânticas. Assim, essa organização é regida por quatro princípios que, segundo Nespor e Vogel (1986), garantem a boa formação dos constituintes prosódicos:

**Princípio 1.** Uma unidade não-terminal dada na hierarquia prosódica,  $XP$ , é composta de uma ou mais unidades da categoria imediatamente mais baixa,  $XP-1$ ;

**Princípio 2.** Uma unidade de um determinado nível da hierarquia está exhaustivamente contida na unidade hierarquicamente superior de que faz parte;

**Princípio 3.** Os constituintes da fonologia prosódica são estruturas n-árias;

**Princípio 4.** A relação de proeminência relativa definida para nós irmãos é tal que a um nó é atribuído o valor forte (s) e a todos os outros nós o valor fraco (w). (Traduzido e adaptado por GAYER (2015) de NESPOR e VOGEL, 1986, p. 7<sup>8</sup>)

De acordo com o primeiro princípio, uma unidade não terminal da hierarquia prosódica ( $X^P$ ) é composta por uma ou mais unidades da categoria imediatamente mais baixa ( $X^{P-1}$ ). Isso quer dizer que o pé métrico, por exemplo, é constituído por uma ou mais sílabas, como ilustrado na Figura 1.

---

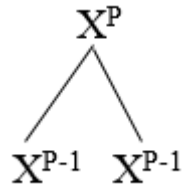
<sup>8</sup> *Principle 1. A given nonterminal unit of the prosodic hierarchy,  $X^P$ , is composed of one or more units of the immediately lower category,  $X^{P-1}$ .*

*Principle 2. A unit of a given level of the hierarchy is exhaustively contained in the superordinate unit of which it is a part.*

*Principle 3. The hierarchical structures of prosodic phonology are n-ary branching.*

*Principle 4. The relative prominence relation defined for sister nodes is such that one node is assigned the value strong (s) and all the other nodes are assigned the value weak (w). (NESPOR e VOGEL, 2007, p. 7).*

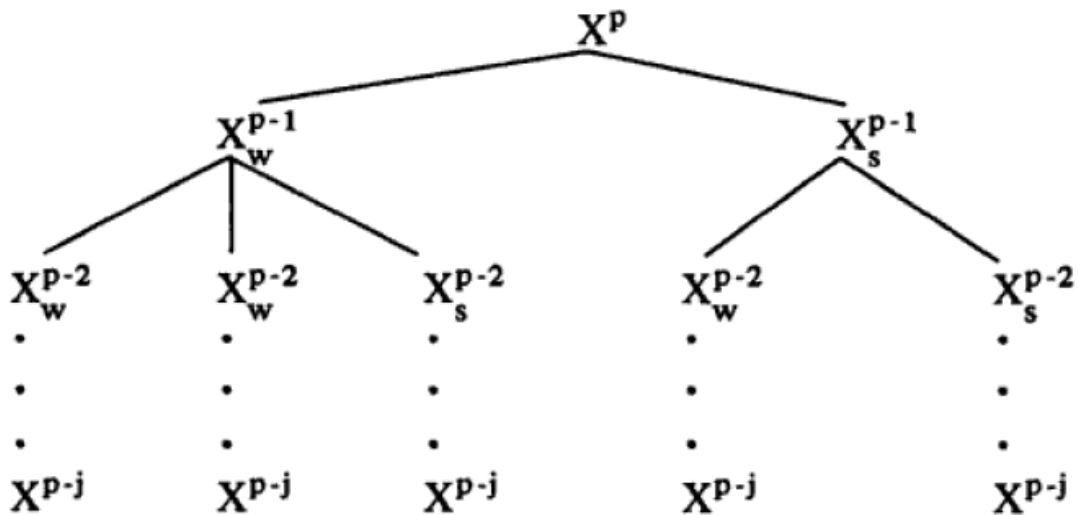
Figura 1. Exemplo do primeiro princípio de constituição da hierarquia prosódica



Fonte: Elaboração própria

O segundo princípio trata da exaustividade, assim, uma unidade de um dado nível da hierarquia está exaustivamente contida na unidade superior de que faz parte. Desse modo, a sílaba, por exemplo, obrigatoriamente deve estar contida no pé métrico, assim como o pé métrico na palavra prosódica e sucessivamente. Essa organização encontra-se ilustrada na Figura 2:

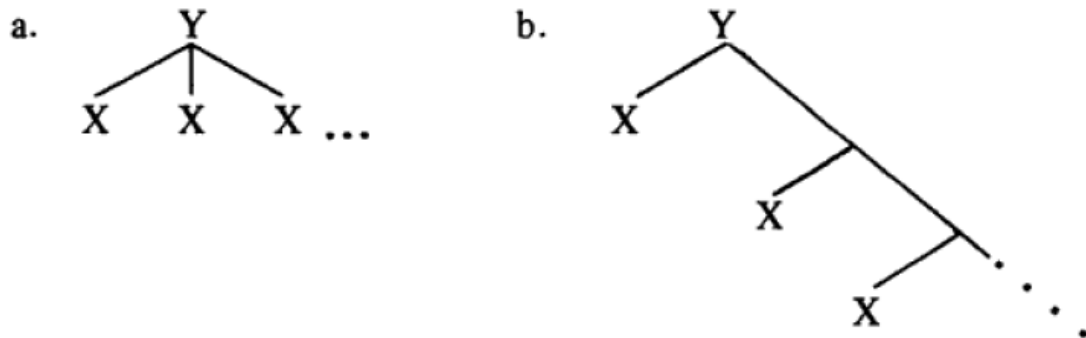
Figura 2. Exemplo do segundo princípio da constituição da hierarquia prosódica



Fonte: Nespor e Vogel (2007, p.8)

O terceiro princípio que rege as estruturas hierárquicas prega que as ramificações devem ser n-árias. Isso significa que todas as ramificações de um dado domínio devem sair de um só nó, diferentemente do que ocorre em estruturas binárias, nas quais um novo nó pode surgir dentro de uma ramificação. Em *a*, na Figura 3, temos um exemplo de uma estrutura n-ária e em *b* um exemplo de uma estrutura binária. No modelo adotado, a configuração em (a) é adotada para construção das estruturas prosódicas.

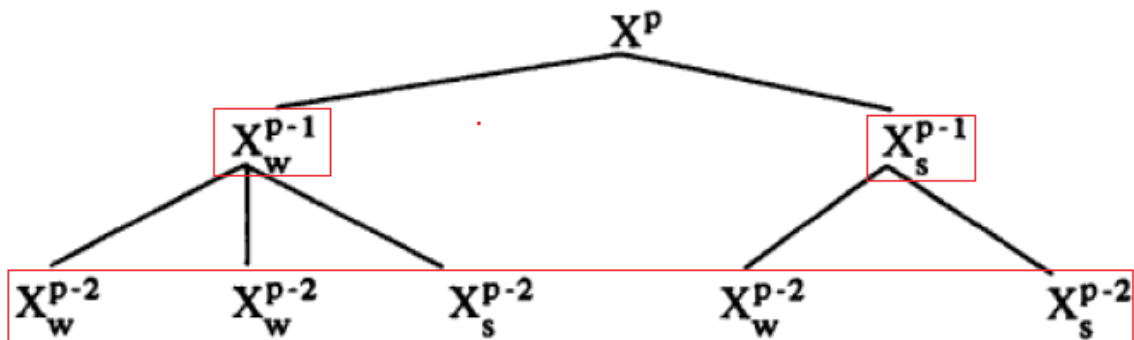
Figura 3. Estruturas n-árias e binárias



Fonte: Nespor e Vogel (2007, p. 9)

No último princípio, afirma-se que entre os nós irmãos há uma relação na qual a um nó é atribuído o valor de forte (s) e a todos os outros o valor de fraco (w)<sup>9</sup>, como na Figura 4 a seguir.

Figura 4. Exemplo do quarto princípio de construção da hierarquia prosódica



Fonte: Nespor e Vogel (2007, p.8, com ajustes nossos)

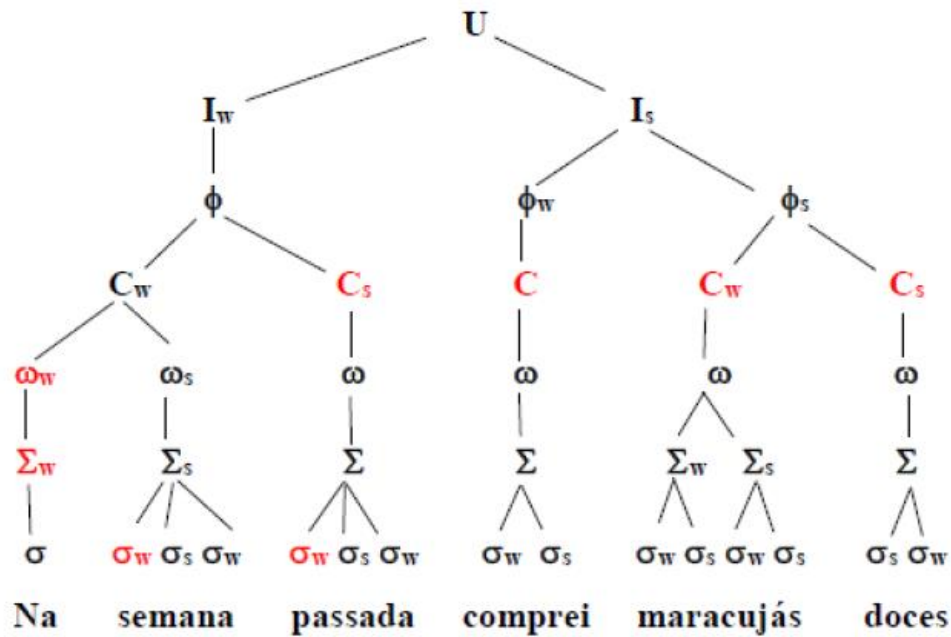
São esses os princípios que garantem que todos os constituintes tenham a mesma configuração geométrica. Além desses princípios, os constituintes seguem uma regra da gramática universal: junte em uma ramificação n-ária  $X^P$  todos os  $X^{P-1}$  incluídos na cadeia delimitada pela definição do domínio de  $X^P$ <sup>10</sup>. É a partir dessa regra e dos princípios apresentados que se organizam hierarquicamente os domínios prosódicos.

A seguir, trazemos um enunciado do PB analisado por Tenani (2016) em constituintes prosódicos, conforme hierarquia de Nespor e Vogel (2007, 1986).

<sup>9</sup> Para a notação de forte e fraco, são utilizadas as letras “s” da palavra em inglês *strong* e “w” da palavra em inglês *weak*.

<sup>10</sup> “*Prosodic Constituent Construction: Join into an n-ary branching  $X^P$  all  $X^{P-1}$  included in a string delimited by the definition of the domain of  $X^P$ .*” (NESPOR; VOGEL, 2007, p. 7).

Figura 5. Hierarquia prosódica segundo Nespor e Vogel (1986, 2007)



Fonte: Tenani (2016, p. 92) a partir da hierarquia de Nespor e Vogel (1986, 2007)

Apresenta-se, a seguir, uma breve caracterização de cada um dos constituintes, levando-se em conta características do PB, com base em Tenani (2016).

A sílaba ( $\sigma$ ) é a unidade basilar da hierarquia prosódica, na qual a cadeia fônica é estruturada. Na Figura 5, elaborada por Tenani (2016) para o PB, temos, por exemplo, a palavra *doce*, que é constituída por duas sílabas que mantêm entre si uma relação de dominância, na qual a sílaba “do” é forte em relação a sílaba “ces”.

Já o pé métrico ( $\Sigma$ ), segundo Bisol (1996), constitui-se da relação de dominância/dominado entre duas ou mais sílabas. A depender dessa relação, o pé métrico recebe uma nomenclatura: pés binários (com duas sílabas), com o acento à esquerda, são chamados de troqueus, como na palavra “doces” contemplada na Figura 5; pés binários, com acento à direita, são denominados iambos, como na palavra “comprei” também presente na Figura 5; e pés ternários, com três sílabas, caso das palavras proparoxítonas, são denominados dátilos, como, por exemplo, a palavra “árvore”.

Logo acima do pé, tem-se a palavra prosódica ( $\omega$ ), que pode ser composta por um ou mais  $\Sigma$ s. A palavra prosódica caracteriza-se por ser portadora de um acento lexical ou acento primário. Um exemplo em PB é a palavra morfológica “guarda-roupa”, que se constitui de duas palavras prosódicas [guarda] $\omega$  [roupa] $\omega$ , por haver dois acentos: um na primeira sílaba de “guarda” e outro na primeira sílaba de “roupa”.

Na hierarquia prosódica de Nespor e Vogel (1996), a palavra prosódica é dominada pelo grupo clítico (CG)<sup>11</sup>, que é formado por uma única palavra de conteúdo acompanhada de clíticos (palavras gramaticais átonas, tais como artigos, preposições, conjunções). Em um enunciado, como o dado por Tenani (2016), tem-se um grupo clítico em “na semana”, no qual o clítico “na” vem acompanhando seu hospedeiro “semana”, que é uma palavra prosódica por apresentar um acento primário.

Imediatamente superior ao grupo clítico, configura-se a frase fonológica ( $\phi$ ). Esse constituinte engloba um ou mais grupos clíticos. Na Figura 5, segundo Tenani (2016), o substantivo “maracujás” e o adjetivo “doces” formam duas frases fonológicas autônomas. Entretanto, há reestruturação, tornando-se uma única frase fonológica, “maracujás doces”, pois “doces” é o complemento de “maracujás”<sup>12</sup>. No PB, quando o nó é ramificado, o elemento que estiver mais à direita será rotulado como forte. No exemplo dado pela autora, o acento da frase fonológica recairá na tônica de “doce”, por ser a sílaba mais à direita em relação à sílaba tônica de “maracujá”.

O constituinte acima da frase fonológica é a frase entoacional (I), a qual pode ser formada por um conjunto de frases fonológicas, como no exemplo dado por Tenani (2016): [[Na semana passada]  $\phi$ ] I [[comprei]  $\phi$  [maracujás doces]  $\phi$ ] I. A frase entoacional pode ser ainda formada por apenas uma frase fonológica, como em: [Cuidado!] I, desde que esta apresente um contorno entoacional ascendente ou descendente.

Por fim, o enunciado fonológico (U), em Nespor e Vogel (1986, 2007), é o constituinte mais alto da hierarquia prosódica. Tem sua proeminência relativa sempre mais à direita em línguas de recursividade sintática à direita, como o PB, e sua identificação é feita através dos limites sintáticos e da pausa. No entanto, nem sempre o enunciado tem o mesmo tamanho do constituinte sintático, no exemplo da Figura 5, temos: [[Na semana passada] I [comprei maracujás doces]I] U. Quando o enunciado fonológico é constituído por duas frases entoacionais, a frase mais à direita da ramificação recebe o valor de forte, como ilustrado na Figura 5.

Essa hierarquia de Nespor e Vogel (1986) foi alvo de muitas críticas, pois as autoras, com o intuito de não ferirem nenhum dos quatro princípios que garantem a boa formação da

<sup>11</sup> Esse constituinte vem fomentando muitas discussões e estudos atualmente. Trataremos dessa problemática mais adiante.

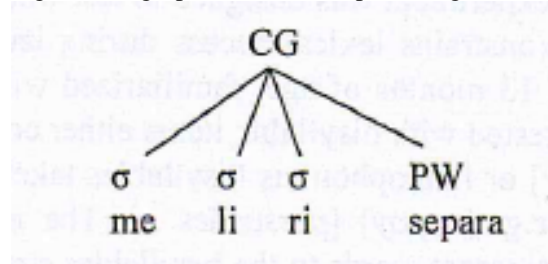
<sup>12</sup> Para essa reestruturação ocorrer, de acordo com Nespor e Vogel (2007), é necessário que o complemento esteja à direita e as frases fonológicas não sejam ramificadas. Tenani (2016) explica que se a estrutura fosse “belos maracujás doces” não haveria reestruturação, pois, “belos maracujás”, forma uma frase fonológica ramificada, ou seja, constituída por duas palavras prosódicas.

hierarquia prosódica, deram ao clítico um *status* prosódico que ele não tem em várias línguas, o de pé métrico e de palavra prosódica.

Na edição de 2007 do livro *Prosodic phonology*, as autoras reconhecem o problema, porém, para elas, eliminar o grupo clítico não seria a melhor solução, já que, se as características dele forem “transmitidas” para a palavra prosódica, não deixaria de existir um problema, uma vez que resultaria em dois tipos de palavras prosódicas e, assim, dois domínios diferentes.

A solução proposta por elas foi “flexibilizar” os princípios 1 e 2. Com a flexibilização, é permitido que um constituinte domine até dois domínios abaixo dele. Dessa maneira, o grupo clítico passa a também dominar sílabas, ao invés de somente a palavra prosódica, como vemos na Figura que segue:

Figura 6. O grupo clítico de Nespor e Vogel (2007)



Fonte: Nespor e Vogel (2007, p. 16)

Na Figura 6, observamos um exemplo em italiano no qual o grupo clítico é constituído de clíticos *me*, *li* e *ri*, e palavra prosódica *separa*. Nesse exemplo, as sílabas dos clíticos não são pés métricos, nem palavras prosódicas, sendo incorporadas à hierarquia prosódica quando da constituição do grupo clítico.

O argumento dado pelas autoras que dá suporte a existência do grupo clítico na hierarquia prosódica é o de que este constituinte deve ser domínio de aplicação de regras segmentais, como em italiano (NESPOR; VOGEL 1986, 2007). Para o PB, Bisol (2005, 2000) também assume a existência do grupo clítico, defendendo que o grupo clítico é o menor constituinte pós-lexical, pois, enquanto os prefixos sempre integram a palavra prosódica, completando-a ou aumentando-a, os clíticos anexam-se a uma palavra pronta, sem integrá-la, emergindo o menor constituinte prosódico pós-lexical.

Bisol (2000) observa que, em PB, o processo de integração que existe entre o prefixo e sua base não existe entre o clítico e seu hospedeiro, uma vez que os membros que compõem essa unidade mantêm sua independência. Ela traz evidências de que os clíticos se comportam diferente dos afixos, já que figuram como independentes na aplicação de regras fonológicas,

como na neutralização de /e, o/, em final de palavra, onde /i, u/ respectivamente se manifestam (exemplos 1.1, 1.2 e 1.3):

- 1.1. do menino > d[u] menin[u];
- 1.2. me leve > m[i] lev[i];
- 1.3. se sente; s[i] sent[i];

Nota-se, nos exemplos acima, que a regra se aplica em ambos os membros do grupo, se aplica tanto em palavras gramaticais, como as preposições “do” e pronomes átonos “me”, quanto em palavras lexicais, como *menino*. Essa regra não se aplica aos prefixos ou sílabas pretônicas, integradas a seu hospedeiro (exemplos 1.4 e 1.5):

- 1.4. reanimar \*rianimar;
- 1.5. pospor \*puspor.

Os prefixos são sempre ligados primeiramente a uma base morfológica, para somente depois **integrarem-se** à palavra como um todo. Os clíticos, por sua vez, nunca se integram à palavra, mas a ela se **anexam**, por adjunção, sob o domínio de um constituinte prosodicamente mais alto. Para Bisol (2000) essa informação sugere que entre o clítico e seu cabeça estabeleceu-se uma relação de adjunção e não de integridade.

A autora ainda traz outras evidências a favor do grupo clítico, mas a principal delas é a elisão da vogal /e/, que só ocorre entre clítico e hospedeiro (exemplos 1.6, 1.7, 1.8 e 1.9):

- 1.6. Em outro dia > noutro dia
- 1.7. De outra vez > doutra vez
- 1.8. De um dia para o outro > dum dia para o outro
- 1.9. Em um espaço curto > num espaço curto

Em (1.6) e (1.7), ocorre o processo de elisão entre o clítico (“em”, “de”) e uma palavra gramatical (“outro”, “outra”); nos exemplos (1.8) e (1.9), entre dois clíticos (“de” + “um”; “em” + “um”). O mesmo processo não é visto na relação entre duas palavras lexicais, como exemplificado em (1.10 e 1.11):

1.10. Cidade antiga \*cidadantiga

1.11. Grande amigo \*grandamigo

A partir dessa evidência, a autora conclui que o clítico, juntamente com a palavra com que se relaciona, portam uma natureza frasal, de uma locução, pois envolvem elementos fonologicamente independentes. Assim, é, sem dúvida para a autora, um constituinte pós-lexical que não tem o tamanho da **palavra prosódica** de Selkirk (1984), mas o tamanho do **grupo clítico** de Nespor e Vogel (1986, 2007). Contrária a essa visão é a interpretação proposta por Simioni (2008). Para essa autora, a elisão da vogal média /e/ pode ser analisada como evidência da prosodização do clítico anexado à frase fonológica. Em termos gerais, Simioni (2008) argumenta que essa prosodização se dá em função de que o clítico: (i) não se comporta como uma palavra prosódica independente; (ii) não se comporta como uma sílaba pretônica (no caso dos proclíticos) ou postônica (no caso dos enclíticos); e (iii) parece não se comportar como se estivesse no início de uma palavra prosódica. Com relação a (ii), a autora defende que em PB não parece haver uma distinção entre próclise e ênclise em termos de estrutura prosódica, ao contrário do que acontece em outras línguas.

Em suma, a prosodização de clíticos – seja em PB, PE ou outras línguas – é tema que gera discussão. Não há consenso na literatura sobre a necessidade de haver um domínio específico na hierarquia prosódica para prosodização de clíticos. Também a definição do *status* do grupo clítico tem causado bastante repercussão e discussões quanto a sua pertinência, quanto à sua formalização. Entretanto, faz-se importante explicitar que, nesta pesquisa de mestrado, reconhecemos essa problemática sem adentrar na discussão de possíveis soluções e nos limitamos a reconhecer que existe uma estrutura prosódica para o clítico e seu hospedeiro. Assumimos que a prosodização do clítico se dá no chamado grupo clítico, proposto por Nespor e Vogel (1986, 2007) e defendido por Bisol (2000, 2005) para o PB.

Na seção seguinte, fazemos revisão de trabalhos que tratam da análise de dados de segmentação não-convencional de palavras.

#### **1.4 Hipossegmentações no EF I e no EF II**

Nosso objetivo, nesta seção, é fazer uma revisão de trabalhos que analisam as hipossegmentações visando situar esta dissertação em relação ao que já foi investigado, uma

vez que esse também é nosso objeto de análise. Para tanto, iniciamos apresentando os trabalhos de Cunha (2004, 2010), de Cunha e Miranda (2007), de Capristano (2007a) e de Torquette (2016) a respeito da escrita de crianças do EF I. Em seguida, apresentamos os trabalhos de Tenani (2011) e de Paranhos (2014), que se referem à escrita de alunos do EF II.

Cunha (2004) e Cunha e Miranda (2007) analisaram dados de hipo e de hipersegmentação em textos produzidos por dez crianças, cinco de uma escola pública e cinco de uma escola particular, ao longo dos quatro anos do EF I. O objetivo das autoras foi descrever e analisar fatos de aquisição da escrita, especificamente, aqueles relacionados à segmentação de palavras, com o intuito de verificar o papel, nessa segmentação, dos constituintes prosódicos, conforme definidos por Nespor e Vogel (1986)<sup>13</sup>.

Na análise que fazem, as autoras optam por dividir os dados de segmentação não-convencional a partir de uma variável linguística que considera o tipo de palavra (se fonológica ou gramatical), a tonicidade e o tipo de sílaba. Com relação à variável tipo de palavras, Cunha (2004) e Cunha e Miranda (2007) mostram duas tendências predominantes nos dados de hipossegmentação: juntura entre uma palavra gramatical<sup>14</sup> e outra fonológica, como em: “olobo” (“o lobo”), “tecomer” (“te comer”), e juntura entre duas palavras fonológicas, como em “chicobento” (“Chico Bento”), “pedefeijão” (“pé de feijão”).

As autoras mostram também que os dados que envolvem uma palavra gramatical mais uma palavra fonológica aparecem em maior quantidade, o que, segundo elas, corrobora a afirmação de Ferreiro e Teberosky (1999) de que a criança que começa a escrever tem dificuldades em reconhecer como palavra conjuntos de uma ou duas letras, como conjunção (“e”, “se”), artigos (“o”, “a”), pronomes (“me”, “te”) e preposições (“de”, “em”) monossilábicas. Por isso, na maioria dos casos, acaba por juntar essas “letras” à palavra seguinte, fazendo uma hipossegmentação.

Os casos de hipossegmentações entre duas palavras fonológicas ocorrem em menor quantidade e formam frases fonológicas ou frases entoacionais. Sobre esse tipo de dado, as autoras explicam que frases fonológicas, como “miaroupa” (“minha roupa”), podem ocorrer devido à mobilização de critérios morfossintáticos e semânticos, como a relação entre o adjetivo “minha” e o substantivo “nome” que constituem um sintagma e, podem também, formar uma

<sup>13</sup> Cunha (2004, 2010), Cunha e Miranda (2007) fundamentam-se na Teoria da Linguagem de Noam Chomsky, de caráter inatista, e a na Teoria da Aprendizagem de Jean Piaget, de caráter construtivista. Assim, nesses trabalhos, a escrita é concebida como um objeto cultural a ser adquirido pelo sujeito (o escrevente) e o sujeito é entendido como ativo, idealizado e universal.

<sup>14</sup>A palavra gramatical em Cunha (2004) é entendida como aquela que não possui significado lexical, como pronomes e preposições, por exemplo. Já a palavra fonológica abarca todas as palavras que possuem um acento primário e que, mesmo não tendo significado conhecido na língua, são candidatas potenciais para tal.

frase fonológica. No caso das frases entoacionais, pode haver uma motivação devido à presença de um contorno entoacional e de pausa que delimitam porções do enunciado, como em “Chico Bento saiu correndo”, grafado como “chicobento saiu correndo”. Observa-se que “chicobento” é um nome que pode funcionar como tópico com contorno entoacional próprio, e “saiu correndo”, uma sentença que pode funcionar como comentário com seu entoacional característico.

No que diz respeito à variável “tonicidade”, Cunha e Miranda (2007) e Cunha (2004) argumentam que a presença de grupos tonais ou de contornos entoacionais pode motivar casos de hipossegmentação. Ademais, a sílaba tônica da palavra também parece ser, de algum modo, levada em conta, pois em quase todos os casos sempre é preservado o pé do acento.

Assim, Cunha e Miranda (2007) e Cunha (2004) concluem que os dados analisados revelaram importantes tendências: (i) há uma possível manutenção da integridade da sílaba e do pé métrico, pois a criança preserva o constituinte silábico ao unir palavras; (ii) os três constituintes de nível mais alto da hierarquia prosódica parecem dirigir processos de hipossegmentação; (iii) a criança, ao perceber a fala como um contínuo, apresenta forte tendência em separar a escrita de acordo com unidades entoacionais.

Em trabalho posterior, Cunha (2010) propõe, a partir de dados segmentações não-convencionais produzidas por crianças brasileiras e portuguesas, uma análise que mostra a relevância dos dados de escrita inicial para a discussão sobre o ritmo linguístico do português brasileiro (PB) e do português europeu (PE). Os dados são advindos de textos escritos por crianças que cursavam, à época das coletas, as séries iniciais do EF em Pelotas (Brasil) e Ensino Básico em Porto (Portugal). Para realizar esse estudo, a autora se baseou em trabalhos sobre o ritmo em fonologia, tais como o de Abaurre e Galves (1998), no qual é defendido que o ritmo de cada língua resultaria da hierarquização de três princípios: “integridade da palavra fonológica”, “pé trocaico” e “binariedade do pé”.

Cunha (2010) mostra que os dados de hipossegmentação ocorrem em maior quantidade se comparado aos casos de hipersegmentação e híbridos. As hipossegmentações envolveram a união de uma palavra fonológica e de uma palavra gramatical, como, por exemplo, “escondeuce” (“escondeu-se”). A autora pontua que esses dados evidenciam importante diferença rítmica entre o PB e PE. No PB, a tendência é a união entre uma palavra gramatical e uma palavra fonológica, enquanto que, no PE, a tendência é a junção entre palavra fonológica e palavra gramatical, evidenciando uma diferença de direcionalidade de prosodização entre as duas variedades. Por exemplo, na sequência “para te ver melhor”, temos, no PB, hipossegmentações como “para tever melhor” e no PE “parate ver melhor”.

Segundo Cunha (2010), essas duas configurações de hipossegmentações são resultados de ritmos distintos das variedades do português, que motivaria a diferença de segmentação: no PE, o ritmo caracteriza-se pela hierarquia: i) pé trocaico, ii) integridade da palavra fonológica e iii) binariedade do pé; já o PB caracteriza-se pela hierarquia: i) integridade da palavra fonológica, ii) binariedade do pé, e iii) pé trocaico. Em ambas as variedades, segundo a autora, a integridade da palavra fonológica precede a posição da binariedade do pé. Desse modo, no PB, há a manutenção da integridade da palavra fonológica em relação à formação de um pé troqueu e preferência em unir o pronome ao verbo pela próclise. No PE, por sua vez, o monossílabo “te”, em função da sua atonicidade, junta-se à palavra “para” formando com ela um pé trocaico.

Apoiada nas considerações de Abaurre e Galves (1998), Cunha (2010) conclui que a diferença entre o ritmo nas duas variedades do português está fundamentada na posição que o pé trocaico ocupa no interior da hierarquia. No PE, ele ocupa o primeiro lugar e, no PB, o último lugar. Assim, o ritmo em PE é baseado no troqueu, enquanto o ritmo no PB se constrói a partir de fronteiras de palavras fonológicas. Esse trabalho demonstra que hipossegmentações são, ao menos em parte, motivadas pela organização prosódica dos enunciados falados, uma vez que a distinção na prosodização de clíticos entre as variedades do português pode também ser observada nos registros escritos por crianças no início do chamado processo de alfabetização.

Ainda sobre a escrita infantil, retomamos o trabalho de Capristano (2007a), no qual é proposta a análise de produções textuais de crianças da primeira série EF I vislumbrando observar marcas na segmentação não-convencional de palavras que permitissem identificar pistas de um imaginário infantil sobre a escrita. A partir da observação dessas marcas, Capristano (2007a) identificou quatro “tipos de funcionamentos”: (i) segmentações não-convencionais resultantes de tentativas de escrita alfabética; (ii) segmentações não-convencionais resultantes da oscilação entre diferentes trânsitos por constituintes prosódicos e informações sobre o código escrito institucionalizado; (iii) segmentações não-convencionais resultantes da oscilação entre constituintes abaixo do domínio da palavra fonológica da hierarquia prosódica (sílabas e pé) e informações sobre o código escrito institucionalizado; e (iv) segmentações não convencionais resultantes de maior percepção de constituintes acima da palavra fonológica na hierarquia prosódica e, talvez, em menor grau, de informações sobre o código escrito institucionalizado (CAPRISTANO, 2007a, p. 71).

De modo específico, sobre o primeiro funcionamento, Capristano (2007a) explica que as segmentações desse tipo se referem a momentos nos quais o escrevente parece circular por

fatos de natureza fonético-fonológicos e pelo código escrito institucionalizado. Nesse conjunto de dados, a autora identificou dois funcionamentos: (1A) segmentações não-convencionais que parecem ser resultado do trânsito entre elementos de natureza fonético-fonológico e de informações sobre o código escrito institucionalizado; (1B) segmentações cuja seleção das letras parece basear-se prioritariamente em informações sobre o código escrito institucionalizado.

Para ilustrar o funcionamento em (1A), citamos o dado “BOIABOL” identificado no *cópus* da autora. Sobre essa ocorrência, Capristano (2007a) afirma que, embora ela não tenha atribuição de sentido, é possível reconhecer a palavra “boi” e uma parte da palavra “bola”. A autora explica que o escrevente, nesse exemplo, parece ter optado por preencher o espaço em branco com palavras com as quais estivesse mais seguro no que se refere ao seu conhecimento de acordo com as convenções ortográficas. O funcionamento de (1B) pode ser exemplificado com o dado “ROORT76”. Nesse exemplo, Capristano (2007a) argumenta que o escrevente parece ter preenchido o espaço gráfico com elementos que poderiam ser escritos (letras e números) e que se referem a um certo imaginário sobre o código escrito institucionalizado que seria, de certo modo, uma resposta ao que ele supõe ser, com base naquilo que aprendeu dentro e fora da escola, o esperado pela instituição escolar (CORRÊA, 1997 *apud* CAPRISTANO, 2007).

As segmentações não-convencionais que evidenciam o segundo funcionamento dizem respeito à oscilação entre diferentes trânsitos por constituintes prosódicos e informações sobre o código escrito institucionalizado. De acordo com Capristano (2007a), um exemplo desse tipo de funcionamento encontra-se na sequência “nao estavanem ai” (“não estava nem aí”). Nessa ocorrência, a autora argumenta que o escrevente parece enfatizar ora uma percepção de fatos prosódicos que possivelmente produz em sua variedade linguística falada, ora um certo imaginário sobre o código escrito institucionalizado.

Ainda no que diz respeito a essa ocorrência, Capristano (2007a) demonstra que o escrevente parece se basear na percepção de uma categoria como a frase entoacional: [não estava nem aí]  $\phi$ . Além dessa explicação, a autora ainda levanta duas hipóteses. A primeira é a de que a delimitação por espaço em branco de “não” e “aí” poderia indicar que o escrevente tenha intuído o funcionamento de “não” e “aí” como unidades gráficas menores, mas que, na escrita, funcionariam como palavras da língua. A segunda hipótese da autora é a de que essas unidades podem ter saliências em decorrência dos diferentes funcionamentos que assumem nas práticas orais nas quais esse escrevente está imerso. No que se refere às hipóteses levantadas,

Capristano (2007a) ressalta que uma hipótese não exclui a outra e que juntas elas corroboram o modo heterogêneo de constituição da escrita (CORRÊA, 1997).

O terceiro tipo de funcionamento abrange as ocorrências em que o escrevente segmentaria mais do que previsto pelas convenções ortográficas. Esse funcionamento desdobra-se em dois outros tipos: (3A) segmentações não-convencionais resultantes de oscilação entre uma percepção de unidades gráficas autônomas e uma percepção de sílabas prosódicas; (3B) segmentações não-convencionais resultantes da oscilação entre uma percepção de unidades gráficas autônomas e uma percepção de pés-métricos.

A ocorrência “é sa muzica” (“essa música”) ilustra o funcionamento em (3A) descrito por Capristano (2007a). Segundo a autora, essa ocorrência parece ter sido atravessada pelo reconhecimento do significado da palavra “é” e pela saliência prosódica das sílabas que constituem essa sequência. O dado “Feliz Bina”, por sua vez, ilustra o funcionamento em (3B) segundo Capristano (2007a). Nesse dado, parece ter havido, por parte do escrevente, a percepção da saliência dos pés-métricos [feliz]  $\Sigma$  [bina]  $\Sigma$ .

Capristano (2007a) mostra que o quarto tipo de funcionamento abarca as segmentações não-convencionais nas quais a localização dos espaços em branco permite pensar que os escreventes tenham, de certo modo, se baseado em uma percepção de fatos prosódicos que produzem em sua variedade linguística falada. De acordo com a autora, esse funcionamento desdobra-se em quatro tipos: (4A) segmentações não-convencionais resultantes de uma maior percepção de um grupo clítico e, em menor grau, de informações sobre o código escrito institucionalizado; (4B) segmentações não-convencionais resultantes de uma maior percepção de uma frase fonológica e, em menor grau, de informações sobre o código escrito institucionalizado; (4C) segmentações não-convencionais resultantes de uma maior percepção de uma frase entoacional e, em menor grau, de informações sobre o código escrito institucionalizado; (4D) segmentações não-convencionais resultantes de uma maior percepção de um enunciado fonológico e, em menor grau, de informações sobre o código escrito institucionalizado.

Para ilustrar o funcionamento de (4A), Capristano (2007a) traz, entre outras, a ocorrência “socorrela” (“socorre-la”). Nessa ocorrência, a autora argumenta que o escrevente parece trabalhar com uma representação baseada na percepção de um grupo clítico e, também, com a imagem que tem da escrita privilegiada pela escola, isto é, parece haver uma tentativa de alçar, por meio da ênclise, uma escrita culta formal (CORRÊA, 1997).

Como exemplo do funcionamento proposto em (4B), Capristano (2007a) traz o dado “fifofe” (“ficou feliz”). Nesse exemplo, a autora explica que o escrevente pretendia escrever

“ficou feliz”, mas interrompeu a escrita da última palavra “feliz”. Segundo Capristano (2007a), esse enunciado pode ser interpretado como uma frase fonológica, pois há a presença do espaço em branco nas fronteiras dessa ocorrência (início e fim) associada à ausência de espaço em branco entre as “palavras” “ficou” e “fe”.

A ocorrência “Estavaveidodehomem” (“estava vendo um monte de homem”) é um dos exemplos caracterizadores do funcionamento proposto em (4C) no qual, segundo a autora, parece haver, por parte do escrevente, uma percepção de uma frase entoacional. Nesse enunciado, Capristano (2007a) argumenta que a presença do espaço em branco nas fronteiras dessa ocorrência (início e fim), associada à ausência de espaço em branco entre as palavras “estava”, “vendo”, “de” e “homem” podem ser interpretadas como indícios de uma representação feita pelo escrevente baseado na percepção do que, em sua variedade linguística, poderia constituir uma frase entoacional.

Por fim, Capristano (2007b) discute os exemplos caracterizadores do funcionamento proposto em (4D) no qual parece haver uma percepção, por parte do escrevente, de um enunciado fonológico: “comariaodeuza xieis” (“qual material que usa? Giz”). Nessa ocorrência, de acordo com a autora, parece ser possível identificar alternância de enunciadores, a qual, de certo modo, indicaria a percepção do escrevente de um certo contorno prosódico de um enunciado fonológico. Desse modo, Capristano (2007a), baseada nos resultados de sua pesquisa, argumenta que dois fatores pareceram atravessar fortemente as estruturas das segmentações não-convencionais: (i) fator ligado aos aspectos prosódicos dos enunciados; (ii) fator ligado àquilo que o escrevente imagina ser o código escrito institucionalizado (CORRÊA, 1997, 2004).

Torquette (2016) também analisa as segmentações de palavras na escrita infantil. Em seu trabalho, a autora compara a distribuição de segmentações não-convencionais de palavras encontradas em textos de alunos do 1º ao 5º ano do EF I. Para tanto, a autora se vale de três variáveis: (i) tipos de escola (pública e privada), (ii) anos letivos e (iii) sexo/gênero (feminino e masculino). Para realizar essa comparação, Torquette (2016) montou um banco de dados (EscIn) composto por 4.230 textos de cinco escolas privadas e de dez escolas públicas do Município de Marília (SP) durante o ano de 2012. Desse material, a autora valeu-se de uma amostra estratificada de 10% dos dados de cada variável de constituição do banco, perfazendo um total de 427 textos.

Em relação à primeira variável, Torquette (2016) mostra que houve diferença estatística significativa em relação ao número de segmentações de palavras das escolas públicas e privadas, pois houve mais ocorrências de segmentações não-convencionais de palavras nos

textos de escreventes de escolas públicas do que de escreventes de escolas privadas. No que diz respeito à hipossegmentações, na escola pública, houve 111 ocorrências, das quais 85 (76,5%) corresponderam à estrutura de um grupo clítico e, na escola privada, sete ocorrências, das quais seis (85,7%) corresponderam à estrutura de um grupo clítico.

Para a variável anos letivos, Torquette (2016) partiu da hipótese de que haveria uma diminuição de casos de segmentação não-convencional ao longo dos cinco anos letivos do EF I. Confirmando essa hipótese, a autora mostra, que em ambas as escolas (pública e privada) houve diferença estatística significativa. Todavia, essa diminuição ao longo do EF I ocorreu de maneira distinta nos tipos de escolas contempladas. A autora explica que, nas escolas públicas, houve uma diminuição contínua, chegando a uma mediana de 0,5% de ocorrências no quinto ano; já nas escolas privadas essa diminuição aconteceu de forma acentuada do primeiro para o segundo ano, chegando a zero no terceiro ano. Entretanto, Torquette (2016) argumenta que os escreventes de escolas públicas iniciaram o primeiro ano do EF I com 12,8% de ocorrências de segmentação não-convencional e os de escolas privadas com 7,3%. Assim, a autora interpreta que a trajetória do desempenho dos escreventes de escolas públicas apresentou ganhos considerados superiores em relação aos de escreventes de escolas privadas, pois, ao final do percurso, os escreventes tanto das escolas públicas, quanto das privadas, mostraram desempenhos próximos.

Torquette (2016) ainda analisou a natureza das segmentações não-convencionais que restaram aos escreventes da escola pública e privada no quinto ano do EF I, e observou que, nas escolas públicas, houve maior recorrência de hipossegmentações, o que parece ser em virtude de dúvidas quanto a questões fonológicas para a determinação das fronteiras de palavra. Por outro lado, nas escolas privadas, essa questão parece ter sido melhor resolvida, restando segmentações não-convencionais que, mais fortemente, tocam em aspectos ortográficos como o do hífen.

A terceira variável analisada pela autora foi a distribuição das segmentações não-convencionais em relação ao sexo/gênero dos escreventes. Para essa análise, Torquette (2016) baseou-se na hipótese de que os textos de escreventes masculinos teriam mais ocorrências de segmentações não-convencionais de palavras do que os textos de escreventes femininas. No entanto, os resultados mostraram que, tanto nas escolas públicas quanto nas privadas, ao longo dos cinco anos letivos, há uma diminuição do percentual de ocorrências de segmentações não-convencionais de palavras nos textos de escreventes femininos e masculinos.

Torquette (2016) também observa que os escreventes do sexo/gênero feminino segmentam de forma similar nas escolas públicas e nas escolas privadas, mas os escreventes do

sexo/gênero masculino não. Nas escolas públicas, os escreventes masculinos apresentaram uma mediana quatro vezes maior de ocorrências do que a mediana dos escreventes femininos. Nas escolas privadas, por outro lado, a diferença entre escreventes masculinos e escreventes femininos, além de bastante pequena, não se mostrou estatisticamente significativa. A autora conclui destacando a importância de se considerar as heterogeneidades das múltiplas tarefas envolvidas na relação que a constituição social dos gêneros tem com a distribuição social dos usos da escrita, pois, como os seus resultados mostram, as formas de segmentar dos escreventes têm forte “relação não apenas com o tipo de escola em que eles estudam, mas, também, com a complexidade da construção social dos gêneros e das tarefas sociais associadas a essa construção” (TORQUETTE, 2016, p. 137)

Passamos a caracterizar a escrita do EF II. Para tanto, descrevemos, a seguir, os resultados dos trabalhos de Tenani (2011) e Paranhos (2014) sobre hipossegmentação.

Tenani (2011) apresenta resultados de estudo sobre hipossegmentação em amostra de textos transversais produzidos por alunos no EF II. Nesse artigo, baseada na hipótese de que é possível observar características fonológicas, especialmente, as prosódicas, a partir de dados de segmentação não-convencional de palavras escritas, a autora propõe explicitar quais características prosódicas podem ser observadas nos dados selecionados. Para tanto, Tenani (2011) valeu-se das noções de pé métrico, palavra prosódica e grupo clítico.

Da amostra transversal do Banco de Dados de Escrita do EF II<sup>15</sup>, constituída de 2469 textos, Tenani (2011) selecionou 424 textos do início dos anos letivos dos quatro anos do EF II. A partir da análise desses textos, foi selecionado um total de 162 ocorrências de segmentações não-convencionais de palavras. De modo geral, Tenani (2011) constatou que há mais dados de hipossegmentação em relação aos de hipersegmentação e que os constituintes prosódicos relevantes para caracterizar as hipossegmentação e hipersegmentações encontradas nos textos do EF II diferem daqueles relevantes para os dados encontrados nos textos do EF I. No que diz respeito aos dados de hipossegmentação, no EF I, há a mobilização de todos os domínios prosódicos e mesclas entre domínios, como mostra Capristano (2007a). Já no EF II, por outro lado, há predomínio de mobilização de fronteiras de pé métrico, palavra prosódica e grupo clítico (como em “meamava” – “me amava”, “levala” – “levá-la”, “perrapado” – “pé-

---

<sup>15</sup> Assim como Tenani (2011) e Paranhos (2014), neste trabalho, também utilizamos o Banco de Dados de Escrita do EF II (como será descrito, de modo detalhado, no capítulo 2 desta dissertação); entretanto, o recorte do material realizado pelas autoras diferencia-se do que fazemos. Em nossa pesquisa, levamos em conta a totalidade de textos da amostra longitudinal que compõem o banco de textos escritos do EF II e fazemos análise estatística dos resultados, de modo que podemos confirmar ou refutar os resultados identificados pelas autoras a partir da consideração de um grupo pequeno de textos do EF II.

rapado”, “oque” – “o que”, “eporiço” – “e por isso”) e não se observou dados de escrita que envolvessem fronteiras de constituintes mais altos da hierarquia prosódica, como frase entoacional e enunciado fonológico.

Tenani (2011) também argumenta que as segmentações identificadas evidenciam aspectos do chamado processo de aquisição da escrita que, com base em CORRÊA (2004), podem ser vistos como marcas que resultam da conjunção das práticas orais/faladas e letradas/escritas e que permitem observar a presença do oral/falado no letrado/escrito. Assim, a autora defende que as características prosódicas das segmentações não-convencionais de palavra identificadas não corroboram a visão dicotômica das relações entre fala e escrita, mas sim a posição que toma a fala e a escrita como modos da língua entrelaçados com as práticas sociais orais e letradas de tal maneira que essas relações são heterogeneamente constituídas (TENANI, 2011).

Paranhos (2014) analisou dados longitudinais de segmentação não-convencional (hipersegmentações e hipossegmentações) em textos de sete alunos que cursaram os quatro últimos anos do EF. No que diz respeito aos resultados gerais, a autora afirma que as segmentações não-convencionais encontradas podem ser motivadas por possíveis estruturas prosódicas da língua, como a palavra fonológica e o grupo clítico. Sobre as hipossegmentações, alvo de nosso interesse, Paranhos (2014) mostrou evidências de o grupo clítico ser analisado como palavra fonológica, como em “derrepente” (“de repente”). A autora explica que, nesse caso, o clítico “de” é analisado como sílaba pretônica prosodizada à palavra prosódica “repente”.

A autora também afirma que os dados de segmentação não-convencional indiciam a organização prosódica da língua e, também, evidenciam características advindas de informações letradas/escritas. Segundo ela, a ausência de fronteira gráfica entre palavras pode ser relacionada à classe gramatical dessas palavras. Em seu estudo, Paranhos (2014) demonstra serem as preposições a classe mais envolvida na grafia dos clíticos nas hipossegmentações. Além desses resultados, a autora demonstra que os constituintes prosódicos relevantes para caracterizar as ocorrências de hipossegmentação envolvem fronteiras de palavra fonológica e grupo clítico, e poucos dados que envolvem constituintes mais altos da hierarquia prosódica (como a frase entoacional e o enunciado fonológico). Paranhos (2014) conclui que sua análise corrobora uma visão não dicotômica das relações entre fala e escrita, evidenciando que a fala e a escrita são modos de enunciação da língua permeados por práticas sociais orais e letradas (CORRÊA, 2004).

Em síntese, as pesquisas apresentadas nessa seção contribuem com essa dissertação, pois fornecem características linguísticas dos dados de hipossegmentação nos anos iniciais e finais do EF. Todas elas demonstram que dados de hipossegmentação mobilizam aspectos prosódicos dos enunciados, além de aspectos letrados/escritos. No EF I, há a mobilização de todos os domínios prosódicos, como mostram Cunha (2004, 2010), Cunha e Miranda (2007), Capristano (2007a) e Torquette (2016). No EF II, há predomínio de mobilização de fronteiras de pé-métrico, palavra prosódica e grupo clítico, como mostram Tenani (2011) e Paranhos (2014).

Além da mobilização de constituintes prosódicos, Capristano (2007a), Tenani (2011), Paranhos (2014) e Torquette (2016) defendem que dados de segmentação não-convencional de palavras podem ser interpretadas como marcas que resultam da conjunção das práticas orais/faladas e letradas/escritas e que permitem observar a presença do oral/falado no letrado/escrito. Nesta dissertação, comungamos dessa perspectiva de interpretação das hipossegmentações.

\*\*\*

Neste capítulo, discorreremos sobre os pilares teóricos que sustentam a presente pesquisa. No capítulo a seguir, descreveremos as principais características do material de análise, como se deu a constituição do corpus, quais foram os passos metodológicos para a identificação das ocorrências de hipossegmentação e as formas de análise dos resultados.

---

## CAPÍTULO 2

---

### MATERIAL E MÉTODOS

Neste capítulo, apresentamos o material do qual selecionamos dados para esta pesquisa e a metodologia empregada para seleção e exclusão de dados. Primeiramente, discorreremos a respeito da caracterização do Banco de Dados de Escrita do Ensino Fundamental II

Logo após, tratamos dos critérios nos quais nos baseamos para estabelecer o recorte do material e o levantamento dos dados de hipossegmentação. Buscamos, nessa seção, ilustrar como se deu a aplicação desses critérios, a fim de evidenciar como foram fundamentais quer para a superação de dificuldades iniciais quer para a identificação das hipossegmentações.

Na sequência, explicitamos as formas de análise dos resultados e os aspectos metodológicos que embasam a execução de tais formas.

#### 2.1 Material analisado

Os dados de hipossegmentação analisados foram identificados em uma amostra longitudinal de textos que integra o “Banco de Dados de Escrita do Ensino Fundamental II” (TENANI, 2015)<sup>16</sup>. O banco é constituído por 5.519 textos escritos por 662 alunos do EF II. Esse banco é composto por uma amostra transversal, constituída por 2.759 textos, e uma amostra longitudinal, constituída por 3.645 textos. Nesta pesquisa, selecionamos 2.435 (67%) textos da amostra longitudinal.

A constituição dessas amostras de textos se deu por meio do projeto de extensão “Desenvolvimento de oficinas pedagógicas de leitura, interpretação e produção textual no ensino fundamental”, coordenado pelas professoras doutoras Luciani Ester Tenani e Sanderléia Longhin-Thomazi (IBILCE/UNESP), e desenvolvido em uma escola pública estadual, situada em São José do Rio Preto, noroeste paulista, com a colaboração de alunos de graduação e de pós-graduação do IBILCE/UNESP.

---

<sup>16</sup> Disponível *online* (<http://www.convenios.grupogbd.com/redacoes/Login>) e pode ser acessado, após autorização das coordenadoras, por meio de um cadastro.

A produção dos textos era precedida de discussão coletiva de variados temas com base na leitura de textos de diversos gêneros em sala de aula. Todas as discussões e atividades desenvolvidas na escola foram pautadas em uma concepção de escrita enquanto modo de enunciação inserida em um processo de letramento vinculado às práticas de oralidade (CORRÊA, 2004), como mostram Tenani e Longhin-Thomazi (2014)<sup>17</sup>:

Nesse quadro teórico, as iniciativas do projeto foram conduzidas pela valorização da multiplicidade de letramentos associada às práticas sociais vivenciadas pelos alunos, pela consideração dos gêneros discursivos e pelo reconhecimento da importância de atentar ao processo de escrita do aluno e não apenas ao produto final. Portanto, visou-se ao alcance da escrita “institucionalizada” como, por exemplo, o texto escrito segundo as convenções ortográficas (TENANI; LONGHIN-THOMAZI, 2014, p. 22).

Com base nessa abordagem teórica sobre escrita, as propostas de produção de texto foram realizadas visando proporcionar aos alunos experiências de leitura e de escrita vinculadas a práticas sociais orais e letradas (TENANI; LONGHIN-THOMAZI, 2014). Essas propostas foram elaboradas, em geral, com base em uma coletânea de textos ou em uma tira de humor e, a partir desse material, os alunos eram convidados a discutir o tema em sala de aula. Posteriormente, a escrever seus textos seguindo as características de diferentes gêneros e tipos textuais, conforme previsto pelo programa curricular de cada ano letivo<sup>18</sup>.

É importante ressaltar que todos os textos foram escritos sem a possibilidade de reescrita. Desse modo produzidos, os textos, por hipótese, poderiam guardar marcas, mais explícitas, do trânsito do sujeito escrevente por práticas orais/faladas e letradas/escritas, como concebidas por Corrêa (2004).

Neste momento, cabe mencionar detalhes da constituição do banco relevantes para análise dos textos. Na época da constituição do banco, houve a ampliação dos anos do EF<sup>19</sup>, que, de oito anos, passou para nove. Nessa reformulação, o EF I passou a ter cinco anos de duração e o EF II continuou com quatro. Dessa maneira, as antigas quinta, sexta, sétima e oitava séries, passaram a ser, respectivamente, sexto, sétimo, oitavo e nono ano do EF. Neste trabalho, optamos por usar a nomenclatura atual, sem prejuízo para a presente pesquisa, uma vez que o conteúdo programático do EF II não sofreu mudanças. Vale ressaltar que os textos que fazem parte do banco de dados foram escritos por alunos que cursaram o EF de oito anos de duração.

---

<sup>17</sup> Tenani e Longhin-Thomazi (2014) trazem uma descrição da metodologia e das atividades desenvolvidas no projeto de extensão.

<sup>18</sup> A descrição dos gêneros e de tipos textuais que constituem o banco está disponível na página *web* do banco.

<sup>19</sup> Lei Federal Nº 11.274, de 2009 instituiu o Ensino Fundamental de nove anos.

Outro aspecto importante relativo aos sujeitos escreventes diz respeito aos índices de rendimento escolar (índice estadual – IDESP – e índice nacional – IDEB) descritos por Tenani (2016) para a escola na qual os textos do banco foram coletados. Segundo a autora, conhecer esses índices permite dimensionar o quão a amostra de textos do EF II é representativa de características da escrita do EF II no Brasil.

De acordo com Tenani (2016), ao longo dos quatro anos da coleta dos textos do banco, o IDESP da escola permaneceu um pouco acima da média geral do estado de São Paulo no mesmo período considerado, como informado na Tabela a seguir.

Tabela 1. IDESP da escola analisada e do estado

<b>Ano</b>	<b>IDESP da escola</b>	<b>IDESP do estado de SP</b>
<b>2008</b>	3,43	2,60
<b>2009</b>	3,70	2,84
<b>2010</b>	3,37	2,52
<b>2011</b>	3,40	2,57

Fonte: Elaboração a partir de dados apresentados por Tenani (2016)

Observa-se que tanto os índices da escola quanto os índices do Estado seguem uma mesma variação de aumento e decréscimo nos anos letivos considerados. Há um aumento de 0,27 pontos no índice de 2009 em relação ao de 2008, um decréscimo de 0,33 pontos em 2010 seguido de um aumento de 0,03 pontos em 2011. Ao descrever esses dados, a autora interpretou que os alunos que produziram os textos da amostra analisada representam as características gerais dos alunos do EF II do estado de São Paulo.

Quanto ao IDEB, Tenani (2016) mostra que a escola em questão alcançou índices um pouco acima da média do país em 2009 (escola 4,7 e país 3,8) e 2011 (escola 5,1 e país 3,9) mostrando, novamente, que os dados da amostra do EF II são representativos de características gerais do Brasil.

Como evidenciado até o momento, o “Banco de Dados de Escrita do Ensino Fundamental II” pode ser tomado como representativo da escrita do EF II e, desse modo, este trabalho visa contribuir com a caracterização da escrita do EF II<sup>20</sup>.

<sup>20</sup> Uma lista de publicações vinculadas ao banco estão disponíveis em: <http://www.ibilce.unesp.br/#!/departamentos/estudos-ling-literarios/laboratorio-de-fonetica/banco-de-dados/pesquisas-e-publicacoes/>

## 2.2 Constituição do objeto de análise

Com o intuito de atender ao objetivo de descrição de dados longitudinais deste trabalho, selecionamos alunos que tivessem ao menos um texto por ano letivo. Desse modo, as produções de um mesmo conjunto de alunos poderiam ser observadas de forma contínua no decorrer dos quatro anos do EF II. A partir desse recorte, chegamos a um total de 2.435 textos produzidos por 115 alunos<sup>21</sup>. A distribuição quantitativa desses textos pode ser observada na Tabela que segue.

Tabela 2. Número de textos selecionados

<b>Nº de textos</b>	<b>Alunos</b>	<b>Total de textos</b>
<b>26</b>	9	234
<b>25</b>	10	250
<b>24</b>	9	216
<b>23</b>	11	253
<b>22</b>	17	374
<b>21</b>	17	357
<b>20</b>	15	300
<b>19</b>	3	57
<b>18</b>	13	234
<b>17</b>	3	51
<b>16</b>	2	32
<b>15</b>	1	15
<b>14</b>	2	28
<b>13</b>	1	13
<b>12</b>	1	12
<b>9</b>	1	9
<b>Totais</b>	115	2435

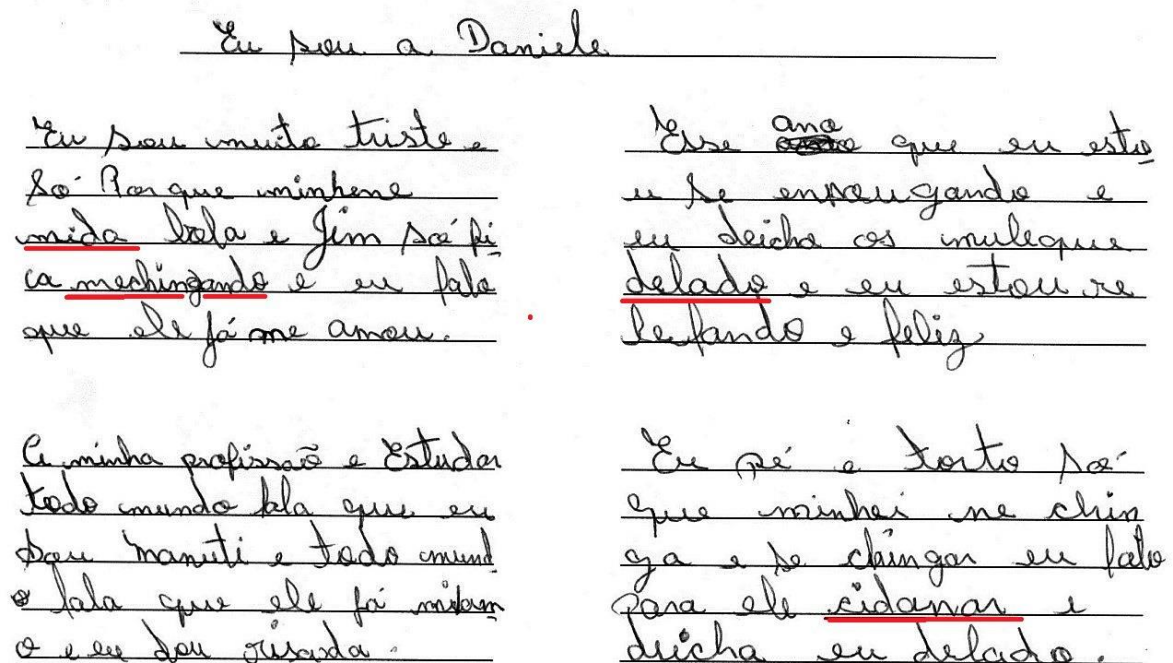
A Tabela 2 mostra que: (i) a maioria dos sujeitos (77%) produziu entre 20 e 26 textos (77% e 100%) nos quatro anos de coleta dos textos; (ii) aproximadamente um quarto dos sujeitos (27% restantes) produziu entre nove textos e 19 textos (25% a 73%) durante os quatro anos de coleta.

<sup>21</sup> Cabe observar que essa tarefa foi facilitada pelo fato de todos os textos do banco estarem digitalizados e transcritos ortograficamente.

Uma vez conhecida a distribuição de textos ao longo do EF, iniciamos a fase de identificação e revisão dos dados de hipossegmentação presentes nos textos selecionados. Para essa fase, adotamos dois critérios para a identificação dos dados: (i) comparação dos espaços em branco entre palavras ao longo do texto; (ii) comparação entre as grafias das mesmas letras dentro de palavras semelhantes que ocorram no mesmo texto, conforme metodologia descrita em Tenani (2016).

A seguir, trazemos exemplos de como os critérios foram empregados nos textos:

Figura 7. Exemplo de critério de identificação de dados de segmentação não-convencional



Fonte: Banco de Dados de Escrita do Ensino Fundamental II (Z08\_5D\_05F\_02)

Na Figura 7, são facilmente identificadas ocorrências de hipossegmentação, já que, onde deveria haver um espaço em branco, há uma ligadura bem visível. Na Figura 8, também é fácil a identificação da hipossegmentação; entretanto, ao invés do espaço em branco, deveria haver um hífen.

Figura 8. Exemplo de critério de identificação de dados de segmentação não-convenional

juvência jogou a mulher de concha se  
 que na hora da juvência ficou com pena  
 pegou ela de volta e tomou a concha  
 E ela ficou falando cora cora e ju-  
 vência conseguiu despertar o la-  
talhão

Fonte: Banco de Dados de Escrita do Ensino Fundamental II (Z08\_5D\_09M\_01)

Nos exemplos acima, os casos de hipossegmentação fáceis de serem identificados se caracterizam por ausência do espaço em branco e do hífen. A seguir, trazemos um texto em que foi difícil a identificação de dados de hipossegmentação por haver dúvida entre um aspecto caligráfico ou um aspecto ortográfico (TENANI, 2016):

Figura 9. Exemplo de critério de identificação de dados de segmentação não-convenional

Oi, primo

Comerá por aí, e a família estábam? Sinto muitas saudades? Você -  
se diverte como?

Eu sempre eu pouco pouco, nada no rio e me divirto muito na  
escala.

Você e se temos que estudar muito e alios. Você me mãina  
 a mecha na internet e manda mensagem pela msn quando você  
vai a qui recita?

thau, primo

Sinto muito calor até o rio está borbulhando nos termos até de  
 colorar gelo na boca os termos até flegma separar  
 gelo na reflexão, o unico jeito e de baixo de uma  
 avale que e floquinho de baixo, de noite eu fico na  
 frente da geladeira a noite, e do tao calor amato eu dormo  
 sem o deito e com a janela aberta. Manda um abraço pra toda  
 mundo

Fonte: Banco de Dados de Escrita do Ensino Fundamental II (Z08\_5D\_04M\_04)

Na Figura 9, há vários casos de difícil classificação, a começar pela sequência “comovai” (“como vai”). Essa sequência trouxe dúvida, já que parece ser tendência do aluno escrever palavras com o espaço em branco de diferentes tamanhos, além de haver flutuação entre letra cursiva e letra de forma. Assim, para resolver essa dúvida quanto à delimitação de

palavra, comparamos as grafias das mesmas letras em palavras semelhantes que ocorriam nesse texto.

Observamos que a caligrafia da letra <o> em todas as palavras do texto tem um traço, como em “o coberto” (“o cobertor”). Por essa razão, a princípio, pensamos que a sequência “comovai” não se tratava de um dado de hipossegmentação; porém, após analisar a letra <v> notamos que ela está inclinada, diferentemente do que acontece com essa letra nas outras palavras, como em “você” e “viu”. Entendemos que essa grafia é indício de que essas duas palavras foram grafadas juntas. Assim, a sequência “comovai” foi classificada como um dado de hipossegmentação.

Outra grafia que causou dúvidas foi a sequência “por aí”, pois nota-se que o espaço entre essas duas palavras é muito pequeno; porém, ao observar outras palavras grafadas com as letras <r> e <a>, como em “pra refletar” (“para refrescar”), chegamos à conclusão de que este escrevente tem tendência de unir essas letras quando as grafa. Assim, nesse caso, “porai” não foi considerado um dado de hipossegmentação.

As grafias “estabem”, “euporezemplo”, “medivirto” e “anoite” (“está bem”, “eu por exemplo”, “me divirto” e “à noite”) foram consideradas dados de hipossegmentação, uma vez que o espaço em branco entre as palavras é quase inexistente.

A última grafia, nesse texto, que gerou dúvida foi “no citio” (“no sítio”). Verifica-se, novamente, a letra <o> com a ligadura; porém, diferentemente do dado “comovai”, a letra <c> não tem nenhum indício que nos faz ter certeza de que as duas palavras foram escritas como uma só. Desse modo, não classificamos a sequência “no citio” como um dado de hipossegmentação.

Esse exercício de análise ancora-se em uma perspectiva que toma os textos escritos como uma fotografia que captura um momento de um processo (TENANI, 2016). Naquilo que nos interessa, é o processo pelo qual o aluno é atravessado pelo código escrito institucionalizado, ou seja, a grafia resulta do que ele imagina ser – a partir de suas experiências com a escrita e com a própria visão escolar sobre a escrita – o código escrito institucionalizado (CORRÊA, 1997, 2004).

Somam-se a essa perspectiva, considerações de Tenani (2016), de que é a partir de informação verbo-visual contida no texto que podem ser elaboradas hipóteses sobre o trabalho epilinguístico do sujeito sobre a linguagem; no caso deste trabalho, sobre a relação entre fala e escrita.

É importante mencionar que, para esta dissertação, adotamos dois critérios de exclusão de dados, a saber: (i) exclusão de palavras homônimas, como “afim”, quando previsto “a fim”; (ii) exclusão de dados de mesclas, como “pora qui” (“por aqui”).

O primeiro critério de exclusão foi adotado por dois motivos: (1) quantitativo, já que há muitos dados de grafias que envolvem homônimos e isso gera enviesamento da amostra; e (2) morfossintático, pois casos de hipossegmentação que geram homonímias se particularizam em relação aos demais dados: há duas grafias que correspondem a uma mesma cadeia fônica. Desse modo, é o critério morfossemântico que garante a escolha de uma e não de outra forma ortográfica, como em “enfrente” e “em frente”.

Mesmo excluindo as grafias que resultam em homonímias, não deixamos de quantificá-las. Assim, nos textos analisados, encontramos um total 272 dados de palavras hipossegmentadas que geram homônimas que ocorrem em todos os quatro anos analisados, como mostra a tabela a seguir:

Tabela 3. Quantidade de dados de palavras homônimas

Ano	Homonímias	Palavras escritas	Índice de homonímias/ palavras escritas
6º	46	69.502	0,07
7º	111	96.069	0,12
8º	41	56.653	0,07
9º	74	12.4513	0,06
<b>Total</b>	<b>272</b>	<b>346.737</b>	<b>0,31</b>

A partir da Tabela 3, notamos que o maior índice (0,12) foi no sétimo ano, tendo um decréscimo no oitavo e no nono ano, com índices semelhantes ao do sexto ano. Portanto, nesse levantamento não houve decréscimo de emprego de hipossegmentações que geram grafias homônimas. Dentre esses dados, 237 (87%) são a junção da preposição “a” mais o substantivo “gente” gerando a grafia “agente”, a qual passa a ser homônima a “agente”, cujo significado amplo é “aquele que age”. Esse número mostra que os alunos do EF II têm dificuldades no uso de “a gente”, grafando “agente”, uma forma para representar a noção gramatical de primeira pessoa plural, como em “agente chegou”. Constatamos haver flutuação<sup>22</sup> entre as duas grafias. Esse fato demanda estudo específico de grafias que geram formas homônimas no EF II, o que já está sendo desenvolvido no âmbito do grupo de pesquisa a que esta dissertação se filia.

<sup>22</sup> Com base em Chacon (2013), utilizamos o termo *flutuação* para designar situações em que uma mesma estrutura é grafada de formas distintas.

O segundo critério relativo à exclusão das mesclas justifica-se por esse tipo de dado envolver tanto a hipossegmentação quanto a hipersegmentação, como em “a olado” (“ao lado”). Esses dados são considerados um terceiro tipo de ocorrência de segmentação não-convencional de palavra, merecendo um estudo específico, o que se encontra em curso no âmbito por membros do grupo de pesquisa a que esta pesquisa se filia. Limitamo-nos a observar que esse tipo de dado apareceu somente duas vezes, uma vez no 6º ano (“pora qui” – “por aqui”) e uma vez no 8º ano (“a olado” – “ao lado”).

Atendendo aos critérios descritos acima, chegamos a um total **750** dados de hipossegmentação, sendo 693 ocorrências com a ausência do branco e 57 ocorrências com a ausência do hífen<sup>23</sup>.

### 2.3 Formas de análise

As hipossegmentações consideradas passaram por análise quantitativa, uma vez que, como já mencionado, nossos objetivos são:

- i. descrever a tendência em ocorrer hipossegmentação longitudinalmente;
- ii. descrever as características prosódicas e morfossintáticas dos dados levantados.

A análise quantitativa foi adotada com o objetivo de observar o comportamento, por sujeito e ao longo dos anos letivos, dos dados de hipossegmentação.

Para o procedimento desta análise, optamos por adotar um tratamento estatístico dos dados, a fim de descrever características do conjunto analisado, possibilitando chegar a conclusões com base na informação obtida de amostras de textos do EF II.

Para o tratamento estatístico realizado neste trabalho, consideramos as seguintes variáveis:

- i. total de hipossegmentação em cada ano do EF II;
- ii. total de palavras escritas em cada ano do EF II;
- iii. índice gerado a partir do total de hipossegmentação e o total de palavras escritas em cada ano do EF II.

---

<sup>23</sup> Além desses dados, também encontramos 17 dados com a presença não-convencional do hífen (*estava-lá, podesse-se*) entre palavras. Não consideraremos esse tipo de grafia como um dado de hipossegmentação, mas dado de uso não-convencional de hífen, objeto de outra pesquisa em curso no âmbito do grupo de pesquisa “Estudos sobre a linguagem”.

Para a contagem de palavras, os textos foram digitalizados e transcritos ortograficamente<sup>24</sup>, assim, o número total de palavras considerado foi o disponibilizado pelo *software word*, o qual considera todas as palavras do texto.

O tratamento estatístico foi realizado por meio dos *softwares Excel* (versão de 2016) e *Minitab* (versão 2017). Além desse tratamento, aplicamos, nos dados, o teste paramétrico ANOVA e o teste *Tukey*, os quais descreveremos mais adiante. Os resultados obtidos a partir desses procedimentos estatísticos serão apresentados no capítulo 3: Análise dos dados.

Na sessão 3.2, apresentamos os dados classificados a partir de dois critérios de análise, a saber: (i) configuração prosódica dos dados: diz respeito à organização das estruturas hipossegmentadas em constituintes prosódicos, como grupo clítico e palavra prosódica, por exemplo; e (ii) configuração morfossintática: se refere à classificação das hipossegmentações em palavras gramaticais e palavras lexicais, identificadas as classes de palavras.

Para a classificação dos dados de hipossegmentação levantados, nos baseamos no aparato teórico-metodológico da Fonologia Prosódica de Nespor e Vogel (1986, 2007). Essa classificação se deu a partir da estrutura prosódica da palavra convencional, por exemplo, “meajude” = clítico + palavra prosódica. A consideração da estrutura prosódica mobilizada nas junções observadas parte da hipótese de que estruturas que apresentam clíticos fonológicos favorecem a hipossegmentação.

Ademais, na análise prosódica também consideramos informações como o número de sílabas da palavra escrita e a localização do acento lexical de palavra. No que se refere ao número de sílabas, as palavras são classificadas, segundo a tradição gramatical, em dissílabos, trissílabos ou polissílabos. Essa informação foi considerada como critério para identificar regularidades com relação ao tamanho tanto das palavras convencionais quanto das pseudopalavras geradas a partir da hipossegmentação.

Segundo Tenani (2016), considerar esse tipo de informação é importante para mensurar o quanto é o tamanho de palavra um fator relevante para a descrição da produção escrita do EF II em comparação ao que já foi afirmado para a produção escrita do EF I. Trabalhos como o de Ferreira et al. (1996), por exemplo, revelam que crianças em fase inicial de aquisição da escrita têm dificuldades em grafar palavras com mais de três sílabas ou com apenas uma sílaba. A questão a ser respondida é se o tamanho da palavra é relevante para a caracterização da produção escrita do EF II.

---

<sup>24</sup> A tarefa de digitalizar e transcrever os textos do Banco de Dados do EF II foi realizada por bolsistas de auxílio à Pesquisa FAPESP (processo 2013/14546-5).

Quanto à informação da localização do acento na palavra, segundo a tradição gramatical, as palavras podem ser classificadas em oxítonas, paroxítonas ou proparoxítonas. Levamos essa informação em consideração, porque a localização do acento de palavra leva à configuração de diferentes pés-métricos, como evidenciado na seção 1.4, na qual descrevemos o funcionamento dos constituintes prosódicos que fazem parte da hierarquia prosódica.

Para realizar a análise morfossintática, consideramos os tipos de palavras, se gramatical ou lexical e efetuamos uma classificação em classes gramaticais. Para tanto, levamos em conta a relação entre as palavras que compõem a estrutura hipossegmentada (por exemplo, “meajude” = pronome + verbo). Essa classificação mostra-se importante, pois, além de evidenciar quais categorias de palavras são lugares de dúvida para os escreventes, dão suporte para análises que revelam a complexidade dos dados de escrita quanto aos aspectos morfossintáticos.

Nesta dissertação, também apresentamos uma análise de cunho linguístico-textual de um conjunto de dados que revelaram, de modo mais evidente ao analista, o trânsito do sujeito por práticas orais/faladas e letradas /escritas e o quão complexo é trabalho do sujeito com a língua(gem). Para essa análise, selecionamos textos nos quais os dados selecionados ocorreram e descrevemos as propostas textuais que os motivaram. Essa análise é apresentada na seção 3.3 e 3.4 deste trabalho.

\*\*\*

Nesse capítulo, descrevemos as características do material de análise, os procedimentos de identificação e delimitação do objetivo de análise e os métodos de análise empregados em nossa pesquisa. No próximo capítulo, analisamos as hipossegmentações levantadas.

---

## CAPÍTULO 3

---

### ANÁLISE DOS DADOS

Neste capítulo, apresentamos a análise dos dados de hipossegmentações de palavras. Por meio dessa análise buscamos: (i) apresentar tendências gerais e possíveis regularidades da escrita de alunos que cursaram os quatro anos do EF II; (ii) descrever características prosódicas e morfossintáticas dos dados apresentados e discutir possíveis motivações para a ocorrência desses dados, (iii) realizar análise de cunho linguístico-textual de quatro textos que, sob nosso olhar, constituem-se importantes indícios da complexidade inerente às segmentações não-convencionais de palavras escritas.

Para tanto, organizamos este capítulo em três seções. Na seção 3.1, apresentamos a análise quantitativa, por meio da qual buscamos estabelecer tendência longitudinal de ocorrência de hipossegmentação no EF II. Em seção subsequente, 3.2, descrevemos características prosódicas e morfossintáticas dos dados e buscamos identificar indícios da circulação do escrevente por práticas orais/faladas e letradas/escritas. Para encerrar essas análises, tecemos considerações sobre os tipos de palavras segmentadas no EF II, apontando regularidades. Na seção 3.3, realizamos análises qualitativas de hipossegmentações com o intuito de argumentar que segmentações não-convencionais de palavras escritas são decorrentes de um complexo processo que envolve a escrita, o sujeito e o outro, enquanto instância de representação do funcionamento da linguagem.

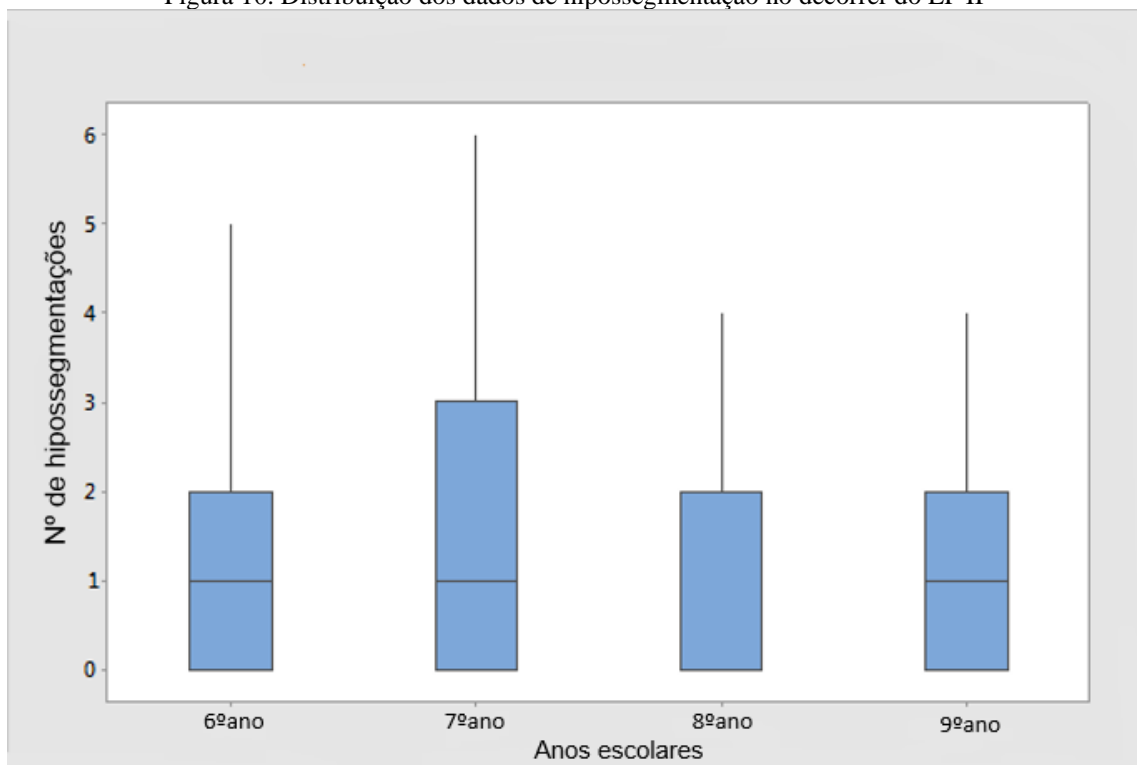
#### **3.1 Hipossegmentação: tendências quantitativas de dados longitudinais**

Nesta seção, tratamos da análise quantitativa dos dados de hipossegmentação. Para tanto, nos baseamos na informação sobre o número de ocorrências de hipossegmentação ao longo dos quatro anos do EF II; posteriormente, levamos em conta, assim como Tenani (2016), o coeficiente gerado a partir da razão entre o número de hipossegmentações e o número de palavras escritas ao longo dos quatro anos analisados.

É nosso objetivo descrever a tendência que os dados de hipossegmentação apresentam no decorrer dos quatro anos do EF II. Pretendemos demonstrar que o aumento da escolarização leva à diminuição das ocorrências de hipossegmentações. Para realizar essa tarefa, os dados de hipossegmentação foram tratados a partir de análise estatística<sup>25</sup>.

No material analisado, levantamos um total de 750 dados de hipossegmentação. Esses dados foram submetidos a tratamento estatístico realizado por meio dos *softwares Excel* (2016) e *Minitab* (17). A Figura 10 apresenta a distribuição dos dados de hipossegmentação no decorrer dos quatro anos considerados.

Figura 10. Distribuição dos dados de hipossegmentação no decorrer do EF II



Na Figura 10, visualizam-se duas informações:

- i. As medianas<sup>26</sup> dos quatro anos do EF II são aproximadamente iguais, com exceção do oitavo ano;
- ii. As amplitudes interquartis<sup>27</sup> também são aproximadas, com exceção do sétimo ano.

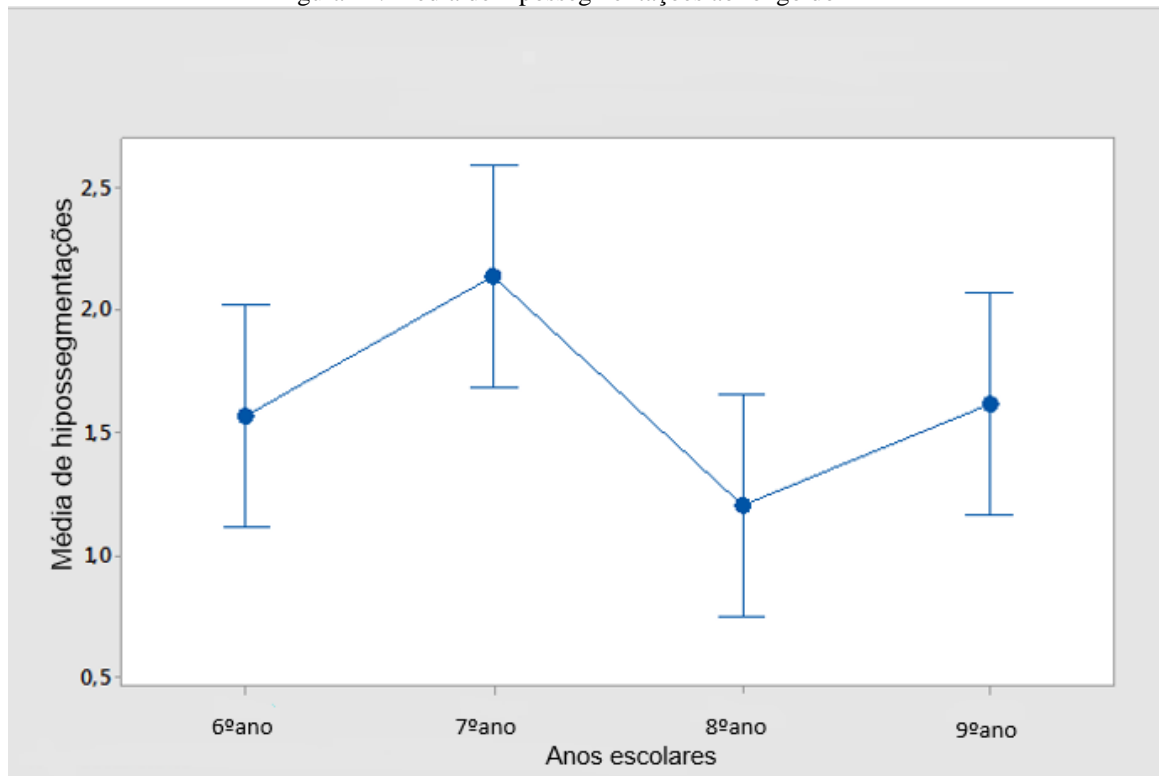
<sup>25</sup> As análises estatísticas foram conduzidas com o auxílio da Profa. Ms. Idalice Carvalho Figueiredo Rillo, a quem muito agradecemos.

<sup>26</sup> A mediana é o valor da variável que ocupa a posição central em um conjunto de dados ordenados. Na Figura 1 ela é caracterizada pela linha horizontal que atravessa as colunas de cada ano.

<sup>27</sup> “Interquartil” significa a diferença entre os quartis  $Q_3$  e  $Q_1$ . O quartil 3 é igual a 75% dos dados e o quartil 1 é igual a 25% dos dados.

Esse resultado traz indícios de uma possível variação entre os quatro anos do EF II que parece se centrar, principalmente, no sétimo e oitavo ano. Essa possível variação também pode ser contemplada a partir de valor das médias calculadas com intervalo de confiança de 95%, como ilustrado na figura seguinte:

Figura 11. Média de hipossegmentações ao longo do EF II



Observamos, na Figura 11<sup>28</sup>, que a maior média de dados ocorre no sétimo ano e a menor no oitavo ano. As médias e os intervalos do sexto e nono se assemelham, havendo uma aproximação entre esses anos. Esse resultado revela que, ao contrário do esperado, os anos de escolarização não contribuem para a diminuição dos dados de hipossegmentação. Os alunos parecem terminar a etapa do EF II com praticamente a mesma média de hipossegmentações com que começaram.

Após essa análise, aplicamos o teste ANOVA<sup>29</sup> (Análise de variância). Esse teste resultou em um valor de p menor que 0,05 ( $p = 0,036^{30}$ ), o que constitui em resultado estatístico para rejeitar a igualdade entre as médias dos grupos. Faz-se necessário pontuar que o teste

<sup>28</sup> Na Figura 11, o ponto azul representa a média de cada ano; entre as hastes superior e inferior, indicamos o intervalo em que as médias dos dados podem ocorrer. Este é chamado de intervalo de confiança da média.

<sup>29</sup> Ao nível de significância de 5% e nível de confiança 95%.

<sup>30</sup> Quanto maior for o valor de F, mais significativa é a diferença entre as amostras. Já o valor de p informa se existem diferenças estatisticamente significantes entre as amostras. Se o valor p for abaixo de 0,05 é bastante provável que qualquer diferença entre os grupos seja uma diferença real e não resultado do acaso.

ANOVA dá o embasamento estatístico necessário para dizer se existem diferenças entre as amostras; entretanto, ele não mostra em quais anos essas diferenças ocorrem. Com o intuito de determinar em qual ano ou anos reside essa diferença, aplicamos o teste paramétrico *post hoc Tukey*<sup>31</sup>. Os resultados do teste mostram diferença significativa entre as médias de hipossegmentação no 7º ano e 8º ano ( $p = 0,022$ ), como vemos na Tabela a seguir:

Tabela 4. Resultado do teste *post hoc Tukey* para 115 alunos.

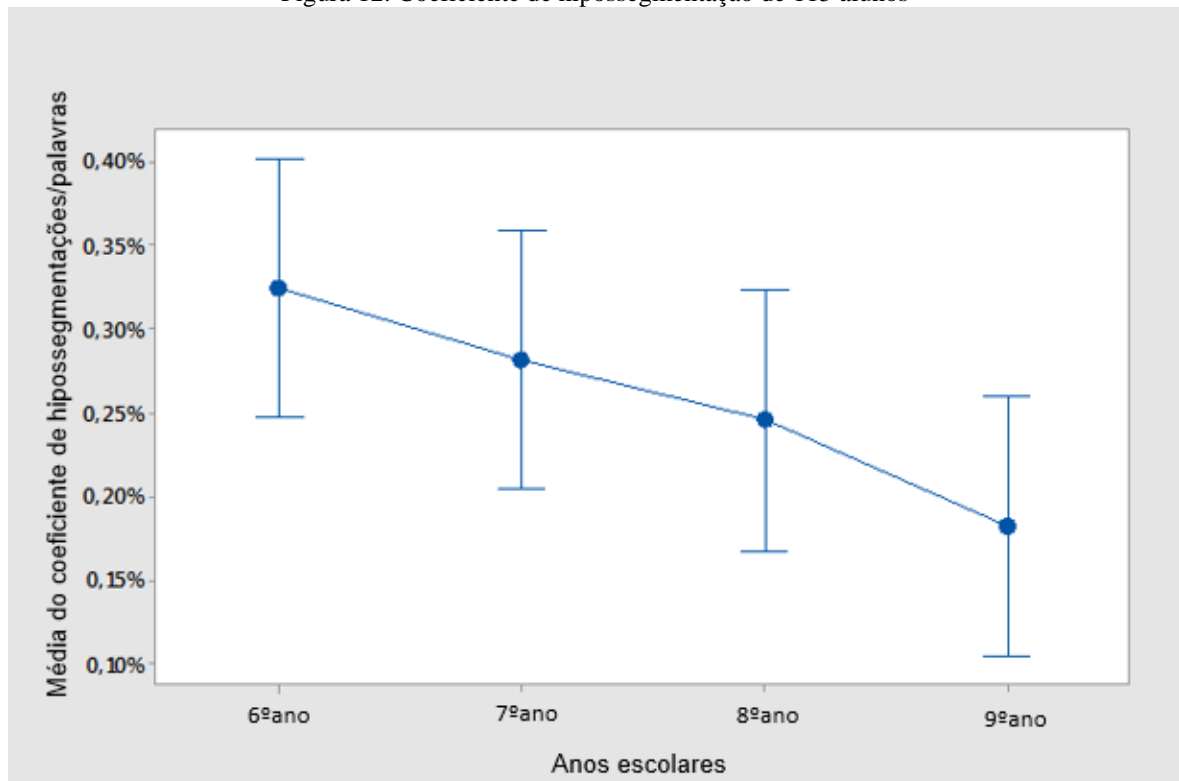
<b>Anos escolares</b>	<b>Valor de T</b>	<b>Valor de p</b>
<b>7º ano - 6º ano</b>	1,75	0,297
<b>8º ano - 6º ano</b>	-1,11	0,680
<b>9º ano - 6º ano</b>	0,16	0,999
<b>8º ano - 7º ano</b>	<b>-2,87</b>	<b>0,022</b>
<b>9º ano - 7º ano</b>	-1,59	0,383
<b>9º ano - 8º ano</b>	1,27	0,579

Por meio desse resultado, a interpretação é que há uma diferença na quantidade de hipossegmentações do sétimo ao nono ano do EF II. Esse resultado muda se, assim como Tenani (2016), considerarmos a quantidade de palavras escritas em cada texto e o coeficiente<sup>32</sup> de ocorrência de hipossegmentação que leva em conta a razão entre o total de hipossegmentação e o total de palavras escritas ao longo dos quatro anos do EF II. O resultado obtido segue uma tendência decrescente de hipossegmentação em função do aumento do tempo de escolarização, como pode ser visualizado na Figura a seguir.

<sup>31</sup> A literatura estatística preconiza que o teste *post hoc Tukey* é indicado para amostras com variâncias aproximadas, como o caso da amostra deste trabalho.

<sup>32</sup> Para obter esse coeficiente, dividimos o total de dados de hipossegmentação de cada sujeito pelo total de palavras e multiplicamos por 100, seguindo metodologia adotada por Tenani (2016). Exemplo: o sujeito 27 tem 29 dados e escreveu 2.228 palavras durante do EF II, assim temos  $(29/2.229) \times 100 = 1,30$ .

Figura 12. Coeficiente de hipossegmentação de 115 alunos



A consideração do coeficiente de hipossegmentações por palavras escritas permite observar que, diferente do exposto anteriormente na Figura 11, quando considerado apenas as hipossegmentações sem se estabelecer relação com o total de palavras escritas por aluno, há resultados que sinalizam para um decréscimo das médias ao longo dos quatro anos em questão. A descrição ora realizada na forma de quantificar os dados demonstra que a afirmação de alto número de ocorrência de junção entre palavras não é critério suficiente para validar diagnóstico de “distúrbio específico de escrita”, uma prática bastante comum em ambiente escolar ou clínico. A consideração da relação da ocorrência da junção e o tamanho do texto escrito se mostra, pois, como abordagem mais segura para afirmações sobre o que pode ser tomado como característico da escrita.

Ao avançarmos no tratamento estatístico aos dados, optamos por eliminar da amostra os alunos cujos textos não apresentaram dados e os alunos em cujos textos havia somente uma hipossegmentação. Esses últimos dados foram eliminados pelo fato de a razão entre o total de hipossegmentações e o total de palavras escritas gerar um coeficiente muito baixo de hipossegmentação por aluno. Dessa maneira, passamos a considerar os textos de alunos com dois ou mais dados de hipossegmentação e ficamos com um total de 737 ocorrências de hipossegmentação, tendo sido eliminadas da amostra apenas 13 ocorrências. A partir desse

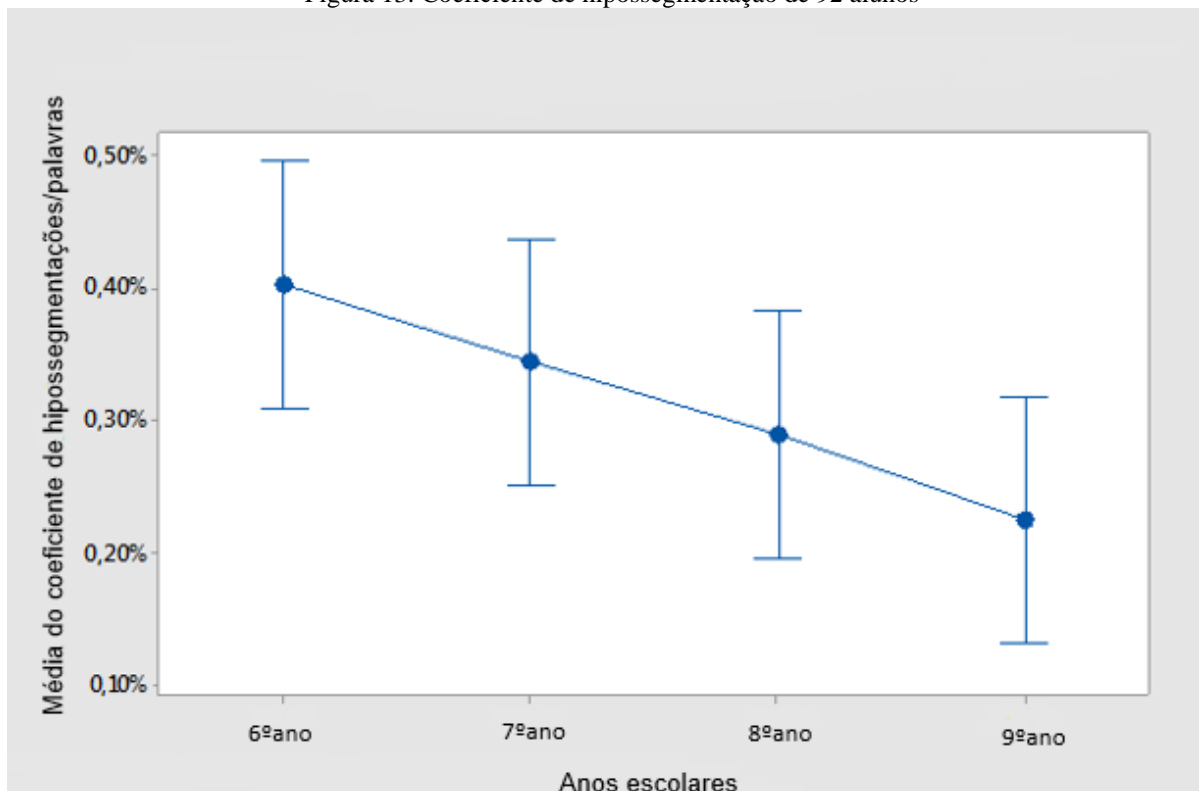
recorte, obtivemos um decréscimo das médias ao longo do EF II, como vemos na Tabela 5 a seguir:

Tabela 5. Coeficiente de hipossegmentação de 92 alunos

Ano	Média	Desvio Padrão	Intervalo de confiança (95%)
6º	0,00403	0,006226	(0,003087; 0,004962)
7º	0,00344	0,003954	(0,002503; 0,004378)
8º	0,00289	0,004226	(0,001953; 0,003828)
9º	0,00225	0,003369	(0,001313; 0,003188)

A Tabela 5 evidencia que o decréscimo das médias, que passaram de 0,00403 no sexto ano, para 0,0025 no nono ano. Esse resultado segue ilustrado na Figura 13:

Figura 13. Coeficiente de hipossegmentação de 92 alunos



Na Figura 13, observamos que o coeficiente de hipossegmentações de 92 alunos com duas ou mais hipossegmentações nos quatro anos em estudo continua gerando um decréscimo nas médias em função do aumento dos anos letivos, resultado que sugere que a diferença entre essas médias existe. Para esse conjunto de alunos, o teste ANOVA resultou em  $p$  igual a 0,052. Devido à proximidade deste valor com o nível de significância  $\alpha = 0,05$ , aplicamos o teste *post hoc Tukey* a fim de verificar as diferenças entre os pares das médias dos coeficientes. O teste

mostrou diferença entre as médias dos coeficientes de hipossegmentações do nono e do sexto ano ( $p = 0,042$ ). Esse resultado pode ser visualizado na Tabela 6:

Tabela 6. Resultado do teste *post hoc Tukey* para 92 alunos.

Anos escolares	Valor de T	Valor de p
<b>7º ano - 6º ano</b>	-0,87	0,822
<b>8º ano - 6º ano</b>	-1,68	0,333
<b>9º ano - 6º ano</b>	<b>-2,63</b>	<b>0,042</b>
<b>8º ano - 7º ano</b>	-0,82	0,847
<b>9º ano - 7º ano</b>	-1,77	0,290
<b>9º ano - 8º ano</b>	-0,95	0,778

Com base nesse resultado estatístico, alcançamos um de nossos objetivos específicos demonstrando que há correlação estatística significativa entre aumento dos anos escolares e diminuição do número hipossegmentações.

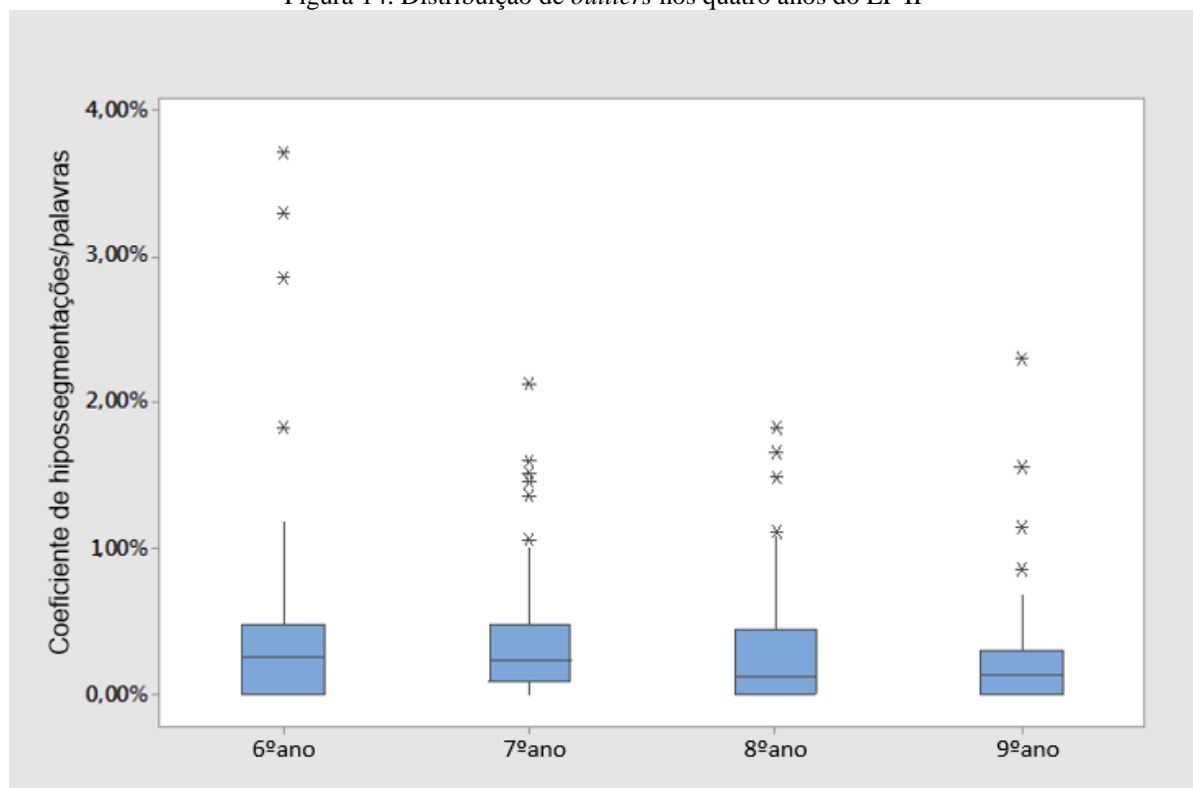
Esse resultado também nos leva a chamar a atenção para a importância de se considerar o total de palavras em trabalhos sobre segmentação de palavras na caracterização de incidência de segmentação não-convencional de palavras. Essa importância também foi evidenciada nos trabalhos de Paranhos (2014) e de Torquette (2016) citados anteriormente. Paranhos (2014) se vale do total de palavras para mostrar que não há uma correlação linear entre número de textos e número de palavras por texto e para mostrar que entre o número de palavras e o número de segmentações não-convencionais de palavras não há uma correlação linear, pois, o percurso de ambas não se dá de forma proporcional. Em Torquette (2016), também foi feita a razão entre o número de segmentações não-convencionais de palavras e o número de palavras dos textos, todavia, o objetivo da autora foi o de realizar comparação entre os tipos de escolas para que não houvesse discrepância em função de o texto ser mais extenso e, por isso, conter mais ocorrências de segmentações não-convencionais.

Cabe observar ainda que nem todos os trabalhos sobre segmentação de palavras escritas têm mesma abordagem metodológica na descrição de ocorrências de dados. Dentre esses, podemos citar os trabalhos de Cunha (2004, 2010), Cunha e Miranda (2007) e Capristano (2007a) descritos anteriormente. Essas diferenças metodológicas são, em parte, motivadas pelos distintos objetivos das pesquisas realizadas, mas devem ser consideradas quando da comparação dos resultados para comparação da escrita de EF I e EF II.

Feita essa caracterização geral, é importante destacar que o aumento da escolarização não significa o fim das hipossegmentações, já que os alunos terminam o EF II com dados de hipossegmentação, como observado por Tenani (2011) em uma amostra transversal. Além dessa ponderação, demonstramos, na seção seguinte, que certos tipos de dados não sofrem esse decréscimo, mas um aumento quantitativo ao longo do EF II.

Por fim, vale observar que, ao lado dessa tendência quantitativa geral, identificamos alunos que produzem muitas hipossegmentações e que estaticamente podem ser considerados *outliers*, como observamos na Figura a seguir.

Figura 14. Distribuição de *outliers* nos quatro anos do EF II



A partir da Figura 14 é possível afirmar que houve alunos estatisticamente *outliers* em todos os anos do EF II. Há uma média de quatro *outliers* por ano, exceto no sétimo ano, no qual se encontram seis *outliers*. Desse modo, 20% (18/92) dos sujeitos investigados apresentaram muitos dados de hipossegmentações e são considerados *outliers*. É importante esclarecer que não são sempre os mesmos alunos que se comportam como *outliers* em todos os anos do EF II, isto é, um mesmo aluno pode ser considerado *outlier* em um ano, mas não em outro devido à quantidade de dados do aluno naquele ano e o total de ocorrências de todos os alunos daquele mesmo ano. Desse modo, dentro desse subconjunto 18 de sujeitos, encontram-se quatro

sujeitos que são *outliers* em mais de um ano do EF II, conforme esquematizado na Tabela seguinte:

Tabela 7. Relação de sujeitos considerados *outliers*

6º ano	7º ano	8º ano	9º ano
Nº	Nº	Nº	Nº
Sujeito	Sujeito	Sujeito	Sujeito
16	90	93	69
60	9	112	107
115	77	27	9
27	73	25	25
	16		
	25		

A partir da Tabela 7, observamos que:

- i. Um sujeito (25) é *outlier* em três anos do EF II;
- ii. Três sujeitos (27, 16 e 9) são *outliers* em dois anos do EF II.
- iii. Não há nenhum sujeito *outlier* em todos os anos do EF II.

Essa breve descrição dos resultados quantitativos obtidos em relação à identificação de *outliers* nos levam a apontar a necessidade de estudo específico para esse conjunto de sujeitos, tendo por base a descrição geral ora realizada. Dados nosso objetivo de descrever características típicas do EF II e, também, a fim de não enviesar a análise quantitativa quando considerada a relação entre número de dado por sujeito da pesquisa, optamos por eliminar do conjunto de dados as hipossegmentações dos alunos considerados *outliers*. Entretanto, essa exclusão foi feita somente no ano em que o aluno foi considerado *outlier*, pois, como já mencionado, um mesmo aluno pode ser considerado *outlier* em um ano, mas não em outro. Assim, após a eliminação de 187 ocorrências dos 18 alunos considerados *outliers*, passamos a analisar, nesta dissertação, o total de **550** dados de hipossegmentação.

A partir desse resultado quantitativo, podemos afirmar que este trabalho contribui com os estudos de segmentação de palavra escrita na medida em que mostra que considerar a quantidade de palavras escritas por texto e a distribuição de ocorrências por sujeito são informações importantes para melhor fundamentar uma análise longitudinal. Essa abordagem metodológica nos auxiliou na identificação de um conjunto de dados produzidos por sujeitos que evidenciaram um comportamento quantitativo diferente dos demais; por essa razão esses

dados não serão considerados no conjunto investigado e devem ser analisados qualitativamente em separado.

Para concluir esta seção, com base nos resultados quantitativos, interpretamos que o fato de os alunos terminarem o EF II ainda produzindo hipossegmentações não significa, necessariamente, que eles não aprenderam as convenções ortográficas quanto à segmentação de palavra. Esse resultado pode ser interpretado, de nossa perspectiva, como decorrendo de atuação de critérios de segmentação de palavras originados por práticas letradas (de leitura e escrita) desenvolvidas, mas não concluídas enquanto processo de aprendizagem, em cada ano escolar do EF II.

Dentre as práticas letradas experienciadas, supomos ser possível identificar, nas produções escritas, parte daquelas práticas que abrangem informações gramaticais trabalhadas em sala de aula. Partindo dessa suposição, faz-se necessário analisar qualitativamente esses dados, quanto à atuação de estruturas prosódicas e morfossintáticas de modo complexo.

Desse modo, na seção seguinte, trazemos a análise das características prosódicas e morfossintáticas dos dados de hipossegmentação considerados.

### **3.2 Hipossegmentação: características prosódicas**

Nesta seção, nosso objetivo é apresentar, a partir de um ponto de vista longitudinal, as principais características linguísticas que os dados de hipossegmentação analisados mobilizaram. Para realizar essa tarefa, os dados foram organizados segundo regularidades prosódicas quanto ao tipo de combinação de palavra prosódica e clítico. Esse critério de classificação dos dados é baseado na hipótese de que as junções entre palavras na escrita observadas nos textos do EF II são ancoradas na configuração de constituintes prosódicos. Essa hipótese não implica afirmar que essas junções gráficas sejam exclusivamente ancoradas em constituintes prosódicos. Ao identificá-las, no entanto, busca-se descrever em que medida há regularidades linguísticas de natureza fonológica. Desse modo, foram cinco as regularidades observadas no cópulus deste trabalho, as quais classificamos de Tipo 1, Tipo 2, Tipo 3, Tipo 4 e Outros, assim denominadas em função da frequência de dados encontrada no cópulus.

De maneira sucinta, o Tipo 1 abarca a combinação de clítico com palavra prosódica (cl + ω), por exemplo: “a favor” > “afavor”. O tipo 2 abrange a combinação de clítico com clítico (cl + cl), por exemplo: “do que” > “doque”. O conjunto de dados do Tipo 3 diz respeito à

combinação de palavra prosódica com palavra prosódica ( $\omega + \omega$ ), por exemplo: “está bom” > “tabom”. Já o conjunto de Tipo 4 compreende a combinação de palavra prosódica com clítico ( $\omega + cl$ ), por exemplo: “pegá-la” > “pegala”. Por fim, a combinação que chamamos de Outros é formada pela junção de mais de uma palavra prosódica e/ou mais de dois clíticos, por exemplo: “filho do cavalo” > “filhodocavalo”. Trazemos, abaixo, uma tabela que contempla longitudinalmente os tipos de combinação:

Tabela 8. Tipos combinação de hipossegmentações

Tipo	Estrutura	Exemplo de hipo	Hipo/ano escolar				Total	
			6 <sup>a</sup>	7 <sup>a</sup>	8 <sup>a</sup>	9 <sup>a</sup>	n <sup>o</sup>	%
1	cl + $\omega$	“acabeça”	72	91	65	70	298	54,3%
2	cl + cl	“oque”	20	20	14	49	103	18,8%
3	$\omega + \omega$	“perai”	20	33	9	12	74	13,5%
4	$\omega + cl$	“tende”	27	19	12	14	72	13,1%
	Outros	euporezempro	2	1	0	0	3	0,5%
<b>Total</b>			141	164	100	145	550	100%

Nas seções (3.2.1) à (3.2.5), discutiremos detalhadamente cada tipo de combinação. De imediato, faz-se importante destacar que, se considerarmos o total bruto de ocorrências, só será possível verificar um decréscimo de ocorrências no conjunto de dados Outros. Nos outros tipos de combinação, observa-se aumento de ocorrências do oitavo ao nono ano, indicando que o grupo de alunos estudado apresenta dificuldades em delimitar palavras ortográficas quando envolvidas sequências de clítico e palavra, principalmente, e sequência de clíticos, em menor porcentagem dos dados.

No entanto, esse cenário muda ao considerarmos o coeficiente gerado a partir da razão entre número de hipossegmentações e palavras escritas em cada ano, como contemplado na tabela seguinte:

Tabela 9. Coeficiente dos dados de hipossegmentação

Tipo	Estrutura	Exemplo de hipo	Hipo/ano escolar				Total
			6 <sup>a</sup>	7 <sup>a</sup>	8 <sup>a</sup>	9 <sup>a</sup>	
1	cl + $\omega$	“acabeça”	0,14	0,13	0,15	0,08	0,50
2	cl + cl	“oque”	0,04	0,03	0,03	0,05	0,15
3	$\omega + \omega$	“perai”	0,04	0,05	0,02	0,01	0,12
4	$\omega + cl$	“tende”	0,05	0,03	0,03	0,02	0,12
	<b>Outros</b>	“euporezempro”	0,004	0,001	0,00	0,00	0,005
<b>Total</b>			0,28	0,24	0,24	0,16	0,91

Na Tabela 9, observamos que quase todos os tipos de combinação apresentam uma tendência decrescente. Essa tendência não foi verificada somente no Tipo 2. Desse modo, considerar a razão hipossegmentações/número de palavras também é importante para identificar tendências longitudinais de tipos de estruturas prosódicas hipossegmentadas. Demonstramos, assim, que essa abordagem metodológica melhor embasa a análise de dados, obtendo-se resultados quantitativos que mais bem descrevem os fatos linguísticos.

É importante salientar que é esperado não haver dados de segmentação não-convencional no EF II, haja vista o tempo de escolarização. De uma perspectiva que vê fala e escrita como modalidades linguísticas dicotômicas, dados de hipossegmentação revelariam uma indesejada ancoragem dos alunos na fala, uma vez que a escolarização não levou a maior domínio das convenções ortográficas, parte do conjunto de práticas letradas escolarizadas. Em outras palavras, nessa visão dicotomizada entre fala e escrita, mais tempo de escolarização levaria à extinção de uma suposta interferência da fala na escrita.

No entanto, de outra perspectiva da relação fala e escrita, interpretamos que a diminuição na quantidade de dados de hipossegmentação e o fato de essas ocorrências não se extinguirem ao final do EF II são evidências de que o trabalho do sujeito com a língua(gem) é constante, heterogêneo e se mostra, de modo privilegiado, nos momentos em que o aluno é chamado a escrever um texto escrito mesmo ao final do EF II. Desse modo, nossa hipótese é a de que os dados de hipossegmentação que iremos apresentar a seguir, evidenciarão (alguns mais, outros menos) a circulação dos alunos enquanto escreventes por diversas práticas orais/faladas e letradas/escritas.

Passemos, então, à descrição e análise dos tipos de estruturas prosódicas mobilizadas nos dados de hipossegmentação identificados.

### **3.2.1 Tipo 1: Clítico + palavra prosódica**

As hipossegmentações que se enquadram no Tipo 1 são resultantes da junção de um clítico com uma palavra prosódica. Essa junção foi a mais recorrente no EF II, totalizando 54,2% (298/550). A seguir apresentamos Tabela de dados desse tipo em cada ano letivo.

Tabela 10. Coeficiente de dados do tipo 1

<b>cl + ω</b>					
<b>Ano</b>	6 <sup>a</sup>	7 <sup>a</sup>	8 <sup>a</sup>	9 <sup>a</sup>	Total
<b>Hipo</b>	72	91	65	70	298
<b>Palavras</b>	50981	68499	42204	93267	254951
<b>Coeficiente</b>	0,14	0,13	0,15	0,08	0,50

A Tabela 10 mostra que as hipossegmentações mantiveram-se com taxa entre 0,13 e 0,15 do sexto ao oitavo ano letivo (do sexto para o sétimo ano houve decréscimo de 0,01 e que do sétimo para o oitavo ano houve aumento de 0,02). Já, no nono ano, houve decréscimo de 0,08, isto é, chegou à metade dos dados. Desse modo, constatamos que os alunos do EF II terminam essa etapa escolar com menos ocorrências do Tipo 1.

Nesse momento, faz-se importante descrever quais as ocorrências encontradas nesse conjunto de dados. Para essa tarefa, passamos a considerar os totais de dados por ano letivo. Chamou-nos a atenção o fato de haver, nesse conjunto de dados do Tipo 1 (correspondente a 67,8% das ocorrências – 202/298), maior número de grafias de combinação de um clítico terminado em vogal e uma palavra prosódica iniciada por consoante (cl vogal + ω consoante). O restante das ocorrências (32,2,9% - 96/298) desse conjunto dividiram-se em:

- combinação de clítico terminado em consoante e palavra prosódica iniciada também por consoante (cl consoante + ω consoante): 16,4% (48/298);
- combinação de clítico terminado em consoante e palavra prosódica iniciada por vogal (cl consoante + ω vogal): 9,4% (28/298);
- combinação de clítico terminado em vogal e palavra prosódica iniciada também por vogal (cl vogal + ω vogal): 6,7% (20/298).

A observação do tipo de segmento na fronteira entre palavras mostra que essa junção envolve menos sequências de palavras em que há possibilidade de aplicação de processos de reestruturação silábica, como o sândi vocálico entre “que” e “eu”, registrado em “queu”, ou a ressilabação do “r” de fim de sílaba em “por” que passa a início de sílaba no registro “porai”.

Na Tabela a seguir, trazemos essas informações sobre os segmentos em fronteira de palavras de clíticos e palavras prosódicas, organizadas longitudinalmente.

Tabela 11. Possibilidades de combinação de dados do tipo 1

Possibilidades de combinação	Ex.	6ano	7ano	8ano	9ano	Totais	%
<b>cl vogal + ω consoante</b>	“denoite”	47	70	37	48	202	67,8
<b>cl consoante + ω consoante</b>	“endia”	11	14	17	6	48	16,1
<b>cl consoante + ω vogal</b>	“porai”	6	4	8	10	28	9,4
<b>cl vogal + ω vogal</b>	“queu”	8	3	3	6	20	6,7
<b>Totais</b>		72	91	65	70	298	100

No sexto ano a combinação cl vogal + ω consoante ocorreu 48 vezes. Nesse conjunto, a maior recorrência foi a junção do clítico “de” com a palavra “novo” > “denovo” (quatro ocorrências) / “dinovo” (quatro ocorrências). A segunda junção mais recorrente foi do clítico “de” com a palavra “repente” > “derrepente” (sete ocorrências) e a terceira foi do clítico “a” com a palavra “noite” > “anoite” (duas ocorrências).

No ano seguinte, a combinação cl vogal + ω consoante aumentou para 70 ocorrências. Assim como no sexto ano, a maioria das ocorrências deram-se a partir da junção do clítico “de” com a palavra “novo” > “dinovo” (seis ocorrências) / “denovo” (nove ocorrências). Do mesmo modo, a segunda maior ocorrência foi a junção do clítico “de” com a palavra “repente” (12 ocorrências) e a terceira junção do clítico “a” com a palavra “noite” > “anoite” (quatro ocorrências).

No oitavo ano, a junção de cl vogal + ω consoante diminuiu para 37 ocorrências, sendo a maioria delas (12 ocorrências) decorrentes da junção do clítico “de” com a palavra “repente”. A junção do clítico “de” com a palavra “novo” ocorreu quatro vezes: “dinovo” (três ocorrências) / “denovo” (uma ocorrência) e o foi a segunda junção mais recorrente nesse ano. A hipossegmentação “anoite” foi a terceira, com três dados.

No último ano do EF II, houve 48 ocorrências da junção de cl vogal + ω consoante, sendo a maioria delas: “dinovo” (três ocorrências) / “denovo” (quatro ocorrências). A segunda junção mais recorrente foi novamente a hipossegmentação “derrepente” (seis ocorrências) e a terceira junção foi a hipossegmentação “anoite” (quatro ocorrências).

A partir desse resultado, notamos que as hipossegmentações decorrentes da junção de cl vogal + ω consoante, apesar de serem muitas, são praticamente sempre as mesmas, evidenciando que a maior dúvida desses alunos está em três sequências específicas: “de novo”, “de repente” e “a noite”.

Esse resultado também mostra que a preposição “de” e o artigo “a” parecem sofrer mais fortemente um processo de prosodização, já que são os clíticos mais hipossegmentadas em todo EF II, como podemos ver na tabela a seguir.

Tabela 12. Quantidade de hipossegmentações de "de" e "a"

	Ex.	6º ano	7º ano	8º ano	9º ano	Total	%
<b>Preposição "de"</b>	“derrepente”	20	36	20	14	90	30,1%
<b>Artigo "a"</b>	“anoite”	11	19	10	23	63	21,0%

Desse modo, a grafia dessas palavras parece ser a principal dúvida para os alunos do EF II e sugere a necessidade de análise aprofundada das estruturas morfossintáticas envolvidas nessas hipossegmentações, a fim de investigar possíveis motivações que podem ser de naturezas distintas, a saber: (i) quantitativa, na medida em que haveria mais preposições e artigos entre os monossílabos empregados nos textos escritos e, assim, as junções de preposições e artigos são apenas efeitos de frequência; (ii) qualitativa, na medida em que preposições e artigos são palavras gramaticais com funcionamento complexo (cf. ILARI, 2015), o que proporcionaria maior dificuldade no seu emprego por alunos no EF<sup>33</sup>.

Sobre os dados “denovo” / “dinovo” e “anoite”, observamos que houve a formação de possíveis palavras trissilábicas. Em “derrepente”, por sua vez, houve a formação de uma palavra polissilábica. Esses três dados equivalem a palavras paroxítonas, isto é, palavras que recebem acento fonológico na penúltima sílaba. Desse modo, esses dados permitem interpretar que há uma percepção, por parte do escrevente, de um padrão métrico da língua, já que o acento na penúltima sílaba é o mais frequente no PB<sup>34</sup>. Verifica-se, portanto, que as hipossegmentações podem ser interpretadas como registros escritos de processo de prosodização de clítico à direita em relação ao seu hospedeiro.

Esse resultado assemelha-se ao mostrado por Abaurre (1991) para escrita infantil. Os dados de segmentação não-convencional analisados por essa autora resultariam em palavras trissilábicas e dissilábicas paroxítonas. Abaurre (1991) argumenta que alguns dados de segmentação não-convencional observados em textos infantis parecem sinalizar para a forma canônica da palavra em português. De modo semelhante, Capristano (2007a) também encontra dados de segmentação não-convencional na escrita infantil que sinalizam que a escrita da criança pode ter sido atravessada por alguma percepção do que seria a forma canônica de palavras na língua.

<sup>33</sup> Pesquisa sobre grafia das preposições (seja hipo ou hipersegmentada) foi feita por (SILVA, 2018).

<sup>34</sup> Araújo et al. (2007, p. 55), baseado em 150.875 palavras registradas no Dicionário Houaiss de Língua Portuguesa, mostra que a frequência relativa de paroxítonas é de pouco mais de 50% entre dissílabos, chega a quase 60% entre os trissílabos e se mantém mais ou menos estável, perto de 70%, para palavras com quatro sílabas ou mais. Já, palavras oxítonas com duas sílabas apresentam uma frequência de um pouco menos de 50% e quanto maior a palavra, menos a chance de ser oxítona. As paroxítonas dissílabas, por sua vez, demonstram uma frequência de menos de 10%, enquanto que a frequência das polissílabas chega quase a 30%.

Em alguns dados desse subconjunto houve grafias que registraram processo de alçamento vocálico<sup>35</sup> que transformou [e, o] em [i, u]<sup>36</sup>. O registro desse processo foi atestado em todos os anos do EF II:

Quadro 1. Hipossegmentações com processo de alçamento

6º ano	7º ano	8º ano	9º ano
“de novo” > “dinovo” (quatro ocorrências) “de nada” > “dinada”	“de novo” > “dinovo” (seis ocorrências) “de nada” > “dinada” “me dar” > “midar”	“de novo” > “dinovo” (três ocorrências) “me beijou” > “mibejo”	“de novo” > “dinovo” (três ocorrências) “no lugar” > “nulugar” “que nem” > “quinem”

Nos dados em que houve o registro de alçamento, foi possível notar, de forma mais evidente, a oscilação do escrevente entre a representação dos sons de sua fala e a convenção ortográfica que a escola ensina. Nessa oscilação, localizamos um modo de emergência da heterogeneidade da escrita (CORRÊA, 1997, 2004): nas grafias das vogais, registram-se suas realizações fonéticas.

As hipossegmentações decorrentes da combinação cl consoante + ω consoante ocorreram 12 vezes no sexto ano, sendo a maioria delas hipossegmentações da sequência “por favor” > “porfavor” (duas ocorrências) e “com certeza” > “concerteza” (duas ocorrências).

No sétimo ano, a combinação cl consoante + ω consoante ocorreu 13 vezes. Nesse conjunto, a hipossegmentação “porcausa” (“por causa”) ocorreu três vezes e ocorreram dados como e “senquere” (“sem querer” – duas ocorrências) e “enquando” (“em quando” – duas ocorrências). No ano seguinte, esse tipo de combinação aumentou para 17 ocorrências. Nesse conjunto de dados, a hipossegmentação “concerteza” (“com certeza”) ocorreu seis vezes. No último ano do EF II, a combinação cl consoante + ω consoante caiu para sete ocorrências, dentre essas, a mais recorrente foi a hipossegmentação da sequência “com certeza” > “concerteza” (duas ocorrências).

Nesse conjunto de dados, os clíticos “com”, “sem” e “em” apresentam característica importante para esta análise. Quando esses clíticos se uniram a palavras prosódicas iniciadas

<sup>35</sup> O alçamento refere-se ao movimento da língua em direção à parte superior da cavidade bucal, no qual as vogais médias-altas [e] e [o] se realizam como vogais altas [i] e [u] (Cf. BISOL, 1981, 2010).

<sup>36</sup> Não é nosso intuito, neste trabalho, investigar quais são os contextos estruturais nos quais ocorrem esses alçamentos. Entretanto, os trabalhos de Tenani e Silveira (2008) e Carmo e Tenani (2013) mostram que o alçamento dessas vogais no dialeto do noroeste paulista é resultado da ação conjunta de fatores traço das consoantes adjacentes à vogal alvo, a contiguidade e tonicidade da vogal seguinte à sílaba pretônica, sujeita ao alçamento.

por consoante, houve a preferência pelo emprego da letra <n> por <m>, grafia que é guiada pela regra contextual de emprego dessas consoantes em final de sílaba seguida por consoante, como descrito no quadro a seguir. O emprego de <n> em fim de sílaba quando seguido por consoante revela a hipótese do escrevente de se tratar de grafia de sílaba dentro de uma palavra.

Quadro 2. Hipossegmentações emprego da letra <m> por <n>

6º ano	7º ano	8º ano	9º ano
“em dia” > “endia” “com certeza” > “concerteza” (duas ocorrências)	“sem querer” > “senquere” (duas ocorrências)	“com certeza” > “concerteza” (seis ocorrências) “em todo” > “intodo”	“com certeza” > “concerteza” (duas ocorrências) “em quanto” > “enquando”

Nas poucas ocorrências observadas em que essa troca de nasal não se verificou, foi possível observar características, como:

- i. a palavra prosódica iniciar pela letra <b> ou <p>: “embusca” (“em busca”) “sempessoa” (“sem pessoa” – uma ocorrência no 8º ano);
- ii. o encontro de duas consoantes iguais, no qual há o apagamento de uma delas – “ummonte” > u[m]onte (uma ocorrência ano 6º ano, uma ocorrência no 7º ano e uma ocorrência no 8º ano).

A troca da nasal e ausência dessa troca – verificadas em todos os anos do EF II – são informações que nos levam a observar a participação do escrevente em práticas letradas/escritas, já que essas grafias revelariam tentativas de o escrevente se “adequar” àquilo que ele imagina ser o código escrito institucionalizado, o que a caracteriza imagem que o escrevente faz da (sua) escrita (CORRÊA, 1997, 2004). Nossa interpretação é que o escrevente seguiu uma regra da convenção ortográfica que estabelece que a nasal em coda, dentro da palavra, é representada por “m” somente quando seguida por “p ou b”. Reflexão dessa natureza já havia sido feita por Tenani (2010) com base em dados de sexto ano do EF, que apontara relação entre escolha de letra segundo as convenções ortográficas e a hipótese de haver junção ou separação da sílaba em relação ao que se segue na escrita.

Ainda decorrente da combinação el consoante + ω consoante, encontramos a hipossegmentação “exanamorada” (“ex-namorada”), único dado em que houve uso não-convencional de hífen no conjunto de dados do tipo 1. Essa hipossegmentação ocorreu uma vez

no sétimo ano. Nesse dado, no lugar do sinal gráfico hífen, o aluno inseriu a vogal <a>, mantendo, assim, a estrutura CVC.

A combinação de cl consoante + ω vogal ocorreu, em todos os anos do EF II, com a preposição “por” somada a uma palavra iniciada por vogal. Essa combinação propiciou a ressilabação de sequências CVC#V de modo a resultar em sequência de duas sílabas canônicas do português: CV.CV. No sexto ano, a hipossegmentação “poriso” / “poriço” (“por isso”) ocorreu três vezes. No sétimo ano, esse número diminuiu para quatro ocorrências. Nesse conjunto, ocorreram os dados “porisso” (“por isso” - duas ocorrências) e “poraqui” (“por aqui” - uma ocorrência). No oitavo ano, houve um aumento para oito ocorrências, sendo que sete delas foram da hipossegmentação “porisso” (“por isso”). No nono ano, houve dez dados da combinação cl consoante + ω vogal e, do mesmo modo que nos anos anteriores, a hipossegmentação “porisso” foi a mais recorrente, com oito ocorrências.

Por fim, a combinação entre cl vogal + ω vogal ocorreu oito vezes no sexto ano, sendo as hipossegmentações “queu” (“que eu”) e “eai” (“e aí”) as mais comuns, com três ocorrências cada. No ano seguinte, verificamos somente três ocorrências decorrentes da combinação cl vogal + ω vogal, foram essas: “queu” (“que eu”), “tespero” (“te espero”) e “meachustei” (“me assustei”). No oitavo ano, esse número se manteve o mesmo, mas as hipossegmentações foram: “euadoro” (“eu adoro” – uma ocorrência) e “eai” (“e aí” – duas ocorrências).

A combinação cl vogal + ω vogal aumentou no nono ano, passando de três para seis ocorrências e, assim como no oitavo ano, a junção da sequência “e aí” foi a mais recorrente (três ocorrências); entretanto, esse dado ocorreu com três grafias distintas: “iaí”, “eae” e “eaí”. Podemos interpretar que as diferentes grafias da sequência “e aí” evidenciam a tentativa, por parte do escrevente, de registrar graficamente a materialidade fônica do oral/falado (CORRÊA, 1997, p. 185).

No sexto, sétimo e nono ano, quando a vogal final do clítico e a vogal inicial da palavra eram idênticas, houve registros de processo de sândi vocálico externo, o qual pode ser interpretado como degeminação<sup>37</sup> da vogal “e”, caso se considere as grafias das formas envolvidas, mas poderia ser considerada uma elisão se considerada a realização alçada da vogal átona final de “que” como [ki eu]

---

<sup>37</sup> A degeminação é um fenômeno fonológico decorrente da ressilabação de duas vogais iguais em fronteiras de vocábulos (Cf. BISOL, 1992).

Quadro 3. Hipossegmentações com processo de degeminação

6º ano	7º ano	9º ano
queu > q[e]u (três ocorrências)	queu > q[e]u tespero > t[e]spero	quele > qu[e]le

Em “queu” e “quele”, observamos o registro de processo de sândi externo entre “que” e “eu” e “que” e “ele”. Em ambos os casos, houve choque entre núcleos silábicos que propiciou a queda da vogal átona de “que”, resultando em [keu] e [keli]. Assim como em “tespero”, em que houve choque entre os núcleos silábicos de “te” e “espero”, gerando a sequência [tis'pɛro].

Interpretamos que o processo de sândi registrado graficamente evidencia a circulação do escrevente pela convenção ortográfica institucionalizada, uma vez que houve tentativas de normatização de sua escrita baseada naquilo que o escrevente imagina ser o código escrito institucionalizado. Em outras palavras, sequências da letra “e” não são comuns na ortografia do PB, isso faz com o aluno busque “adequar” a sua escrita para atender àquilo que ele imagina que a ser o código escrito institucionalizado.

Essa tentativa de normatização às convenções ortográficas não ocorreu quando as vogais eram diferentes, como exemplificado no quadro a seguir:

Quadro 4. Hipossegmentações sem processo de degeminação

6º ano	7º ano	8º ano	9º ano
"e aí" > “eai” (três ocorrências) "me abandone" > “meabandone”	"me assustei" > “meachustei”	"e aí" > “eai” (duas ocorrências) "eu adoro" > “euadoro”	"e aí" > “eai” (três ocorrências)

De modo geral, os dados de tipo 1 indiciam que há um possível não-reconhecimento do estatuto autônomo dos clíticos, que seriam interpretados, pelos escreventes, como sílabas de palavra de conteúdo às quais se juntam graficamente e prosodicamente. Esse não-reconhecimento ocorreu em todos os anos do EF II.

O conjunto de dados do tipo 1, apresenta as mesmas características em todos os anos, evidenciando a manutenção do trânsito dos escreventes por elementos de natureza fonético-fonológico e pelo código escrito institucionalizado.

A seguir, descrevemos as características da junção entre dois clíticos, segundo tipo de junção mais recorrente no EF II.

### 3.2.2 Tipo 2: Clítico + clítico

As combinações de Tipo 2 são resultantes da junção de dois clíticos. Houve 103 ocorrências (18,8%) desse tipo. Na tabela a seguir, apresentamos as ocorrências ao longo dos quatro anos do EF

Tabela 13. Coeficiente de dados do tipo 2

<b>cl + cl</b>					
<b>Ano</b>	6 <sup>a</sup>	7 <sup>a</sup>	8 <sup>a</sup>	9 <sup>a</sup>	Total
<b>Hipo</b>	20	20	14	49	103
<b>Palavras</b>	50981	68499	42204	93267	254951
<b>Coeficiente</b>	0,04	0,03	0,03	0,05	0,15

O coeficiente de dados do Tipo 2 decaiu 0,01 no sétimo e no oitavo ano em relação ao sexto ano; todavia, no nono ano, houve um aumento de 0,02. Vale observar que, apesar da grande quantidade de ocorrências, esse conjunto apresenta pequena variedade de dados, pois a junção mais frequente, com 85 ocorrências (82,5%), é do artigo “o” com a conjunção “que”, recorrente em todos os anos letivos, como se pode atestar na próxima Tabela.

Tabela 14. Número de ocorrência da hipossegmentação *oque*

<b>Hipossegmentação "o que"</b>						
<b>Ano</b>	6 <sup>o</sup>	7 <sup>o</sup>	8 <sup>o</sup>	9 <sup>o</sup>	Total	%
<b>N<sup>o</sup></b>	16	13	10	46	85	82,5%

Como podemos ver na Tabela 14, no sexto ano, ocorreram 16 hipossegmentações da sequência “o que”. No sétimo ano, esse número caiu para 13 ocorrências, sendo que duas delas foram grafadas como “uque” e uma como “oq”. No ano seguinte, houve diminuição para dez ocorrências. E, no último ano do EF II, esse número quadruplicou para 46 ocorrências, sendo uma delas grafadas como “oq”.

Vale mencionar que a grafia “oq” é indício do entrelaçamento da escrita com práticas sociais orais e letradas, já que essa estrutura é comumente usada por crianças e adolescentes em mensagens via internet, uma prática letrada digital, e representa [´uki], uma realização fonética de “o que”, advinda de práticas orais. Já, na grafia “uque”, por sua vez, nota-se um registro do processo de alçamento vocálico que transforma a vogal média-alta <o> na vogal alta <u>, evidenciando a tentativa do escrevente de registrar, no gráfico, a materialidade fônico-pragmática do oral/falado (CORRÊA, 1997, p. 185).

Essa grande frequência de hipossegmentações da estrutura “o que” que aumenta no final do EF II, por ser diferente do esperado, levou-nos a ampliar o olhar para flutuações entre grafia convencional “o que” e a grafia não-convencional “oque” em um mesmo texto, investigando alguma motivação linguística nessa flutuação. A fim de verificar qual ou quais seriam essas possíveis motivações, selecionamos dois textos em que ocorreram dados de flutuação de “o que”. Essa análise será apresentada na seção 3.4, quando consideraremos as hipossegmentações e flutuação com a grafia convencional no texto em que ocorreram.

No conjunto de dados de Tipo 2, também ocorreu, mas em menor número, a junção da preposição “do” e da conjunção “que” > “doque”. Esse dado ocorreu em todos os anos do EF II: uma vez no sexto ano, duas vezes no sétimo, três vezes no oitavo e uma vez nono.

Baseada em Tenani (2011b), interpretamos que as hipossegmentações “oque” e “doque”, foram motivadas pela conjunção “que”, que sozinha não apresenta acento lexical, mas que, se unida ao artigo “o” ou à preposição “do”, pode receber acento prosódico e passar a ser o hospedeiro dos clíticos “o” e “de”.

Além desses dados, ocorreram também junções da conjunção “e” com os artigos “o” ou “a” em três dos quatro anos letivos, como se verifica no quadro a seguir:

Quadro 5. Hipossegmentações das sequências "e o" e "e a"

6º	7º	9º
“e o” > “eo” (duas ocorrências)	“e a” > “ea” (cinco ocorrências)	“e o” > “eo” “e a” > “ea”

Esse tipo de ocorrência também foi encontrado por Cunha e Miranda (2007) em textos do EF I. Segundo essas autoras, a criança mostra resistência à presença de duas letras grafadas isoladamente no início de sentenças, pois esse tipo de hipossegmentação só ocorreu em início de sentenças no material analisado pelas autoras. Nos textos do EF II, por outro lado, esses dados ocorreram, em sua maioria, entre duas sentenças coordenadas, conforme os exemplos (3.1) e (3.2):

3.1. “[...] moro com a minha mãe ‘eo’ meu padrasto [...]” (6º ano)

3.2. “[...] além do fumo causar doença no pulmão e bronquite ‘ea’ pessoa fica com os dentes amarelos [...]” (9º ano)

O único caso em que esse tipo de junção não se deu em início de sentença foi em (3.3):

### 3.3. “Eu ‘ea’ Nathália brigamos [...]” (7º ano)

No que concerne às ocorrências apresentadas acima, nossa interpretação é a de que foram motivadas pelo fato de a conjunção “e”, um monossílabo átono, receber acento prosódico em (3.1) e (3.2) por ocupar a posição de início de frase entoacional, formando um ditongo ao ser grafado junto aos artigos “o” e “a”, resultando uma sílaba na qual seria possível notar a relação forte/fraco entre os elementos vocálicos.

Tendo concluída a análise do conjunto de dados do Tipo 2, consideramos que as junções desse tipo, assim como as de Tipo 1, são resultantes do não-reconhecimento do estatuto autônomo dos clíticos. Dessa forma, nesse conjunto, as hipossegmentações seriam resultados da interpretação, pelos escreventes, como uma única palavra prosódica.

Podemos argumentar que, de modo geral, essas ocorrências são motivadas simultaneamente por: (i) informações de natureza prosódica, uma vez que mobilizam componentes átonos da língua e sua prosodização em relação a seu hospedeiro; (ii) informações letradas/escritas e orais/faladas, já que algumas ocorrências evidenciaram, de forma mais marcada, o trânsito do escrevente por práticas letradas/escritas e orais/faladas.

A seguir, descreveremos o Tipo 3 de hipossegmentações, que compreende as características das junções entre duas palavras prosódicas.

#### 3.2.3 Tipo 3: Palavra prosódica + palavra prosódica

O conjunto de dados do Tipo 3 diz respeito à junção de duas palavras prosódicas. Em termos gerais, somente houve um decréscimo nos dois últimos anos: no oitavo ano, o coeficiente era de 0,02 e decaiu para 0,01 no nono ano, como se confirma na tabela a seguir.

Tabela 15. Coeficiente de dados do tipo 3

Ano	ω + ω				Total
	6 <sup>a</sup>	7 <sup>a</sup>	8 <sup>a</sup>	9 <sup>a</sup>	
<b>Hipo</b>	20	33	9	12	74
<b>Palavras</b>	50981	68499	42204	93267	254951
<b>Coeficiente</b>	0,04	0,05	0,02	0,01	0,12

No que diz respeito aos tipos de ocorrências, nos dois primeiros anos do EF II, ocorreu um maior número de hipossegmentações que envolveram verbos. Das 20 ocorrências

identificadas no sexto ano, 11 foram decorrentes da junção de um verbo e uma palavra prosódica, que poderia ser outro verbo (duas ocorrências), um adjetivo (seis ocorrências), um advérbio (duas ocorrências) ou um substantivo (uma ocorrência). Abaixo, transcrevemos exemplos:

- 3.4. “vai ser” > “vaiser” (uma ocorrência)
- 3.5. “está bom” > “tabom” (cinco ocorrências)
- 3.6. “espera aí” > “perai” (uma ocorrência)
- 3.7. “tem fim” > “tenfim” (uma ocorrência)

No ano seguinte, as ocorrências envolvendo verbos aumentaram, assim, dos 33 dados encontrados nesse ano, 26 foram decorrentes da junção de verbo com outro verbo (cinco ocorrências), de um verbo com um advérbio (cinco ocorrências) e de um verbo com um adjetivo (16 ocorrências). Os exemplos abaixo demonstram algumas delas:

- 3.8. “fui lá” > “fuila” (cinco ocorrências)
- 3.9. “está tirando” > “tatirando” (cinco ocorrências)
- 3.10. “está bom” > “tabom” (16 ocorrências)

Nesse subconjunto de dados, em algumas hipossegmentações ocorreu a redução do verbo “estar” que, após perder a sílaba pretônica, é grafado como “ta” (sem acento gráfico) ou “to”. Esse tipo de redução ocorreu em quatro ocorrências no sexto ano e em 19 ocorrências no sétimo ano, abaixo exemplificadas:

Quadro 6. Hipossegmentações com redução do verbo "estar"

6º ano	7º ano
“está bem” > “tabem” “está bom” > “tabom” (três ocorrências)	“está bom” > “tabom” (16 ocorrências) “estou dando” > “todando” “estou indo” > “toindo” “está tirando” > “tatirano”

Em “tabem”, “tabom”, “todando” e “toindo”, há choques de acentos entre o verbo “está” e as palavras “bem”, “bom”, “dando” e “indo”. Sobre isso, Matzenauer (1996) explica que as línguas têm tendência a eliminar choque de acentos através de uma reorganização da grade métrica. Tenani (2002) discute três possibilidades de resolução de choques de acentos com base

em dados do Português falado na região do noroeste paulista. A autora conclui que, além da retração de choque de acento mais à direita, também é possível haver aumento da duração das sílabas em choque acentual e diferença de altura tonal das sílabas acentuadas ou, ainda, haver desacentuação. Nesse caso, o choque acentual parece ter sido solucionado com a transformação da forma verbal “está” em clítico, portanto, a redução e a desacentuação de formas flexionadas. Vale ressaltar que essa possibilidade de cliticização de “está” já foi observada por Bisol (2005) com base em dados de fala. Tais grafias registram essa possível cliticização da forma “está”.

Além da cliticização de “está”, outras motivações podem estar presentes, como a relação com gênero textual solicitado. Nas produções em que os dados ocorrem, os gêneros são contos, relatos de experiências e cartas pessoais. Esses gêneros propiciam que os alunos criem diálogos entre personagens, possibilitando registros gráficos mais característicos de enunciados falados informais em interação face a face, por exemplo.

Interpretamos que esses dados são exemplos da circulação do escrevente por práticas orais, nos quais, de acordo com Corrêa (1997, 2004), há a “tentativa, por parte do escrevente, de representação do planejamento conversacional e do jogo argumentativo prosodicamente marcado ou, finalmente, sua tentativa de registrar graficamente a materialidade fônico-pragmática do oral/falado” (CORRÊA, 1997, p. 185).

É possível observar também que em “tadando”, “toindo”, “tatirano”, a forma reduzida do verbo “estar” nas perífrases de gerúndio é característica de uma mudança em curso. Segundo Beline (1999), a forma “tá” pode tornar-se um clítico em relação ao gerúndio e pode vir a ser reinterpretado como “forma presa” nessas estruturas morfossintáticas. De um ponto de vista diacrônico, segundo Hopper e Traugott (1993), há uma tendência de formas perifrásticas virem a substituir formas simples. Em “tatirano”, a redução “ta”, forma átona para “está”, ocorre independentemente de haver choque acentual; assim, nesse caso, a forma reduzida “ta” parece ser cliticizada por motivos morfossintáticos e a hipossegmentação é o registro dessa mudança em curso no português.

As hipossegmentações mostradas ocorreram somente nos dois primeiros anos do EF II. No oitavo ano houve somente uma hipossegmentação envolvendo um verbo: “tinha vez” > “tinaves”. Nesse dado, ocorreu uma simplificação gráfica do dígrafo “nh”, o qual foi substituído pela letra “n”, que apresenta um valor sonoro nasal semelhante ao do dígrafo. Poderíamos supor que, nesse dado, houve uma manutenção do mesmo padrão CV da sílaba, alterando a representação ortográfica da consoante nasal, o que pode indicar a circulação do escrevente por práticas letradas/escritas e orais faladas, uma vez que ele parece demonstrar um certo domínio do sistema alfabético e do valor sonoro dos grafemas.

No nono ano houve três ocorrências envolvendo verbos, a seguir transcritas:

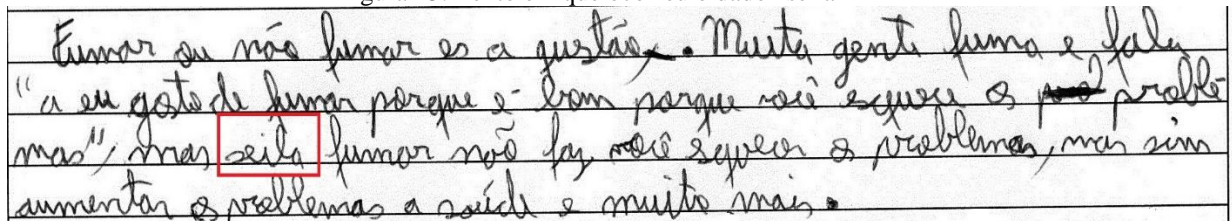
- 3.11. “sei lá” > “seila”
- 3.12. “importa para” > “enportapara”
- 3.13. “fui para” > “fuipra”

No dado em (3.11), observamos a formação de uma possível frase fonológica que evidencia a circulação do escrevente por práticas orais, já que “sei lá” poderia ser considerado uma expressão utilizada em práticas orais informais, para organizar a linearidade do discurso, como aponta Araújo (2014, p. 348):

[...] o *lá* modalizador se discursiviza, quando acompanhado do verbo saber no presente do indicativo, na terceira pessoa do singular. Além do seu valor modalizador original, que denota insegurança ou incerteza do falante em relação ao que fala, o *sei lá*, em determinados contextos, exerce a função de reestruturar o discurso, quando sua linearidade é rapidamente interrompida, ou de preenchedor de pausa. Nesses contextos, o *sei lá* é utilizado para viabilizar o processamento da fala, pois seu uso está mais para organizar a linearidade do discurso do que da estrutura significativa da cláusula.

No excerto em que ocorreu esse dado, abaixo copiado, podemos ver o funcionamento descrito por Araújo (2014).

Figura 15. Texto em que ocorreu o dado “seila”



Fonte: Banco de Dados de Escrita do EFII (Texto: Z11\_8E\_04M\_06)<sup>38</sup>

Nessa ocorrência, observamos que a hipossegmentação “seila” assume uma função pragmática de marcador discursivo, pois indica uma hesitação e é usada pelo escrevente para argumentar e sinalizar seu ponto de vista. Nota-se que o contexto em que ela ocorre é semelhante a contextos de fala nos quais o falante faz uso dessa expressão para não “perder o

<sup>38</sup> Possível leitura: Fumar ou não fumar es a questão. Muita gente fuma e fala “a eu gosto de fumar porque é bom você esquece os ~~pro~~ proble-mas”, mas seila fumar não faz você esquecer os problemas, mas sim aumentar os problemas a saúde e muito mais.

turno da fala e ganhar tempo, enquanto procura as palavras adequadas para expressar as suas ideias” (ARAÚJO, 2014, p. 348).

Os dados em (3.12) e (3.13) também envolvem verbos; entretanto, a configuração de ambos é distinta das anteriores, uma vez que os verbos “importa” e “fui” se juntam à preposição “para”. Essa junção não se mostrou comum no cópulus analisado, pois todos os demais dados que envolvem a preposição “para” (que serão descritos em breve) se unem à palavra que está mais à direita. Essa direção de cliticização não foi verificada nas ocorrências em (3.12) e (3.13). Nessas ocorrências, a palavra “para” se une à palavra que está a sua esquerda.

Devido às suas características únicas no conjunto de dados do Tipo 3, trazemos, a seguir, os textos nos quais os dados “enportapara” e “fuipra” ocorreram, a fim de verificarmos possíveis motivações para essas grafias.

Na proposta de produção textual em que “enportapara” ocorreu, era solicitada a produção de um artigo de opinião, no qual o aluno teria que argumentar a favor ou contra a internacionalização da Amazônia<sup>39</sup>. A seguir, apresentamos o texto a ser considerado na análise da hipossegmentação:

---

<sup>39</sup> Em anexo, a proposta é transcrita na íntegra.

Figura 16. Texto em que ocorreu o dado “enportapara”

Amazônias

Eu acho que isso é injusto os outros países vêm para o Brasil simplesmente para se apoderar de nossos pertences para ser levados para os outros países para fazer descobertas como remédios etc. Mais também isso inclui o fato de eles enportapara seu país sem pagar nada sobre aquilo que pegou os maiores prejudicados são os brasileiros por que esses produtos são levados da qui ~~de~~ para (graça) (e volta com) lá e depois volta para o Brasil caríssimos, eu acho que o Brasil tem condições de Bancar isso só que nos Brasileiros não se despoem a fazer. Também a o fato de que o Brasil não se enporta com a natureza e já começou a repartir a Amazônia, vendeu para um senhor de outro país, mais eu acho que a Amazônia deveria ser só dos Brasileiros. E os brasileiros ~~se acordassem~~ acordassem para poder over mais descoberta não fora do Brasil mais sim nele.

Fonte: Banco de Dados de Escrita do EFII (Texto: Z11\_8E\_07M\_05)<sup>40</sup>

Como podemos observar, a palavra “enporta” aparece duas vezes nesse texto: na primeira ocorrência (3.12.1), está hipossegmentada junto à palavra “para”, na segunda ocorrência (3.12.2), as fronteiras de palavra seguem a convenção, mesmo havendo um clítico que lhe antecede (“se”) e uma preposição que lhe segue (“the”). É possível notar que cada ocorrência tem significados diferentes.

<sup>40</sup> Possível leitura: Eu acho que isso é injusto os outros países vim para o Brasil simplesmente para se apoderar de nossos pertences para ser levados para os outros países para fazer descobertas como remédios etc. Mais também isso inclui o fato de que eles enportapara seu país sem pagar nada sobre aquilo que pegou os maiores prejudicados são os brasileiros por que esses produtos são levados da qui de para (graça) (e volta com) lá e depois volta para o Brasil caríssimos, eu acho que o Brasil tem condições de Bancar isso só que nos Brasileiros não se despoem a fazer. Também a o fato de que o Brasil não se enporta com a natureza e já começou a repartir a Amazônia, vendeu para um senhor de outro país, mais eu acho que a Amazônia deveria ser só dos Brasileiro. E os brasileiros se acordassem para poder over mais descoberta não fora do Brasil mais sim nele.

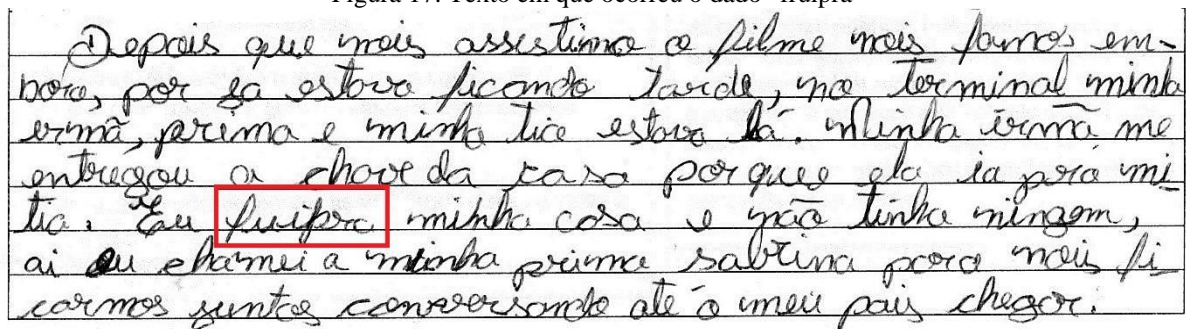
3.12.1 “Mais também isso inclui o fato de \que/ eles ‘enportapara’ seu país sem pagar nada sobre aquilo que pegou [...]”

3.12.2 “Também a o fato de que o Brasil não se ‘enporta’ com a natureas e já começou a repartir a Amazônia, [...]”

Em (3.12.1), “importa” é forma de verbo transitivo com significado de “mandar trazer de fora do país”; no caso, “importar produtos da Amazônia para seu país”. Em (3.12.2), “importa” é forma flexionada de verbo pronominal com o significado de “dar importância a”; no caso, “o Brasil não **se importa** com a natureza”. Observamos que a flutuação gráfica quanto aos limites de palavra está articulada a uma diferença de significado das formas verbais.

A seguir, trazemos trecho do texto no qual o dado “fuipra” ocorreu:

Figura 17. Texto em que ocorreu o dado “fuipra”



Fonte: Banco de Dados de Escrita do EFII (Texto: Z11\_8C\_20F\_02)<sup>41</sup>

Nesse trecho, observamos a hipossegmentação “fuipra” (3.13.1), constituída de “fui”, forma flexionada do verbo “ir”, e “pra”, forma reduzida da preposição “para”. Em trecho anterior, está grafada convencionalmente a sequência “ia pra”, também constituída de forma do verbo “ir” seguida de forma reduzida da preposição “para” (3.13.2).

3.13.1 “Eu ‘fuipra’ minha casa e não tinha ningem, [...]”

3.13.2 “Minha irmã me entregou a chave da casa porque ela ia pra mi\_tia”

Nos dois casos, há a mesma estrutura de sujeito seguida do verbo “ir” flexionado (em diferentes tempos e modos verbais) e forma reduzida da preposição “para”, seguida de “minha”,

<sup>41</sup> Possível leitura: Depois que nois assistimo o filme nois famos em-bora, por ja estava ficando tarde, no terminal minha irmã, prima e minha tia estava lá. Minha irmã me entregou a chave da casa porque ela ia pra mi\_tia. Eu fuipra minha casa e não tinha ningem, ai eu chamei a minha prima sabrina pra nois fi\_carmos juntas conversando até o meu pais chegar.

potencial hospedeiro prosódico, à direita da preposição “para”. A hipossegmentação “fuipra” evidencia a relação morfossintática entre verbo e preposição, grafada em sua forma reduzida, que possivelmente átona. Além dessa relação, uma configuração rítmica também se mostra relevante, pois, em (3.13.1) e (3.13.2), há sequências de pés troqueus binários, mas apenas a segunda ocorrência é hipossegmentada.

Em suma, a ausência de junção em (3.12.2) e (3.13.2) em comparação ao observado em (3.12.1) e (3.13.1) revela característica oscilante do processo de escrita quanto à representação gráfica dos limites de palavra que derivam das possibilidades da língua e evidenciam o trabalho do escrevente com a língua(gem).

Outro subconjunto de hipossegmentações encontrado em todos os anos do EF II compreende registros da forma reduzida da preposição “para” seguida de um hospedeiro. Quando há redução da preposição nos enunciados falados informais, ocorrendo “pra” ou “pa”, as formas podem se comportar como clíticos. Nas hipossegmentações, fica evidenciado o fato de “pra”, “pa” se unirem às palavras que os seguem, mostrando um comportamento de uma sílaba pretônica.

Em estudo sobre dados de fala de informantes da variedade do português do noroeste paulista, mesma variedade falada pelos escreventes sujeitos desta pesquisa, Marcato (2013) identificou a redução “para” > “pra” > “pa” e interpretou esse processo como indício de prosodização da preposição “para”. Essa autora mostrou que há 75% (1428/1982) de ocorrências da forma “pra” e de 25,3% (502/1982) de “pa” nos dados de fala. A forma “para”, no corpus analisado pela autora, alcançou somente 2,6% (52/1982) dos dados.

Nos dados de hipossegmentação deste trabalho, houve a grafia da forma “pra” em oito dados e somente em dois dados ocorreu a grafia “pa”. Já a forma “para” ocorreu em somente um dado, como observado no quadro que se segue:

Quadro 7. Hipossegmentações da sequência "para mim"

6º ano	7º ano	8º ano	9º ano
“para ninguém” > “paninguem” “para mim” > “pramim”	“para conversar” > “paraconversar” “para fora” > “prafora” “para mim” > “pramim”	“para frente” > “pafente” “para mim” > “pramin” “pramim” “para lá” > “pralá”	“para mim” > “prami” “pramim”

Do quadro acima, observamos uma informação interessante: em todos os anos do EF II ocorreu a hipossegmentação da sequência “pra mim”. Interpretamos essa recorrência como decorrente da resposta do escrevente às propostas textuais, pois esses dados ocorreram em

textos do tipo relato de experiência, no quais os alunos eram convidados a dar a sua opinião sobre determinado assunto.

Identificamos ainda processo semelhante quanto à redução descrita acima em mais três dados, identificados a seguir:

Quadro 8. Hipossegmentações com redução de palavra

6º ano	7º ano	8º ano	9º ano
“você criou” > “vocriou” “espera aí” > “perai”	“todo mundo” > “tomundo”	“tudo bem” > “tubem”	“todo dia” > “todia”

Nos dados do quadro acima, há indícios de que a palavra prosódica mais à esquerda perdeu conteúdo fonológico e passou a se comportar como um clítico. Os dados resultantes desse processo podem ser interpretados como possíveis palavras prosódicas.

Esse processo ocorreu, ao menos uma vez, em todos os anos do EF II. No sexto ano, em “vocriou”, notamos que a sílaba tônica “cê” é apagada. A sílaba átona se comporta como resto da forma de tratamento “você” e se une à palavra “criou”, seu hospedeiro. No dado “perai”, ocorreu registro do apagamento da sílaba pretônica “es” da forma verbal “espera” seguido do registro da degeminação das vogais “a”, entre “espera” e “aí”.

Nos dados do sétimo, oitavo e nono ano, as sílabas átonas das palavras mais à esquerda não são grafadas e as sílabas tônicas se mantêm. Entretanto, interpretamos que essas sílabas tônicas passam a ter proeminência enfraquecida em relação à tonicidade da sílaba da palavra prosódica à direita que também é elemento proeminente da frase fonológica. Nessa interpretação, as sílabas que restam da palavra prosódica mais à esquerda se comportam como clíticos.

Ademais, em “tubem” e “todia”, pode ter ocorrido haplogogia, um processo fonológico em que há a queda de uma sílaba em consequência do encontro de duas sílabas semelhantes em fronteiras de palavras (PAVEZI, 2005).

No conjunto de dados do tipo 3, alguns monossílabos acentuados também parecem evidenciar processo de cliticização. Esse processo ocorreu no sexto, sétimo e oitavo ano, como apresentado no quadro a seguir.

Quadro 9. Hipossegmentações com choque de acento

6º ano	7º ano	8º ano
“nós todos” > “nostodos”	“já já” > “jajá”	“já tinha” > “jatinha”

Nessas ocorrências, os acentos gráficos dos monossílabos não são registrados. A ausência desse sinal gráfico dá indícios de que os monossílabos podem ter passado por processo de cliticização, tendo seu acento enfraquecido, uma das possíveis soluções para resolução de choque acentual.

Por fim, no conjunto de ocorrências do Tipo 3, encontramos, no sexto e sétimo ano, hipossegmentações de estruturas perifrásticas com o verbo “ir” e um verbo infinitivo:

Quadro 10. Hipossegmentações de estruturas perifrásticas

6º ano	7º ano
“vai ser” > “vaiser”	“vou dormir” > “voudormir” “fui brincar” > “fuibrincar”

Sobre esse tipo de estrutura, Neves (2000) pontua que as perífrases de “ir” + infinitivo são perfeitamente gramaticais na língua e constituem um recurso muito produtivo para a expressão da futuridade. Ilari (2001) também explica que as perífrases constituídas com o auxiliar “ir” + infinitivo assemelham-se a uma forma verbal simples acrescida de morfemas, já que a referência temporal localiza-se em relação ao momento da fala. O autor ainda afirma que é provável que seja essa semelhança funcional com a forma simples de futuro que tem contribuído para que, atualmente, registre-se uma preferência por se expressar o tempo futuro por meio de expressões perifrásticas. Hopper e Traugott (1993) afirmam que, diacronicamente, há uma tendência de as formas perifrásticas virem a substituir as formas simples.

Os resultados apresentados para o conjunto de dados do Tipo 3 permitem afirmar que esses dados apontam para o possível reconhecimento de palavras como clíticos. Esse reconhecimento pode estar relacionado à representação que o escrevente faz de vocábulos relacionados a práticas mais informais de oralidade/fala e de letramento/escrita.

Passamos, a seguir, à combinação de Tipo 4, referente à junção de palavra prosódica e clítico.

#### 3.2.4 Tipo 4: Palavra prosódica + clítico

Passemos à descrição do conjunto de dados do Tipo 4, que diz respeito à junção de uma palavra prosódica com clítico. Esse tipo de juntura teve decréscimo de ocorrências ao longo do

EF II, como observado na Tabela 16. No sexto ano, houve 0,05 ocorrências, no sétimo e oitavo ano, 0,03 ocorrências cada e no nono ano, somente 0,01 ocorrência.

Tabela 16. Coeficiente de dados do tipo 4

Ano	ω + cl				Total
	6 <sup>a</sup>	7 <sup>a</sup>	8 <sup>a</sup>	9 <sup>a</sup>	
Hipo	27	19	12	14	72
Palavras	50981	68499	42204	93267	254951
Coeficiente	0,05	0,03	0,03	0,01	0,12

A tendência nesse conjunto de dados é a hipossegmentação de um verbo e um pronome em posição enclítica. No sexto ano, ocorreram 21 hipossegmentações desse tipo de sequência, no sétimo ano, foram somente sete, no oitavo ano, oito e no nono ano, foram 12 hipossegmentações, como descrito no quadro a seguir.

Quadro 11. Hipossegmentações de verbo + pronome

6º ano	7º ano	8º ano	9º ano
“achá-los” > “achalos”	“lembrá-lo” >	“vendê-lo” >	“preenchê-las” >
“ajudá-los” > “ajudalos”	“lembra-lo”	“vendelo	“preenchelas”
“despistá-lo” > “depistalo”	“ajudá-lo” >	“ajudá-las” >	“ajudá-lo” > “ajudalo”
“derrotá-los” > “derrotalos”	“ajudalo”	ajudalas	“fazê-lo” > “fazêlo”
“descrevê-lo” >	“pagá-la” >	“compra-lo” >	“foi-se” “foise”
“descrevelo	“pagala”	“compralo”	“levá-lo” > “levalo”
“estudá-los” > “estudalos”	“acalmá-la” >	“defende-los” >	“matá-lo” > “matalo”
“imitá-los” > “imitalos”	“acalmala”	“defenlos”	“passa-las” > “pasalas”
“pegá-lo” > “pedalo”	“cumpri-las” >	“seguí-lo” >	“perde-lo” > “perdelo”
“matá-lo” > “matalo”	“cumprilas”	“seguilo”	“seguí-lo” > “seguilo”
(duas ocorrências)	“tê-la” > tela	“tratá-la” > “tratala”	“sustenta-las” >
“vê-los” > “velos”	“estava-me” >	(duas ocorrências)	“sustentalas”
(duas ocorrências)	“tavame”	“pedi-la” > “pedila”	“tê-las” > “telas”
“buscá-la” > “buscala”			“testá-lo” > “testalo”
“levá-la” > “levala”			
“socorrê-la” > “socorrela”			
“pegá-la” > “pegala”			
(duas ocorrências)			
“casarem-se” > “casareense”			
“ajude-me” > “ajudime”			
(três ocorrências)			

Esse tipo de junção parece menos motivada na configuração prosódica preferencial do PB, pois a direção da prosodização de pronomes em relação ao verbo no PB é proclítica. Desse modo, essa estrutura pode ser considerada como marcada quando se levam em conta enunciados falados do PB (cf. ABAURRE e GALVES, 1996). Por outro lado, essa colocação pronominal é ensinada e, sobretudo, valorizada em ambiente escolar e em registros formais. A nossa

interpretação é a de que essa seja a razão da grande quantidade hipossegmentações de verbo + pronome no EF II, principalmente no sexto ano, que seria o ano escolar em que esse conteúdo programático relativo à colocação enclítica do verbo é ensinada<sup>42</sup>

Retomando a descrição dos dados do Quadro 11, observamos que a maioria das junções de verbo e pronome (15 ocorrências no sexto ano, cinco no sétimo ano, oito no oitavo ano e dez no nono ano) são de verbos com duas ou três sílabas, como em “pagala” e “estudalos”, resultando em palavras paroxítonas com pés ternários, uma vez que essa junção resulta em uma configuração métrica de troqueu silábico. Segundo Bisol (1996), essa configuração métrica é a mais geral do português quanto à atribuição de acento.

Já a junção de um verbo monossilábico a um clítico, como “velos” e “tela”, ocorreu bem menos (duas ocorrências no nono ano, uma no sétimo ano, e duas no nono ano) e resultou em pés binários, também trocaicos. Identificamos, ainda, a formação de possíveis palavras proparoxítonas, como “tavame”, (duas ocorrências no sexto ano e uma no sétimo ano). Nesse caso, o verbo, ao juntar-se com o pronome, forma um pé ternário (BISOL, 1996).

Cenário semelhante a esse foi encontrado nos dados de escrita infantil analisados por Cunha e Miranda (2007). Segundo elas, o uso da ênclise não é o mais característico do falar brasileiro e, por essa razão, a criança tende a representar essa estrutura, na escrita, como uma só palavra, conforme mostrado nos exemplos. Cunha e Miranda (2007) explicam que, ao ter de decidir entre as duas situações não tão comuns na língua, a ênclise ou o acento proparoxítono, o aluno opta pela ênclise, grafia que gera uma configuração métrica pouco frequente na variedade brasileira do português.

Ainda no subconjunto de dados do Tipo 4, encontramos, no sexto, sétimo e nono ano, um total de sete dados decorrentes da junção de um verbo com uma preposição ou um artigo, como exemplificado no quadro a seguir.

Quadro 12. Hipossegmentações de verbo + preposição ou artigo

6º ano	7º ano	9º ano
“desci da” > “desida” “bateu a” > “bateua” “achei o” > “acheio” “amei o” > “ameio”	“tem que” > “tenque” “vi o” > “vio”	“tem de” > “tende”

<sup>42</sup> Segundo os PCN (Parâmetros Curriculares Nacionais de Língua Portuguesa, 1998), no EF II, deve ser trabalhado o “sistema pronominal (diferentes quadros pronominais em função do gênero): preenchimento da posição de sujeito, extensão do emprego dos pronomes tônicos na posição de objeto, desaparecimento dos clíticos, emprego dos reflexivos etc.,” (PCN, 1998, p. 60).

A partir do Quadro 12, observamos que no sexto ano, predomina a junção de verbo (todos no pretérito perfeito) e artigo. Os verbos que a princípio são constituídos de um pé binário iâmbico, após a junção, se transformam em pés ternários. O mesmo ocorreu com o dado “desida” (“desci da”); entretanto, nesse caso, há motivação lexical e ortográfica, uma vez que esse dado se assemelha às palavras “descida” e “decida”.

Ademais, esses dados, como já comentado, apresentam estrutura diferente do que predomina no conjunto de dados analisados, pois a prosodização dos clíticos em português, geralmente, se dá mais à direita e, nestes dados, os clíticos se uniram à palavra mais à esquerda, equivalendo a uma possível estrutura enclítica, como “ame-o”. Consideramos que esse tipo de ocorrência ilustra a complexidade que envolve os dados de hipossegmentação do EF II, por essa razão, na seção 3.3, exploraremos alguns desses exemplos.

No sétimo ano, também ocorreu a hipossegmentação de um verbo e um artigo; entretanto, a estrutura de “vio” (“vi o”) parece ser distinta das apresentadas anteriormente, pois, nessa ocorrência, temos a formação de um pé binário trocaico. Esse dado também carrega uma informação lexical e ortográfica, pois se assemelha à forma verbal “viu”. Além dessa ocorrência, no sétimo houve a junção do verbo “tem” com o clítico “que”. Notamos, no caso desse dado, que o clítico “que” tem valor de preposição, já que, de acordo com dicionário Houaiss (2009), modernamente, a estrutura “ter que” mais infinitivo funciona como “ter de” mais infinitivo. Notamos, nesse exemplo, que também ocorre a troca da nasal <m> pela nasal <n>, uma consideração da convenção ortográfica quanto ao emprego de <M, N> em meio de palavra.

Essa estrutura também pode ser interpretada como evidência de emergência de gramaticalização de construções morfossintáticas. Segundo Barros (2012), a estrutura “ter que” sofreu um processo de gramaticalização, na qual o verbo “ter” passou por uma alteração categorial, inserindo-se no paradigma dos auxiliares. Esse é o mesmo caso do dado “tende” (“tem de”) no nono ano.

Voltando à descrição dos dados do tipo 4, no sexto, sétimo e oitavo ano, ocorreram sete hipossegmentações de um advérbio com uma conjunção, uma preposição ou um pronome, como ilustrado no quadro a seguir:

Quadro 13. Hipossegmentações de advérbio + conjunção, preposição ou pronome

6º ano	7º ano	8º ano
“só que” > “soque” “além de” > “alende”	“sempre que” > “sempreque” “não me” > “nãome” “só me” > “some” “lá em” > “lem” (quatro ocorrências)	“agora se” > “agorase”

Nesses casos, o advérbio funciona como hospedeiro dos clíticos que o seguem, constituindo estruturas enclíticas distintas daquelas formadas de verbo e pronome, como “ajude-me”. Esses registros, mesmo poucos, trazem indícios de prosodização de clíticos à esquerda e se assemelham aos anteriormente descritos por configurarem indícios de cliticização à esquerda do hospedeiro, resultando uma configuração de ênclise que envolveria palavras distintas de verbos e pronomes. No entanto, como vimos, o maior conjunto de dados envolve, principalmente, verbos e pronomes hipossegmentados, que deveriam ser separados por hífen. Essa junção possibilita, assim como a combinação anterior, a formação de possíveis palavras prosódicas, na sua maioria paroxítonas constituídas por duas ou três sílabas.

Para finalizar, vale mencionar que nesse conjunto também encontramos a hipossegmentação da sequência “para que” > “praque” (duas ocorrências no oitavo ano e uma no nono ano), na qual verificamos a redução “para” > “pra”. Esses dados, como comentado na seção anterior, indiciam o processo de prosodização da preposição “para” junto a um pronome.

Os dados apresentados parecem revelar, em alguma medida, a representação que o sujeito escrevente faz do escrito institucionalizado na sua escrita. Nas palavras de Corrêa (1997, p. 271), “o escrevente lida, basicamente, com o que supõe ser – a partir não só do que aprendeu na escola, mas, em grande parte, do que assimilou fora dela – a visão escolarizada de código institucionalmente reconhecido”.

Desse modo, consideramos a junção do clítico com o verbo em posição enclítica como marca linguística que indicia a tentativa do escrevente de “adequar” a (sua) escrita ao que ele imagina ser a escrita escolar. Essa “adequação”, segundo Corrêa (1997), é estabelecida por meio dos entrelaçamentos do oral/falado e do letrado/escrito que “se evidenciam sempre que o escrevente leva a extremos uma tal imagem sobre o código escrito” (CORRÊA, 1997, p. 271).

A seguir, apresentaremos as hipossegmentações de tipo 5, que envolvem a combinação de mais de duas palavras gráficas.

### 3.2.5 Tipo 5: Outras combinações de palavras

O conjunto de dados do tipo 5 se refere à junção de mais de uma palavra e/ou clíticos. Junturas desse tipo não se mostraram comuns no EF II: houve somente duas ocorrências no sexto ano e uma no sétimo ano, como se visualiza na Tabela seguir.

Tabela 17. Coeficiente de dados do tipo 5

<b>Outros</b>					
<b>Ano</b>	6 <sup>a</sup>	7 <sup>a</sup>	8 <sup>a</sup>	9 <sup>a</sup>	Total
<b>Hipo</b>	2	1	0	0	3
<b>Palavras</b>	50981	74399	44713	97452	267545
<b>Coeficiente</b>	0,004	0,001	0,00	0,000	0,005

No tipo 5, encontram-se dados gerados a partir da junção de mais de duas palavras e/ou clíticos. No quadro a seguir, são apresentadas as ocorrências identificadas.

Quadro 14. Hipossegmentações de mais de uma palavra e/ou clíticos

<b>6º ano</b>	<b>7º ano</b>
“eu por exemplo” > “euporezemplo” “filho do cavalo” > “filhodocavalo”	“a gente brigou” > “ajenteprigou”

Em “euporezemplo”, há a formação de uma possível frase entoacional, identificada pela estrutura sintático-semântica da sentença, de modo que é possível atribuir contorno entoacional ao trecho hipossegmentado, considerado o enunciado em que ocorre, como se observa a seguir. Vale destacar que “por exemplo” deveria ser separado por vírgulas, segundo as convenções de pontuação. A hipossegmentação dá indícios de resultar da configuração de uma frase entoacional “eu por exemplo”, já que a frase entoacional constituída por “eu”, pode se reestruturar por ser relativamente pequena, formando com a frase entoacional seguinte uma única frase (cf. TENANI, 2002).

3.14 [Eu] I, [por exemplo] I, [pesco peixe] I, [nado no rio]I, [e me divirto muito na escola] I

3.15 [Eu por exemplo] I, [pesco peixe] I, [nado no rio]I, [e me divirto muito na escola] I

Similarmente, em “ajenteprigou”, a sentença hipossegmentada constituída de sujeito e verbo dá indícios de ter sido motivada pela configuração de uma única frase entoacional: [A gente brigou] I.

Já em “filhodocavalo”, há uma estrutura sintática composta por sintagmas complexos, o que leva à configuração de sintagmas fonológicos, tal como previsto na abordagem da fonologia prosódica de Nespor e Vogel (1986, 2007). Cada sintagma tem configuração entoacional característica (TENANI, 2002), que pode ter sido característica prosódica motivadora da hipossegmentação observada.

Esses dados só ocorreram nos dois primeiros anos do EF II e foram escritos por sujeitos diferentes. Neles é possível observar resquícios da escrita do EF I, uma vez que essa é uma característica da escrita dessa etapa escolar, como mostrado por Capristano (2007a, 2007b), para dados de escrita do noroeste paulista, e Cunha (2010), para dados do interior gaúcho. Essas autoras demonstram que, no EF I, a maioria das junções entre mais de dois elementos propicia a formação de constituintes mais altos da hierarquia prosódica, tal como propostos por Nespor e Vogel (1986, 2007), como frases entoacionais e enunciados fonológicos.

As hipossegmentações de tipo 5 parecem ser relacionadas, mais diretamente, às práticas orais/faladas, uma vez que esse tipo de estrutura envolve características entoacionais típicas de constituintes mais altos da hierarquia prosódica de Nespor e Vogel (1986, 2007).

### **3.3 Casos especiais de hipossegmentação**

No conjunto do material analisado, identificamos dados que instigam uma análise um pouco mais detalhada em função da complexidade que se revela de modo mais generoso ao analista. Nesta seção, realizamos análises qualitativas de hipossegmentações com o intuito de argumentar que segmentações não-convencionais de palavras escritas são decorrentes de um complexo processo que envolve três aspectos essenciais: (i) o Outro, instância de representação do funcionamento da linguagem como modo de enunciação escrito; (ii) a escrita, concebida como constituída de modo heterogêneo; (iii) o sujeito, concebido sócio-historicamente, permeado por ideologias, e responsável por movimentar diversas posições-sujeito em seus textos (CAPRISTANO, 2007b).

Faremos, a seguir, análises de cunho linguístico-textual que não se limitam somente às hipossegmentações, mas também abrangem outras características relevantes para descrição de

possíveis motivações para sua ocorrência, como as propostas de produção textual. Essas análises encontram suporte na perspectiva sobre a relação entre fala e escrita adotada neste trabalho, apresentada na seção de fundamentação teórica


Faz-se importante explicar que considerar a dimensão textual-discursiva não implica em descartar aspectos da dimensão fônica apresentados anteriormente, uma vez que, de acordo com Chacon (1998), ambas as dimensões constituem a organização rítmica da escrita.

Selecionamos uma proposta textual apresentada aos alunos no sexto ano do EF II e dois textos elaborados a partir dessa proposta. A escolha dessa proposta textual se deu por essa apresentar o cordel, texto cujas características deixam de modo mais explícito ao leitor o entrelaçamento de práticas orais/faladas e letradas escritas.

Apresentaremos, inicialmente, a proposta textual na qual os alunos se basearam para produzir seus textos e, em seguida, analisaremos os textos “O rei do Hip-Hop” e “A minha vida”.

Na imagem a seguir trazemos a proposta textual selecionada:

Figura 18. Enunciado da proposta - Cordel (6º ano - 2008)

**unesp**  **UNIVERSIDADE ESTADUAL PAULISTA**  
**“JÚLIO DE MESQUITA FILHO”**  
 Campus de São José do Rio Preto

Projeto de extensão: Oficinas de Leitura, Interpretação e Produção Textual – Coordenação Luciani Tenani

**E.E. Professora Zulmira da Silva Salles**

Nome: \_\_\_\_\_ Série/Turma: 501

Proposta 2 Data: 04/04/08

- O cordel abaixo conta a história do Mestre Bimba, um capoeirista. Leia-o com atenção.

**Bimba espalhou capoeira nas praças do mundo inteiro**

Manoel dos Reis Machado	Ganhou a vida, com tudo
Famoso na capoeira	Fez carvão, cortou madeira
É nossa árvore do bem	Foi trapicheiro e carpina
Nosso grande mestre Bimba	Estivador de primeira
Nome que até hoje vem	Mas o que fez com mais classe
Por não perder pra ninguém	Só foi jogar capoeira.

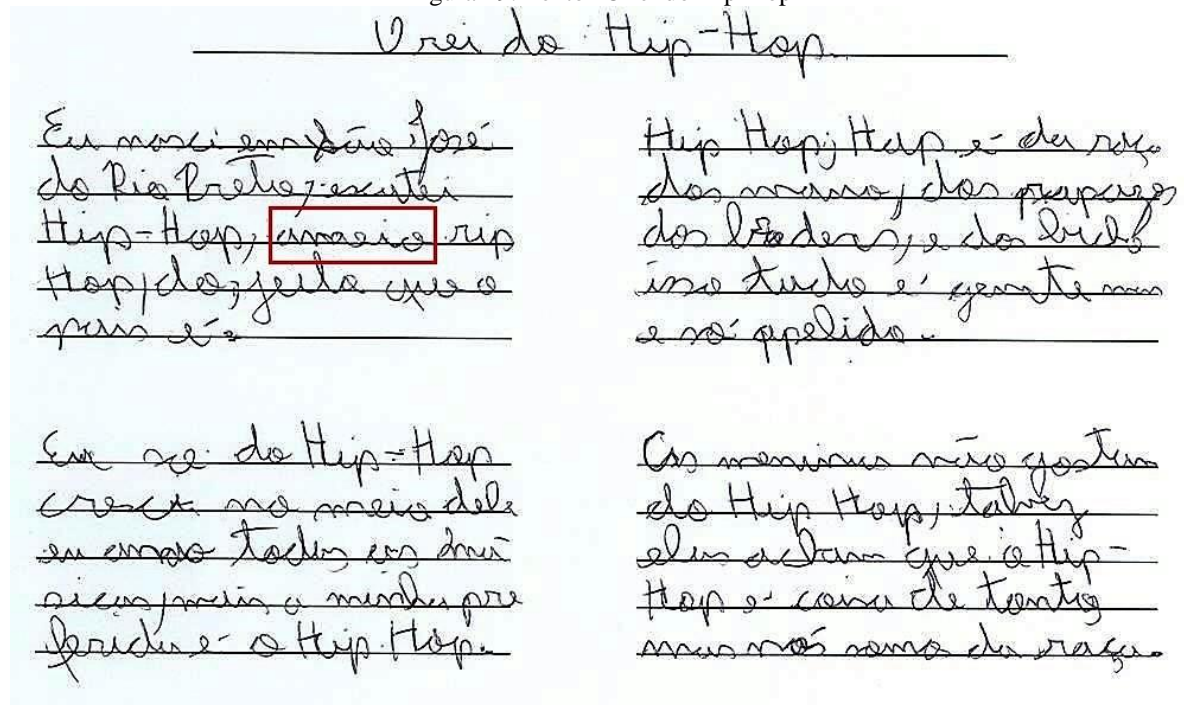
Engenho Velho de Brotas	Luís Gonzaga foi rei
Local do seu nascimento	Cantando mulher rendeira
Salvador sua cidade	Pelé foi o rei da bola
Onde alegria e tormento	Com meião e com chuteira
Lhe deram vivacidade	E mestre Bimba sem dúvida
Força, coragem e talento.	Foi o rei da capoeira.

- Como visto, o cordel é um tipo de texto em que se pode identificar tanto elementos da narração (personagens e ações) quanto elementos da poesia (rima).
- Levando em conta esses aspectos, escreva um cordel que conte um pouco de sua história.
- Você deve escrevê-lo em primeira pessoa, no espaço de quatro estrofes, abaixo delimitado.
- Dê um título a seu texto.

Fonte: Banco de Dados de Escrita do EF II (TENANI, 2015)

Seguindo as instruções, o sujeito 30 escreveu o seguinte texto:

Figura 19. Texto "O rei do Hip-Hop"



Fonte: Banco de Dados de Escrita do EFII (Texto: Texto Z08\_5A\_09M\_02)<sup>43</sup>

Em termos gerais, nesse texto, o aluno responde ao solicitado pela proposta ao descrever o rei do *hip-hop*. Interpretamos que essa escolha dialoga com o tema principal do texto apresentado na proposta que trata do rei da capoeira. *Hip-hop* e capoeira, além de estarem associados a origens africanas, são danças coreografadas realizadas coletivamente e podem ser consideradas, relativamente, marginais na sociedade brasileira.

Do ponto de vista prosódico, o texto do aluno também atende à proposta; embora não tenha rimas, não podemos dizer que o texto não tem métrica. A métrica pode ser observada por meio do uso de vírgulas que demarcam fronteira de frases entoacionais e, ainda, pela segmentação não-convencional: “ameio” (“amei o”).

Na primeira estrofe deste texto, temos duas frases entoacionais bem semelhantes quanto à estrutura sintática e número de sílabas, uma seguida da outra:

3.16 [Escutei hip-hop,] I

3.17 [ameio hip hop] I

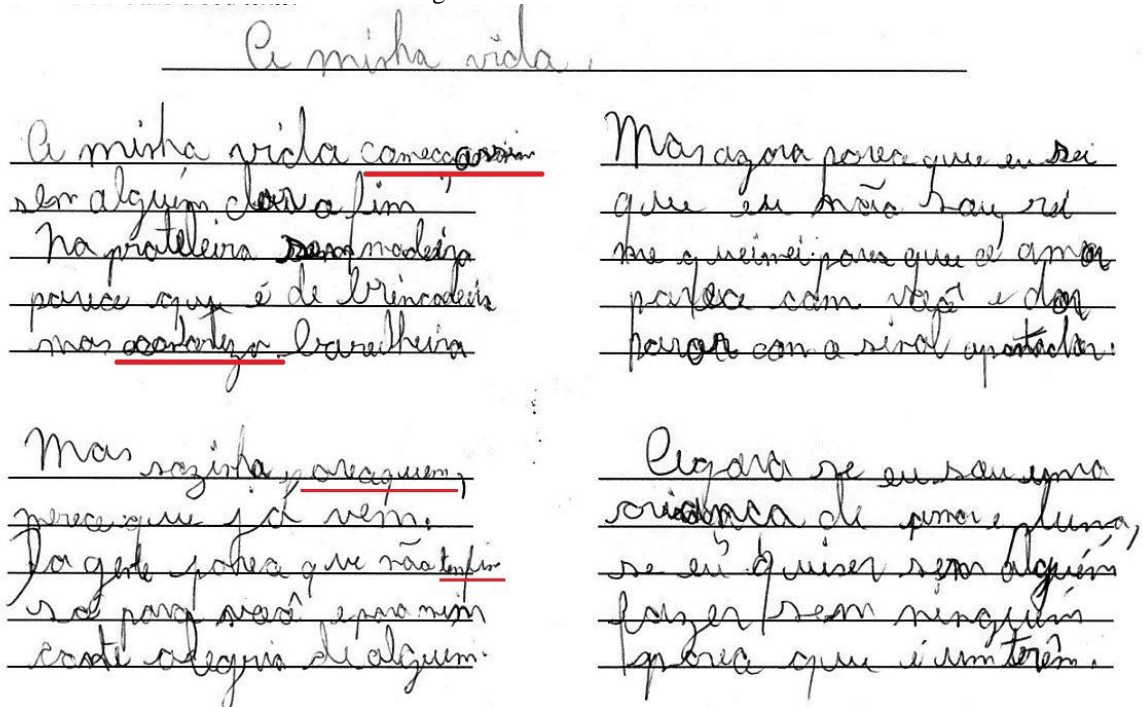
<sup>43</sup> Leitura possível: Eu nasci em São José/do Rio Preto, escutei/Hip-Hop, ameio rip/Hop, do, jeito que o/país é./Eu so do Hip-Hop/cresci no meio dele/eu amo todas as mu-/sicas, mais a minha pre/ferida é o Hip-Hop./Hip Hop, Hap é da raça/dos mano, dos rapazes/dos broders, e dos bicho/isso tudo é gente mas/e só apelido./As meninas não gostam/do Hip Hop, talvez/elas acham que o Hip-/Hop é coisa de tonto,/mas nós somos da raça.

Na comparação entre (3.16) e (3.17), observamos que, em (3.16), há uso não-convencional de hífen para unir “hip-hop”, configurando-se duas porções gráficas “escutei” e “hip-hop”; em (3.17), o artigo, caso não fosse hipossegmentado, seria diferente do enunciado (3.16) quanto à organização gráfica das sílabas em relação à “escutei”, nesse verso, não há, porém, uso de hífen em “hip hop”. Dessa forma, o registro gráfico em (3.17) parece indiciar a busca por regularidade gráfica.

Ademais, como mostrado nas seções anteriores, a prosodização dos clíticos em português geralmente se dá mais à direita e, neste dado, um clítico se uniu à palavra mais à esquerda, equivalendo a uma possível estrutura enclítica, como “ame-o”, na qual o artigo “o” também poderia ser interpretado como pronome anafórico, uma vez que retoma as palavras “hip-hop”. Esse dado é único em relação a outros dados do EF II, e nos permite interpretar que a sua ocorrência tem relação com a organização verbo-visual do enunciado escrito.

Passamos, a seguir, para a análise do texto “A minha vida” escrito por outro sujeito.

Figura 20. Texto “A minha vida”



Fonte: Banco de Dados de Escrita do EFII (Texto: Z08\_5B\_13M\_02)<sup>44</sup>

<sup>44</sup> Leitura possível: A minha vida começa assim/sem alguém dar o fim/ Na prateleira sem madeira/ parece que é de brincadeira/ mas concerteza barulheira/ Mas sozinha parecequem,/ merece que já vem./ Da gente parece que não temfim/ só para você e para mim/ cante alegria de alguém./Mas agora parece que eu sei/que eu não sou, rei/me queimei parece que o amor/parece com você e dor/parar com o sinal apontador./Agora se eu sou uma/criança de amor e pluma,/se eu quiser ser alguém/fazer sem ninguém/parece que é um trêm.

Esse texto, de um ponto de vista linguístico-textual, é muito rico em indícios que apontam para a circulação do escrevente por práticas orais/faladas e letradas/escritas. A seguir, apresentamos alguns desses indícios:

Neste texto, identificamos as seguintes hipossegmentações: “começaassim” (começa assim), “concerteza” (com certeza), “parecequem” (parece quem), “temfim” (tem fim). Essas hipossegmentações podem ser analisadas em duas estruturas prosódicas distintas, a saber:

3.18  $\omega + \omega > \Phi$ : “começaassim”, “tenfim”, “parecequem”

3.19  $cl + \omega > \omega$ : “concerteza”

Em (3.18), temos a junção de palavras prosódicas, formando possíveis frases fonológicas, constituídas de verbos e complementos, sendo os primeiros cabeças lexicais. Os dois versos nos quais os dados “começaassim” e “tenfim” ocorrem podem ser interpretados como possíveis frases entoacionais não-finais, pois ambas apresentam um contorno entoacional delimitado pelo fim e início de um verso. Notamos que essas frases entoacionais rimam com as frases entoacionais finais, a saber:

3.18.1 “[A minha vida ‘começaassim’] I [sem alguém dar o fim] I”;

3.18.2 “[Da gente parece que não ‘tenfim’] I [só para você e para mim] I”.

Esse jogo de rimas proporciona uma estrutura métrica a esse texto que é fortemente marcada, pois nota-se claramente preocupação com a palavra visando atender ao gênero cordel no que diz respeito à sua organização em versos rimados. A atenção com a organização métrica dos enunciados escritos pode ter, por hipótese, colocado em segundo plano a articulação sintático-semântico das sentenças. Podemos notar ainda, no registro gráfico do texto, que essas duas hipossegmentações parecem ter sido inseridas em um segundo momento da escrita com o intuito de se formarem rimas, como uma possível busca do sujeito para atender ao que lhe foi proposto.

O terceiro dado parece seguir a mesma tendência dos anteriores, já que a união da palavra “quem” com a palavra “parece” indicia um contorno entoacional que pode ser atribuído a esse verso, formando, junto com as demais palavras, uma possível frase entoacional:

3.18.3 “[Mas sozinha ‘parecequem’,] I [merece que já vem.] I”.

A vírgula que segue a hipossegmentação corrobora a interpretação desse verso como uma possível frase entoacional, pois, como já mencionado, as vírgulas, tanto convencionais quanto não-convencionais, como o é nesse caso, marcam posições onde se encontram, subjacente à escrita, fronteiras de estrutura de frases entoacionais não-finais (SONCIN; TENANI, 2015).

Outro aspecto interessante é que as palavras com as quais essas hipossegmentações rimam são sempre monossilábicas: “começaassim – fim”, “parecequem – vem”, “tenfim – mim”. O aluno parece fazer rimas de uma porção maior com uma porção menor, atribuindo ritmo poético ao seu texto. Essa valorização do ritmo poético também faz com que o sujeito escrevente use rimas em todas as estrofes, sendo a maioria dessas rimas provenientes das terminações usadas no texto da proposta “-em”, “-eira”.

Notamos ainda a flutuação na grafia das palavras “fim” e “parece”, grafadas de modo convencional e não-convencional. Quando a palavra “fim” é grafada de forma convencional, temos um clítico (o artigo “o”) antes dela (o fim) e não uma palavra prosódica como em “tenfim”. Observa-se que a palavra “parece”, nas duas vezes em que foi grafada convencionalmente, estava no meio dos versos, não era seguida pela palavra “quem” e não havia a necessidade de rima.

Dessa maneira, a partir da análise desses dados, podemos elencar possíveis razões para a ocorrência de hipossegmentações neste texto: (i) a posição da palavra no verso, ocupando posição adjacente a fronteiras não-finais de I; (ii) necessidade de construção de versos com rimas; visando à construção de um texto poético.

A única hipossegmentação que fugiu a essa característica foi “com certeza”, que segue a tendência geral dos dados de hipossegmentação. Esse tipo de hipossegmentação é uma tendência nos textos do EF II e já foi explicada anteriormente, quando tecemos considerações a favor da interpretação desse tipo ocorrência como uma possível palavra prosódica, já que o clítico funciona como a sílaba pretônica da palavra prosódica “concerteza”.

A ênfase no uso das rimas pelo sujeito escrevente do texto “A minha vida” e a não priorização da rima pelo sujeito escrevente do “O rei do hip-hop” possibilita-nos observar a interação entre leitura e escrita. A maneira como o aluno lê a proposta de produção textual o orienta na identificação do que seja mais significativo para ser retomado no texto que será redigido a partir dessa proposta. O escrevente do primeiro texto demonstra, em sua leitura da proposta, ter priorizado o desenvolvimento do tema solicitado; já o escrevente desse segundo

texto parece ter priorizado a composição de rimas na elaboração de seu texto em resposta ao que lhe foi proposto.

Para concluir esta análise, observamos que os dados de hipossegmentação apontam para a tensão dos sujeitos escreventes em lidarem com a convenção ortográfica aprendida na escola e, ao mesmo tempo, com o conhecimento linguístico que dispõem enquanto falantes e escreventes. E que os conhecimentos mobilizados pelos alunos na produção do texto são permeados por diferentes práticas orais e letradas de uso da linguagem.

Assim, a análise linguístico-textual mostra que o trabalho com a heterogeneidade da escrita possibilita identificar a circulação do sujeito com a linguagem, com seu texto escrito. E também corrobora a afirmação de Tenani (2016) de que grafias não-convencionais de palavras não são apenas dados de língua, mas especialmente dados de discurso.

Na próxima seção, apresentaremos a análise de mais dois textos que apresentam a estrutura “o que” hipossegmentada.

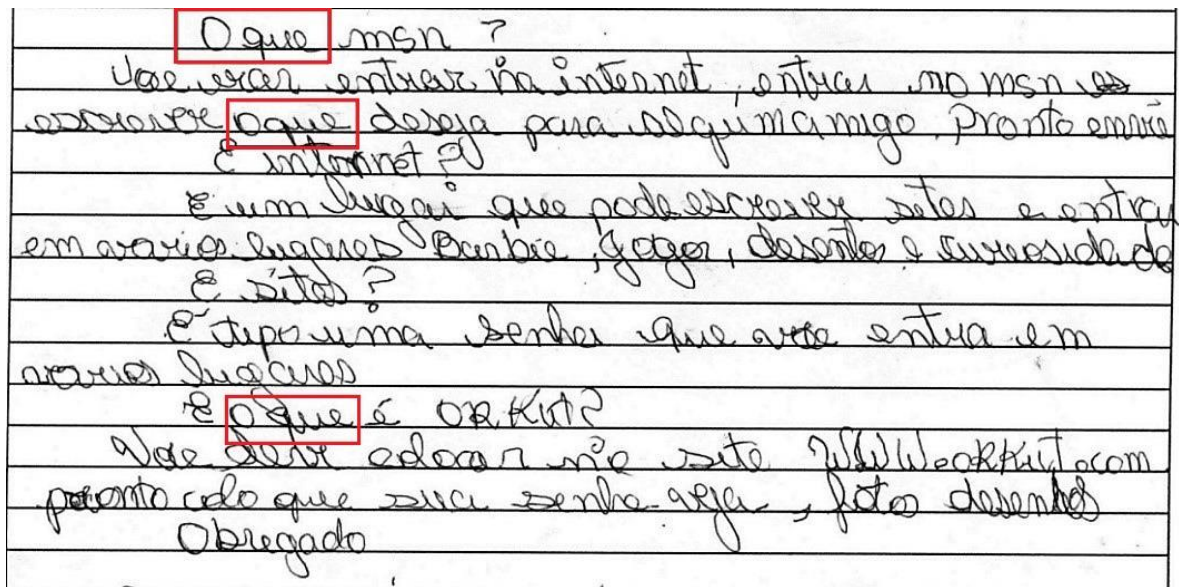
### **3.4 Flutuações entre o convencional e não-convencional**

Nesta seção, por meio da análise que apresentaremos a seguir, buscaremos mostrar qual o funcionamento da estrutura “o que” hipossegmentada, visto que esse tipo de junção ocorreu em grande quantidade em textos do EF II, flutuando os registros entre convencional e hipossegmentado. Além desse grande número de ocorrências, observamos também que essa flutuação não parece ser aleatória. Dessa maneira, nossa interpretação para essas flutuações é a de elas demonstram certa regularidade que diz respeito à mobilização de critérios linguísticos diversos.

Para investigar o funcionamento dessas ocorrências, analisaremos dois textos em que há a flutuação na grafia da estrutura “o que”.

O primeiro deles, apresentado na figura abaixo, foi feito por um aluno que cursava o sexto ano do EF II. Neste texto, o aluno foi convidado a escrever uma carta pessoal que tratasse do funcionamento da internet e do uso do MSN:

Figura 21. Flutuação de segmentação de “o que”



Fonte: Banco de Dados de Escrita do EFII (Texto: Z08\_5E\_28F\_04)<sup>45</sup>

No texto dado é possível observar que “o que” é segmentado ora de forma convencional, ora de forma não-convencional:

- 3.20 “Você vai entrar na internet, entrar no MSN escrever ‘oque’ deseja para algum amigo”.
- 3.21 “O que (é) MSN?”
- 3.22 “E o que é Orkut?”


Em (3.20), de um ponto de vista prosódico, há junção de, potencialmente, dois elementos clínicos. Segundo Tenani (2011b) a partícula “que”, sozinha não apresenta acento lexical; entretanto, ao se unir a partícula “o”, recebe um acento prosódico e passa a se comportar como um hospedeiro. De um ponto de vista sintático, o verbo “escrever”, conforme o dicionário Houaiss (2009), pode ser classificado como transitivo direto, isto é, ele se conecta diretamente com seu complemento, sem a necessidade de preposições. Essa informação permite levantar a hipótese de que “oque” funciona como complemento do verbo, estando mais opaca ao aluno a construção de artigo “o” seguido de conjunção “que”. Já em (3.21) e (3.22), “o” e “que” funcionam como construção sintática mais transparente quanto à sua composição em dois elementos, característica que favoreceria seu registro convencional. De um ponto de vista

<sup>45</sup> Leitura possível: O que é MSN?. Você vai entrar na internet, entrar no MSN escrever oque deseja para algum amigo. Pronto envie. E a internet? É um lugar que pode escrever sites e entrar em vários lugares Barbie, jogos, desenhos e curiosidade. E sites? É tipo uma senha que você entra em vários lugares. E o que é Orkut? Você deve colocar no site [www.orkut.com](http://www.orkut.com) pronto colo que sua senha veja, fotos desenhos. Obrigado.

sintático, nos enunciados (3.21) e (3.22), o artigo “o” e a conjunção interrogativa “que” funcionam como orações interrogativas do tipo Q- (PERINI, 2010). É importante mencionar que, de acordo com Perini (2010), uma oração interrogativa não é necessariamente semelhante a uma pergunta. “A oração interrogativa é um tipo de construção gramatical, não um enunciado com a função locucionária de pergunta.” (PERINI, 2010, p. 124).

Esse texto também possibilita considerar características linguístico-textuais do texto que parecem motivar a hipossegmentação observada. Como feito na seção anterior, consideremos a proposta textual apresentada ao aluno.

Figura 22. Enunciado da proposta de redação - Internet e o uso do MSN (6º ano - 2008)

**unesp**  UNIVERSIDADE ESTADUAL PAULISTA  
“JÚLIO DE MESQUITA FILHO”  
Campus de São José do Rio Preto


Projeto de extensão: Oficinas de Leitura, Interpretação e Produção Textual – Coordenação Luciani Tenani

E.E. Professora Zulmira da Silva Salles

Nome: \_\_\_\_\_ Série/Turma: 5a C

Proposta 4 Data: 14/10/08

- Observe os quadrinhos abaixo do Chico Bento e seu primo da cidade.



- Chico Bento e seu primo fazem coisas parecidas ao longo do dia, mas cada um não sabe o que o outro faz. Suponha que você, como o Chico Bento, tenha ouvido falar sobre a possibilidade de contar sobre as coisas que faz por meio da internet.

- Escreva uma carta ao seu primo da cidade, pedindo para ele contar **o que é a internet** e como se faz para mandar mensagens por meio do MSN.
- Seu texto deve conter de 20 a 25 linhas e deve ser escrito à tinta. Seu texto não deve ultrapassar os limites designados para a escrita.

Fonte: Banco de Dados de Escrita do EF II (TENANI, 2015)

O enunciado da proposta que motivou a produção do texto em análise solicita que o escrevente se coloque na posição do personagem Chico Bento e pergunte a seu primo da cidade como funciona a internet e o MSN. Seguindo essa solicitação, o aluno busca mobilizar duas vozes: a voz do primo do sítio que não sabe o que é internet e a voz do primo da cidade que explica o funcionamento dessa tecnologia.

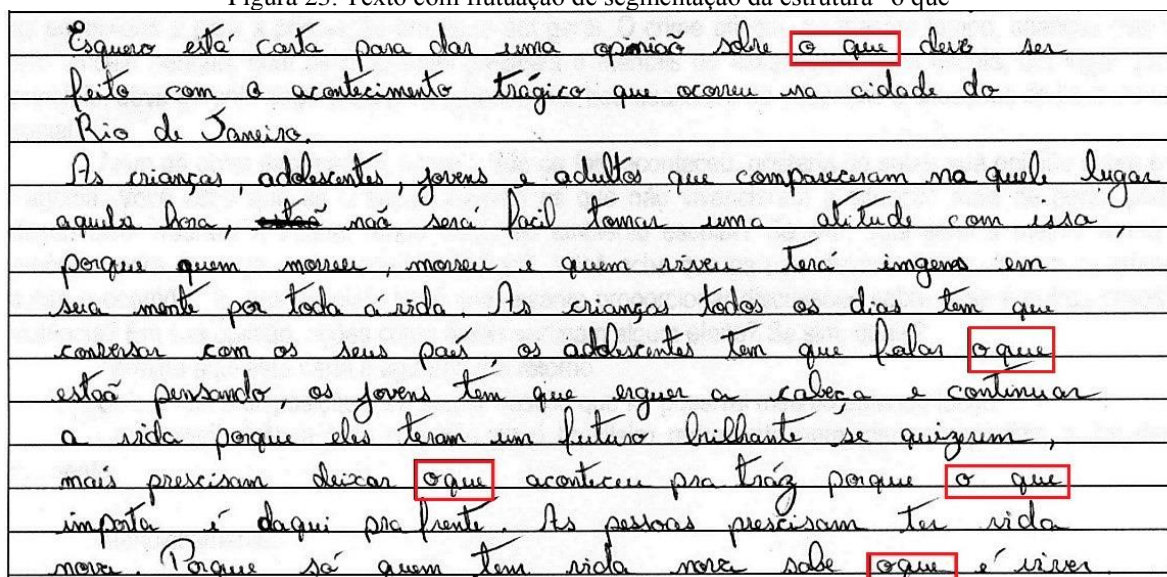
Notamos que os dados de “o que” segmentados de forma convencional decorrem do falar do primo do sítio que é, de certa forma, atravessado pelo enunciado da proposta de redação (“escreva uma carta ao seu primo da cidade, pedindo para ele contar o que é internet [...]”). O

dado hipossegmentado ocorre na fala do primo da cidade que explica “oque” fazer com a internet. Assim, é possível levantar a hipótese de que essas estruturas são mobilizadas a depender da posição-sujeito que o escrevente assume ao escrever<sup>46</sup>.

Notamos ainda que há, na penúltima linha do texto, um dado de hipersegmentação na fala do primo da cidade (“colo que” – “coloque”). Dessa maneira, os dois únicos dados de segmentação não-convencional que ocorrem neste texto aparecem no falar do primo da cidade. Esse é um indício de que há, neste texto, uma movimentação de vozes que se mostra, de forma sutil, por meio da segmentação de palavras.

O segundo texto a ser analisado, é uma carta de opinião, escrito por um aluno do nono ano do EF II, sobre o massacre ocorrido em Realengo no ano de 2011. A seguir, o texto é apresentado.

Figura 23. Texto com flutuação de segmentação da estrutura “o que”



Fonte: Banco de Dados de Escrita do EFII (Texto: Z11\_8E\_12F\_03)<sup>47</sup>

Neste texto, a flutuação da grafia de “o que” ocorre de modo que se identificam duas grafias convencionais (3.23) e (3.24) e duas não-convencionais (3.25) e (3.26):

<sup>46</sup> Segundo Ferreira (2001), a posição-sujeito é um objeto imaginário que representa, no discurso, os lugares ocupados pelos sujeitos na estrutura de uma formação social. Desta maneira, não há um sujeito único, mas diversas posições-sujeito, as quais estão relacionadas com determinadas formações discursivas e ideológicas.

<sup>47</sup> Leitura possível: Esquevo está carta para dar uma opinião sobre o que deve ser feito com o acontecimento trágico que ocorreu na cidade do Rio de Janeiro. As crianças, adolescentes, jovens e adultos que compareceram na quele lugar aquela hora, ~~estão~~ não será fácil tomar uma atitude com isso, porque quem morreu, morreu e quem viveu terá imagens em sua mente por toda a vida. As crianças todos os dias tem que conversar com os seus pais, os adolescentes tem que falar o que estão pensando, os jovens tem que erguer a cabeça e continuar a vida porque eles tem um futuro brilhante se quiserem, mais precisam deixar o que aconteceu pra trás porque o que importa é daqui pra frente. As pessoas precisam ter vida nova. Porque só quem tem vida nova sabe o que é viver.

3.23 “Escrevo está carta para dar opinião sobre o que deve ser feito com o acontecimento trágico que ocorreu na cidade do Rio de Janeiro”.

3.24 “porque o que importa é daqui pra frente.”

3.25 “[...] os jovens tem que erguer a cabeça e continuar a vida porque eles teram um futuro brilhante se quiserem, mas precisam deixar “oque” aconteceu pra trás [...]”

3.26 “Porque só quem tem vida nova sabe “oque” é viver.”

Há uma ocorrência de “o que” que gera dúvida quanto ao espaçamento entre as palavras, o que não permite afirmar, com certeza, se se trata de um dado de segmentação não-convencional ou convencional. Entretanto, teceremos argumentos para considerá-la como uma hipossegmentação:

3.27 “[...] os adolescentes tem que falar o que\* estão pensando [...]”

Em (3.23) e (3.24), o pronome “o” funciona como clítico, do mesmo modo que nos enunciados (3.21) e (3.22). No entanto, diferente desses enunciados, as ocorrências em (3.23) e (3.24) não são frases interrogativas. Dessa maneira, o que parece estar em jogo, nesses casos, é o funcionamento de cada uma das partículas (“o” e “que”) em separado: a partícula “o” tem valor de pronome demonstrativo e funciona como especificador da partícula “que”, que, nesses enunciados, opera como um pronome relativo que faz referência à informação que vem logo em seguida: “o que deve ser feito com o acontecimento trágico que ocorreu na cidade do Rio de Janeiro” e “o que importa é daqui pra frente”. Em outras palavras, as ocorrências “o que”, tanto em (3.23) quanto em (3.24), funcionam como antecipadores da informação que serão tratadas no decorrer do texto.


Nos enunciados (3.25) e (3.26), as ocorrências de “oque” funcionam como uma única palavra prosódica. Ambas parecem operar de forma a retomar o assunto já tratado no texto: em “deixar ‘oque’ aconteceu pra trás [...]”, “o que” retoma o acontecimento referido pelo aluno, o qual já foi tratado de forma indireta no texto e de forma direta na proposta textual; e, em “Porque só quem tem vida nova sabe ‘oque’ é viver.”, “o que” retoma o trecho em que o aluno fala de quem morreu e viveu (“porque quem morreu, morreu e quem viveu terá imagens em sua mente por toda à vida.”) e o fato de que só quem tem vida – isto é, sobreviveu à tragédia de Realengo – sabe o que é viver.

Esse mesmo funcionamento é observado na ocorrência em que ficamos em dúvida. Em (3.27), de certa maneira, a ocorrência “o que” retoma a mensagem de motivação para as crianças

e jovens que passaram pela tragédia. De um ponto de vista gráfico, o espaçamento entre os elementos é muito menor se comparado com (3.23) e (3.24). Por essas razões, consideramos a ocorrência em (4.27) como um caso de hipossegmentação.

A proposta de produção textual (abaixo transcrita e disponibilizada por completo em anexo), que embasou a escrita desse texto, orienta o aluno a escrever uma carta em resposta à carta enviada por Francisco González<sup>48</sup>, na qual ele faz várias perguntas sobre a tragédia que ocorreu em uma escola do bairro Realengo, Rio de Janeiro:

Figura 24. Enunciado da proposta - Massacre em Realengo (9º ano - 2011)

**unesp**  **UNIVERSIDADE ESTADUAL PAULISTA**  
**"JÚLIO DE MESQUITA FILHO"**  
 Campus de São José do Rio Preto

Projeto de extensão: **Oficinas de Leitura, Interpretação e Produção Textual** – Coordenação Luciani Tenani

**E.E. Professora Zulmira da Silva Salles**

Nome: \_\_\_\_\_ Série/Turma: 8ª C Nº 28

Proposta 3 Data: 16/05/2011

- Você percebeu que, em sua carta, Francisco pede mais informações sobre o ocorrido em Realengo e faz uma série de perguntas. Produza uma carta em resposta ao texto escrito por Francisco González. Para atender às expectativas de Francisco, é preciso expor algumas informações sobre a tragédia a que ele se refere e apresentar o que você acha sobre esse tema, levando em consideração as perguntas que ele faz. Tente, ao responder as perguntas, justificar a sua opinião apresentando alguns argumentos.
- Seu texto deve conter de 25 a 30 linhas e deve ser escrito à tinta. Seu texto não deve ultrapassar os limites designados para a escritura.

Fonte: Banco de Dados de Escrita do EF II (TENANI, 2015)

Como podemos ver na Figura 24, na produção de carta, o aluno deveria apresentar fatos e expor argumentos sobre a tragédia, visando responder às perguntas de González. No entanto, nesse texto, o aluno não expõe os fatos da tragédia e muito menos responde às dúvidas de González. O aluno expõe sua opinião sobre o que deve ser feito à respeito da tragédia, isto é, ele apresenta argumentos de motivação para as crianças e adolescentes que sobreviveram ao massacre.

Contudo, como vimos, o aluno mobiliza mecanismos linguísticos que dialogam com as informações sobre o massacre, mostrando que, mesmo indiretamente, faz referência à tragédia, às vítimas que morreram, às vítimas que sobreviveram, e, de alguma forma, tenta responder às perguntas de González. Sobre essa referência aos acontecimentos que envolvem a tragédia, isto

<sup>48</sup> Nesta carta, Francisco González apresenta-se como um espanhol que viveu algum tempo no Brasil e manifesta interesse sobre a tragédia que ocorreu em Realengo. Para saber mais sobre o ocorrido, o espanhol faz algumas perguntas, como “o que ocorreu, como aconteceu, quantas crianças foram mortas e qual a idade média dessas crianças” (cf. Anexo 2 desta dissertação).

é, à coletânea da proposta, Corrêa (1997) explica que essa pressuposição de que o leitor conhece os textos da coletânea é um indício da relação que o sujeito escrevente mantém com o já falado/escrito e ouvido/lido.

A partir dessa análise, podemos afirmar que a flutuação entre “oque” hipossegmentado e “o que” segmentado, nesses textos, não é simplesmente aleatória. Ela demonstra certa regularidade que diz respeito à mobilização de critérios linguísticos diversos.

Em resumo, notamos que a estrutura “oque” hipossegmentada:

- i. funciona como uma possível palavra prosódica;
- ii. funciona como um objeto direto;
- iii. mobiliza construções gramaticais distintas: uma construção gramatical explicativa e estrutura com um elemento anafórico, uma vez que retoma um assunto já tratado no texto.

A estrutura de “o que” segmentada de forma convencional:

- i. funciona como dois elementos clíticos;
- ii. funciona como interrogativas abertas e como elementos catafóricos que antecipam assuntos que o escrevente ainda tratará nos textos.

As grafias “o que” e “oque” evidenciam diversas características linguísticas-discursivas decorrentes das propostas textuais sugeridas:

- i. há mobilização de vozes e de posições-sujeito nos textos;
- ii. há mobilização de informações da proposta textual não presentes diretamente no texto escrito do aluno.

No que diz respeito aos gêneros textuais, notamos que os dados de segmentação analisados ocorreram em gêneros textuais semelhantes: no sexto ano, ocorreram em cartas pessoais e, no nono ano, em uma carta de opinião. Esses gêneros têm em comum o fato de ser proposto um possível diálogo. No sexto ano, o diálogo deveria ser com um primo da cidade e, no nono ano, um diálogo, por meio de cartas, com um estrangeiro interessado em saber mais sobre uma tragédia ocorrida no Brasil.

Dessa maneira, podemos interpretar que textos em que se instauram diálogos, mesmo que indiretos, são motivadores de ocorrências de flutuação de segmentação de “o que”, pois, as relações de diálogos suscitam diversos sentidos e posições-sujeito na materialidade textual.

Em termos gerais, afirmamos, com base nessa análise, que as flutuações entre segmentações convencionais e não-convencionais mostradas:

- i. se distinguem entre si pela configuração prosódica, gramatical e linguística discursiva;

- ii. dão indícios da inserção dos alunos em práticas orais/faladas e letradas/escritas;
- iii. dão indícios do trânsito do sujeito escrevente pelo código escrito institucionalizado, no qual ele faz uso do que imagina ser – a partir de suas experiências com a escrita e com a própria visão escolar sobre a escrita – o código escrito institucionalizado (CORRÊA, 2006);
- iv. são marcas do “complexo processo que envolve o outro como instância representativa da linguagem (e da escrita em particular), a escrita na complexidade de seu funcionamento (heterogeneamente constituída) e a criança enquanto sujeito escrevente” (CAPRISTANO, 2007a, p. 160).

\*\*\*

Os resultados apresentados no Capítulo 3 apontam para a necessidade de estudos específicos: (i) sobre o estatuto linguístico das estruturas de “o que” no EF II; (ii) sobre casos de flutuações entre segmentações convencionais e não-convencionais no EF II.

Para concluir este capítulo, retomamos algumas questões centrais. A partir das análises apresentadas, buscamos descrever longitudinalmente as principais características no que diz respeito à organização prosódica e morfossintática de hipossegmentações identificadas em textos de alunos do EF II. Podemos afirmar que, de modo geral, os alunos do EF II demonstram dificuldades para segmentar palavras gramaticais, já que a maioria dos dados (85,7% – 472/550) envolveu esse tipo de palavra.

No sexto ano, 120 ocorrências (85,1%) envolveram palavras gramaticais. Além dessas, nove (6,4%) ocorrências são de palavras lexicais que se comportaram como clíticos. No sétimo ano, assim como no ano anterior, 120 ocorrências envolveram palavras gramaticais e 22 (13,4%) palavras lexicais evidenciaram comportamento de clíticos. No ano seguinte, as ocorrências envolvendo palavras gramaticais diminuíram para 91 (91%) e somente seis (6%) dados foram de palavras lexicais que podem ser interpretadas como clíticos. No último ano do EF II, aumentaram em 133 as ocorrências (91,7%) envolvendo palavras gramaticais; e palavras lexicais que se configuraram como clíticos diminuíram para 5 ocorrências (3,4%).

As análises conduzidas neste capítulo também apontaram para a possível circulação dos escreventes por prática orais/faladas e letradas/escritas. Algumas ocorrências pareceram demonstrar de modo mais marcado a “tentativa, por parte do escrevente, de representação do planejamento conversacional” (CORRÊA, 1997, p. 185), como, por exemplo, o dado “tabem”, que, de acordo com nossa interpretação, evidencia características de enunciados falados em

interação face a face, nos quais é possível notar diálogos entre personagens. Outro exemplo seria o dado *seila*, que demonstraria o jogo argumentativo prosodicamente marcado, indicando uma hesitação usada pelo escrevente para argumentar e sinalizar seu ponto de vista.

Ainda observamos alçamento a uma escrita institucionalizada que se evidenciou, por exemplo, em dados de hipossegmentações como “concerteza” e “endia”, nos quais os escreventes grafam a letra <n> ao invés de <m> – “com certeza” e “em dia” – indiciando uma possível ‘adequação’ ao que eles imaginam ser o código escrito institucionalizado.

A escrita institucionalizada também foi observada, principalmente no sexto ano, em ocorrências da sequência verbo + pronome, como “ajudalos” e “levala”, pois a estrutura enclítica não é comum no PB, mas é ensinada e valorizada pela instituição escolar; desse modo, interpretamos que o escrevente a usa a fim de “adequar” a sua escrita ao que ele imagina ser uma escrita institucionalmente aceita.

No próximo capítulo, apresentamos as considerações finais, nas quais retomamos os resultados encontrados e sugeriremos possíveis questões de investigação e desenvolvimentos futuros.

---

## CAPÍTULO 4

---

### CONSIDERAÇÕES FINAIS

Nesta dissertação, apresentamos uma descrição longitudinal das principais características da escrita de alunos do EF II no que diz respeito ao funcionamento das hipossegmentações, caracterizadas pela ausência de fronteira gráfica entre palavras.

Inicialmente, trouxemos uma discussão sobre fronteira de palavra, a fim de problematizar o fato de não haver, necessariamente, sobreposições entre fronteiras de palavra prosódica e morfológica. Mostramos que a noção de palavra é complexa, uma vez que se trata de uma unidade linguística constituída de características fonológicas, morfológicas, sintáticas e semânticas.

Também discutimos diferentes perspectivas e suas respectivas interpretações atribuídas aos casos de segmentação não-convencional, entre as quais citamos: (i) perspectiva escolar, para a qual os casos de hipo e hipersegmentação são interpretados como erros ortográficos decorrentes de indesejada interferência do oral no escrito, resquícios de problemas de alfabetização, e/ou decorrentes de distúrbio específico de escrita, a disortografia, e, (ii) perspectiva linguística, a partir da qual as grafias não-convencionais são interpretadas como efeito do trânsito dos sujeitos escreventes por práticas sociais orais/faladas e letradas/escritas.

No âmbito dessa discussão, assumimos perspectiva linguística, a qual interpreta os casos de segmentação não-convencional de palavras como pistas do trabalho do sujeito com a língua(gem). Mostramos ainda que essa assunção implica em mobilizar conceitos de língua e escrita constituídas heterogeneamente pelos modos de enunciação falado e escrito, marcados pela diversidade linguístico-histórica dos escreventes (CORRÊA, 1997, 2004).

Uma vez que características dos enunciados falados também são observáveis nos enunciados escritos, descrevemos que características prosódicas se mostram de modo regular no conjunto de textos considerados, por meio de análise prosódica dos dados de hipossegmentação, desenvolvida com base no arcabouço teórico da Fonologia Prosódica de Nespor e Vogel (1986, 2007), que organiza os enunciados falados em uma série de constituintes fonológicos arranjados hierarquicamente. Com base nessa abordagem teórica, explicamos os modos pelos quais a organização rítmico-entoacional dos enunciados constituem os textos escritos e motivam os casos de hipossegmentação apresentados.

Quanto aos resultados, inicialmente, por meio de análise estatística dos dados, concluímos que considerar o coeficiente gerado a partir da razão entre quantidade de hipossegmentações e número de palavras escritas é importante, pois chegamos ao resultado que demonstra que os casos de hipossegmentações observadas no EF II decrescem em função do aumento de anos de escolarização. Essa informação confirma a nossa hipótese de que o aumento da escolarização leva à diminuição das ocorrências de hipossegmentações.

Esse resultado também vai ao encontro de trabalhos sobre hipersegmentações em textos do EF II, como o trabalho de Tenani (2016) que demonstra, apoiada em análise estatística, que as hipersegmentações diminuem ao longo dos anos do EF II. Resultado semelhante é constatado em trabalhos sobre hipossegmentações em textos do EF I. Conforme demonstram Capristano (2007a, b) e Cunha (2010, 2004), há decréscimo de hipossegmentação em função do aumento de anos de escolarização, embora não tenha sido feito tratamento estatísticos dos dados. Portanto, os resultados dos diferentes estudos apontam para o efeito positivo dos anos de escolarização na diminuição da ocorrência de grafias não-convencionais de palavra quanto à localização de fronteira ortográfica entre palavras.

No que se refere à análise que levou em conta características prosódicas e morfossintáticas, observamos que a junção entre clítico e palavra prosódica foi a estrutura que mais gerou dados de hipossegmentação em todos os anos do EF II. Ao considerar o coeficiente gerado a partir da razão entre número de hipossegmentações e o número de palavras escritas, constatamos que esse tipo de junção apresentou um decréscimo ao longo dos anos do EF II.

Nesse conjunto de dados, observamos, em todos os anos do EF II, um possível não-reconhecimento do estatuto autônomo dos clíticos, que seriam interpretados, pelos escreventes, como sílabas que integram a palavras que os acompanham. Observamos ainda que nas hipossegmentações há formação de sequências CV, como “denoite”, “tevejo” e “porai”. A formação de sequências CC na fronteira onde há a juntura morfológica ocorreu em contextos específicos, como em “envez” e “concerteza”, em que o clítico apresenta coda nasal. Em sequências VV decorrente da junção entre palavras, como em “que eu”, ocorreu registro de processo de sândi vocálico externo, como “queu”. Interpretamos que esses dados indiciam, de modo mais marcado, a busca pelo código escrito institucionalizado, pois concorrências como, por exemplo, “queu” e “quele” classificadas como Tipo 1, evidenciam alçamento do escrevente a uma escrita convencional. De modo geral, o conjunto de dados do Tipo 1 é constante em todos os anos, evidenciando a manutenção do trânsito dos escreventes por elementos de natureza fonético-fonológico e pelo código escrito institucionalizado.

O segundo tipo de hipossegmentação que mais ocorreu foi a junção entre dois clíticos, como “oque” e “doque”. Mesmo considerando o coeficiente de ocorrência do dado, esse tipo de junção não apresentou decréscimo, isto é, os escreventes terminaram o EF II com mais dados desse tipo do que começaram. Todavia, os dados desse grupo são, em sua maioria, a junção dos clíticos “o” e “que”. Observamos grande flutuação entre a grafia convencional da estrutura “o que” e a grafia não-convencional “oque”. Essas flutuações, muitas vezes, ocorrem em um mesmo texto, sugerindo, por hipótese, fatores linguísticos atuando nessa flutuação. A grafia hipossegmentada, nos dois textos analisados, funcionou como uma possível palavra prosódica com função de objeto direto e mobilizou diversas construções gramaticais – no sexto ano funcionou como uma construção gramatical explicativa e no nono ano como um elemento anafórico, uma vez que retoma um assunto já tratado no texto. Já a grafia convencional funcionou como elemento clítico que ora desempenhou função de interrogativa aberta e ora funcionou como elemento catafórico. Ademais, argumentamos que as grafias “o que” e “oque” evidenciaram mobilização de (i) vozes e de posições-sujeito nos textos; (ii) informações da proposta textual não presentes diretamente no texto escrito do aluno. Portanto, esse se mostra como um tipo de dado que merece estudo específico, dada sua complexidade morfosintática e semântico-discursiva.

Interpretamos, ainda, que esse conjunto de dados do Tipo 2 também é resultante do não-reconhecimento do estatuto autônomo dos clíticos. Assim, nesse conjunto, as hipossegmentações seriam interpretadas, pelo escrevente, como uma única palavra prosódica. Argumentamos ainda que as ocorrências desse conjunto seriam motivadas simultaneamente por: (i) informações de natureza prosódica, uma vez que mobilizam componentes átonos da língua, e (ii) informações letradas/escritas e orais/faladas, já que algumas ocorrências, como “uque” e “oq”, evidenciaram o trânsito do escrevente por práticas letradas/escritas e orais/faladas.

Os resultados do conjunto de dados do Tipo 2 apontam para a necessidade de estudos que contemplem casos de flutuação entre o convencional e o não-convencional na escrita do EF II a fim de defender, de modo mais contundente, que dados de flutuação de segmentação de palavras escritas não são simples erros e/ou acertos, eles evidenciam a complexidade linguístico-textual que subjaz as relações entre prosódia e escrita e possibilitam a interpretação das grafias analisadas enquanto dados de discurso (TENANI, 2016).

O terceiro tipo de estrutura prosódica encontrada nos textos do EF II foi a junção entre duas palavras prosódicas, como “tabem”. O coeficiente dessa estrutura apresenta decréscimo ao longo, chegando à 0,01 no nono ano. No conjunto do Tipo 3, observamos a prosodização

das palavras prosódicas “para” (“pramim”, “pafente”) e “está” (“tabem”, “tabom”) e a hipossegmentação de estruturas perifrásticas, como “voudormir”, “fuibrincar”. Em termos gerais, as ocorrências do Tipo 3 apontam para o possível reconhecimento de palavras como clítics. Esse reconhecimento pode estar relacionado à representação que o escrevente faz de vocábulos relacionados a práticas mais informais de oralidade/fala e de letramento/escrita.

A quarta estrutura prosódica encontrada nos textos do EF II foi a junção de palavra prosódica e clítico. O coeficiente dessa junção também apresentou um decréscimo ao longo do EF II. No sexto ano, houve 0,05 ocorrências e nono ano, somente 0,02. Nesse conjunto de dados, houve maior mobilização, principalmente no sexto ano, de combinações com a ausência do hífen, que possibilitou a formação de possíveis palavras prosódicas, na sua maioria paroxítonas constituídas por três sílabas e quatro sílabas, como em “buscala” e “socorrela” respectivamente.

Argumentamos que os dados apresentados parecem referir-se à representação que o sujeito escrevente faz do escrito institucionalizado para a sua escrita, pois consideramos a junção do clítico com o verbo em posição enclítica como tentativa do escrevente de alçar a (sua) escrita ao que ele imagina ser a escrita convencional. Essa tentativa do escrevente, da perspectiva de Corrêa (1997) que adotamos, é estabelecida por meio dos entrelaçamentos do oral/falado e do letrado/escrito que “se evidenciam sempre que o escrevente leva a extremos uma tal imagem sobre o código escrito” (CORRÊA, 1997, p. 271).

No cópús de análise também encontramos três casos de junção envolvendo mais de uma palavra prosódica e/ ou clítics, como “euporezemplo”. Esse tipo de combinação ocorreu apenas duas vezes no sexto ano e uma vez no sétimo ano. Mostramos que esses dados, além de se constituírem como resquícios da escrita do EF I, abrangem estruturas como a frase entoacional e o enunciado fonológico. Interpretamos que a mobilização desses constituintes dá pistas da circulação do escrevente por práticas orais/faladas, pois parecem relacionadas à possibilidade de organização rítmica do enunciado.

A partir da análise qualitativa de dados selecionados também foi possível observar, além da circulação do escrevente por práticas orais/faladas e letradas escritas, a (sua) relação com o já falado/escrito e ouvido/lido (CORRÊA, 1997, 2004). Essa relação se mostrou de modo mais evidente nos momentos em que ocorreram hipossegmentações advindas da tentativa de o escrevente atender as expectativas escolar ao responder ao que é pedido na proposta, bem como pelo uso da coletânea de textos disponíveis na proposta.

Concluimos as considerações finais acerca do estudo desenvolvido, defendendo que esta dissertação traz importante contribuição metodológica para os estudos de segmentação de palavra escrita, na medida em que evidencia que considerar a quantidade de palavras escritas

por texto e a distribuição de ocorrências por escrevente são informações importantes para melhor fundamentar uma análise longitudinal. Considerar o coeficiente gerado a partir da razão palavras escritas/hipossegmentações nos permitiu mostrar que mais anos de práticas letradas escolares levam à diminuição dos casos de hipossegmentação, mas não necessariamente levam à extinção desse tipo de grafias não-convencionais. Essa informação também nos auxiliou na identificação de um conjunto de sujeitos que evidenciaram comportamento quantitativo diferente dos demais, fazendo-se necessária investigação de caso considerando cada sujeito em particular.

Além dessas contribuições, nossas análises evidenciaram que dados de hipossegmentação e de flutuação mobilizam diversos discursos que perpassam e constituem o texto e o sujeito escrevente. Esses discursos concretizam-se, na materialidade textual, por meio da mobilização de estruturas prosódicas dos enunciados em confluência com aspectos morfológicos e linguístico-textuais.

Ressaltamos que nosso estudo abre perspectiva de trabalhos futuros, no que diz respeito ao estudo de dados de flutuação de segmentação de palavras escritas, haja vista que essas flutuações gráficas evidenciam a complexidade da noção de palavra, unidade linguística que se constitui de características fonológicas, morfológicas, sintáticas e semânticas.

---

**REFERÊNCIAS**

---

ABAURRE, M. B. M. A relevância dos critérios prosódicos e semânticos na elaboração de hipóteses sobre segmentação na escrita inicial. **Boletim da Abralín**, Campinas, v. 11, p. 203-217, 1991.

\_\_\_\_\_.; GALVES, C. As diferenças rítmicas entre o português europeu e o português brasileiro: uma abordagem otimalista e minimalista. **DELTA: documentação de estudos em linguística teórica e aplicada**, São Paulo, v. 14, n. 2, p. 377-403, 1998.

\_\_\_\_\_.; FIAD, R. S.; MAYRINK-SABINSON, M. L. T. **Cenas de aquisição da escrita**: o trabalho do sujeito com o texto. Campinas: Mercado de Letras, 1997.

\_\_\_\_\_. O que revelam os textos espontâneos sobre a representação que faz a criança do objeto escrito? In: KATO, M. A. (Org.). **A concepção da escrita pela criança**. 2. ed. Campinas: Pontes Editores, 1992. p. 135-142.

\_\_\_\_\_.; GALVES, C. Os clíticos no português brasileiro: uma abordagem sintático-fonológica. In: CASTILHO, A. T.; BASÍLIO, M. (Org.). **Gramática do português falado**: estudos descritivos. Campinas: Ed. da UNICAMP, 1996. p. 273-320.

\_\_\_\_\_.; CAGLIARI, L. C. Textos espontâneos na primeira série: evidências da utilização, pela criança, de sua percepção fonética da fala para representar e segmentar a escrita. **Cadernos CEDES**: São Paulo, v. 14, p. 25-29, 1985.

ARAÚJO, D. P. B. A gramaticalização da partícula lá. **Linguística Textual e Pragmática**. Rio de Janeiro: Círculo Fluminense de Estudos Filológicos e Linguísticos, 2014.

BARROS, E. C. M. **Construções modais com “ter”**: gramaticalização e variação. 2012. 246 f. Tese (Doutorado em Linguística) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2012.

BELINE, R. M. **A gramaticalização da perífrase estar + gerúndio no português falado**. 1999. Dissertação (Mestrado em Linguística) - Universidade Estadual de Campinas, Campinas, 1999.

BISOL, L. (Org.). **Introdução a estudos de fonologia do português brasileiro**. Porto Alegre: EDIPUCRS, 1996.

BISOL, L. Mattoso Câmara Jr. e a Palavra Prosódica. **DELTA: documentação de estudos em linguística teórica e aplicada**, São Paulo. v. 20, n. especial, p. 59-70, 2004. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S0102-44502004000300006](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0102-44502004000300006)>. Acesso em: 21 dez. 2017.

\_\_\_\_\_. O clítico e o seu hospedeiro. **Letras de Hoje**, Porto Alegre, v. 40, n. 3, p. 163-184. 2005.

\_\_\_\_\_. O clítico e seu status prosódico. **Revista de Estudos da Linguagem**, Belo Horizonte, v. 9, n. 1, p. 5-20, 2000.

BOOP DA SILVA, T. **Formação de palavras compostas em português brasileiro: uma análise de interfaces**. 2010. 195 f, Tese (Doutorado em Letras) - Universidade Federal do Rio Grande do Sul, Porto Alegre, 2010.

BRASIL. Ministério da Educação. Secretaria de Educação Fundamental. **Parâmetros curriculares nacionais: língua portuguesa**. Brasília, 1997.

CAPRISTANO, C. C. A propósito da escrita infantil: uma reflexão sobre segmentações não-convencionais. **Letras de Hoje**, Porto Alegre. v 39, n. 3, 2004, p. 245-260.

\_\_\_\_\_. **Aspectos de segmentação na escrita infantil**. São José do Rio Preto. 166 f. Dissertação (Mestrado em Estudos Linguísticos) - Instituto de Biociências, Letras e Ciências Exatas, UNESP, 2003.

\_\_\_\_\_; ANGELO, C. M. P. Concepções de escrita de professores em formação de um curso de Letras. **Revista Todas as Letras**, São Paulo, v. 17, p. 29-42, 2015.

\_\_\_\_\_. **Mudanças na trajetória da criança em direção à palavra escrita**. 2007. 252 f. Tese (Doutorado em Linguística Aplicada) - Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, 2007b.

\_\_\_\_\_. **Segmentação na escrita infantil**. São Paulo: Martins Fontes, 2007a. v. 1.

CARMO, M. C. do; TENANI, L. E. Vogais médias pretônicas na variedade do noroeste paulista: Uma análise sociolinguística. **Alfa: revista de linguística**, São Paulo, v. 57, p. 607-637, 2013.

CHACON, L. Constituintes prosódicos e letramento em segmentações não-convencionais. **Letras de Hoje**, Porto Alegre, v. 39, n. 3, p. 223-232, 2004.

\_\_\_\_\_. Flutuação na segmentação de palavras: relações entre constituintes prosódicos e convenções ortográficas na escrita infantil. **Filologia e Linguística Portuguesa**, São Paulo, v. 15, p. 369-383, 2013.

CHACON, L. **Ritmo da escrita**: uma organização do heterogêneo da linguagem. São Paulo: Martins Fontes, 1998.

\_\_\_\_\_. Prosódia e letramento em hipersegmentações: reflexões sobre a aquisição da noção de palavra. In: CORRÊA, M. L. G. (Org.). **Ensino de língua**: representação e letramento. Campinas: Mercado de Letras, 2006. p. 155-167.

CORRÊA, M. L. G. Heterogeneidade da escrita: a novidade da adequação e a experiência do acontecimento. **Filologia e Linguística Portuguesa**, São Paulo, n. 2, p. 269-286, 2006.

\_\_\_\_\_. **O modo heterogêneo de constituição da escrita**. 1997. 434f. Tese de Doutorado. Instituto de Estudos da Linguagem, Unicamp, 1997.

\_\_\_\_\_. **O modo heterogêneo de constituição da escrita**. São Paulo: Martins Fontes, 2004.

CUNHA, A. P. N. **A hipo e a hipersegmentação nos dados de aquisição da escrita**: um estudo sobre a influência da prosódia. 2004. 132 f. Dissertação (Mestrado em Educação) – Universidade Federal de Pelotas. Pelota, 2004.

\_\_\_\_\_.; MIRANDA, A. R. M. A influência da hierarquia prosódica em hipossegmentações da escrita de crianças de séries iniciais. **Revista Virtual de Estudos da Linguagem**, [S. l.], v. 1, p. 1-19, 2007.

\_\_\_\_\_. **As segmentações não-convencionais da escrita inicial**: uma discussão sobre o ritmo linguístico do português brasileiro e europeu. 2004. 188 f. Tese (Doutorado em Educação) – Universidade Federal de Pelotas, Pelotas, 2010.

DE LEMOS, Cláudia Thereza Guimarães. Das vicissitudes da fala da criança e de sua investigação. **Cadernos de Estudos Linguísticos**, Campinas, p. 41-69, 2002.

FERNANDES-SVARTMAN, F. R. A entoação das sentenças clivadas em português brasileiro e a interface sintaxe-fonologia. **Filologia e Linguística Portuguesa**, São Paulo, v. 14, p. 37-56, 2012.

FERREIRO, E.; TEBEROSKY, A. **Psicogênese da língua escrita**. Porto Alegre: Artes Médicas, 1999.

\_\_\_\_\_. et al. **Chapeuzinho vermelho aprende a escrever: estudos psicolinguísticos comparativos em três línguas**. São Paulo: Ática, 1996.

FERNÁNDEZ, A. Y. et al. Avaliação e intervenção da disortografia baseada na semiologia dos erros: revisão da literatura. **Revista CEFAC**, São Paulo, v. 12, n. 3, p. 499-504, 2010.

GUZZO, N. B. **A prosodização de clíticos e compostos em português brasileiro**. 2015. 232 f. Tese (Doutorado em Letras) – Universidade Federal do Rio Grande do Sul, Instituto de Letras, Porto Alegre, 2015.

HOPPER, P. J.; TRAUGOTT, E. C. **Grammaticalization**. Cambridge: Cambridge University Press, 1993.

HOUAISS, A. **Dicionário eletrônico Houaiss da língua portuguesa: versão 3.0.1**. Rio de Janeiro: Objetiva, 2009. 1 CD-ROM.

ILARI, R. (Org.). **Gramática do português culto falado no Brasil: palavras de classe fechada**. São Paulo: Contexto: FAPESP, 2015.

MARCATO, F. **Análise prosódica de preposições monossilábicas**. 2013. 158 f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista Júlio de Mesquita Filho, Instituto de Biociências, Letras e Ciências Exatas, 2013.

NESPOR, M.; VOGEL, I. **Prosodic phonology**. Dordrecht: Foris Publications, 1986.

\_\_\_\_\_. **Prosodic phonology: with a new foreword**. Berlin: New York: Mouton de Gruyter. 2007.

NEVES, M. H. M. **Gramática de usos do português**. São Paulo: Ed. UNESP, 2000.

PARANHOS, F. C. **Segmentações não-convencionais de palavra nos quatro últimos anos do ensino fundamental: um estudo longitudinal**. 2014. 170 f. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Estadual Paulista. São José do Rio Preto, 2014

PAVEZI, V. C. Haplogia, elisão e monomorfema. **Estudos Linguísticos**, São Paulo. v. 34, p. 750-755, 2005.

PERINI, M. A. Interrogativas abertas (interrogativas –q). In: \_\_\_\_\_. **Gramática do português brasileiro**. Parábola, 2010. p. 124-125.

QUADROS, E. S. de; SCHWINDT, L. C. Um estudo sobre a relação entre palavra morfológica e palavra fonológica em vocábulos complexos do português brasileiro. In: **Anais Círculo de Estudos Linguísticos do Sul**, 2008, Porto Alegre, 2008, p. 1-18.

ROSA, M. C. **Introdução à morfologia**. São Paulo: Contexto, 2000.

SILVA, L. M.A prosodização de clíticos preposicionais do português brasileiro: o que dados de escrita podem revelar?. **Raído**, Dourados, v. 10, p. 11-25, 2016.

SIMIONI, T. O clítico e seu lugar na estrutura prosódica em português brasileiro. **Alfa: revista de linguística**, São Paulo, n. 52(2), p. 431-446, 2008. Disponível em: <[www.alfa.ibilce.unesp.br](http://www.alfa.ibilce.unesp.br). > Acesso em: 12 dez. 2015.

SONCIN, G. C. N.; TENANI, L. E. Emprego de vírgula e prosódia do Português brasileiro: aspectos teórico-analíticos e implicações didáticas. **Filologia e Linguística Portuguesa**, São Paulo, v. 17, p. 305-326, 2015.

TENANI, L. E. A grafia dos erros de segmentação não-convencional de palavras. **Cadernos de Educação**, Pelotas, v. 35, p. 247-269, 2010.

\_\_\_\_\_; PARANHOS, F. C. Análise prosódica de segmentações não-convencionais de palavras em textos do sexto ano do EF. **Filologia e Linguística Portuguesa**, São Paulo, v. 13, p. 477-504, 2011.

\_\_\_\_\_. A segmentação não convencional de palavras: evidências de representações de relações entre enunciados falados e escritos. In: **Anais 2. Simpósio Mundial de Estudos da Língua Portuguesa: ULTRAPASSAR FRONTEIRAS, JUNTAR CULTURAS**, 2009 p.107-127.

\_\_\_\_\_. A segmentação não-convencional de palavras em textos do ciclo II do Ensino Fundamental. **Revista da ABRALIN**, Brasília, v. 10, n. 2, p. 91-119, 2011b.

\_\_\_\_\_. **Banco de dados de escrita do ensino fundamental II**. São José do Rio Preto: Instituto de Biociências, Letras e Ciências Exatas, UNESP, 2015. Disponível em: <<http://www.convenios.grupogbd.com/redacoes/Login>>. Acesso em: 10 nov. 2015.

TENANI, L. E. **Domínios prosódicos no português do Brasil:** implicações para a prosódia e para a aplicação de processos fonológicos. 2002. 331 f. Tese (Doutorado em Linguística) – Universidade Estadual de Campinas, Campinas, 2002.

\_\_\_\_\_. Hipersegmentação de palavras: análise de aspectos prosódicos e discursivos. **Linguagem & Ensino**, Pelotas, v. 16, n. 2, 2013, p. 305-326.

\_\_\_\_\_. Letramento e segmentações não-convencionais de palavras. In: TFOUNI, L. V. (Org.). **Letramento, escrita e leitura:** Questões Contemporâneas. Campinas: Mercado de Letras, 2011a. p. 229-243.

\_\_\_\_\_. Notas sobre a relação entre constituintes prosódicos e ortografia. **Revista de Estudos da Linguagem**, Belo Horizonte, v. 16, p. 231-245, 2008.

\_\_\_\_\_; SILVEIRA, A. A. M. da. O alçamento das vogais médias na variedade culta do noroeste paulista. **Alfa: Revista de Linguística**, São Paulo, v. 52, p. 447-464, 2008.

\_\_\_\_\_; LONGHIN-THOMAZI, S. Oficinas de leitura, interpretação e produção textual no Ensino Fundamental. **Em Extensão**, Uberlândia, v. 13, n. 1, p. 20 - 34, 2014.

\_\_\_\_\_; FIEL, R. P. O papel de estruturas prosódicas e morfossintáticas em hipossegmentações do Ensino Fundamental II. **Revista do GEL**, Araraquara, v. 13, p. 107-129, 2016.

\_\_\_\_\_. **Prosódia e escrita:** uma análise a partir de (hiper) segmentações de palavra. 2016. 213 f. Tese (Livre-docência) – Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto, 2016.

\_\_\_\_\_. Segmentações não-convencionais e teorias fonológicas. **Letras de Hoje**, Porto Alegre. v. 39, n. 3, p. 233-244, 2004.

TORQUETTE, A. O. **Segmentações não-convencionais de palavras em escolas públicas e privadas.** 2016. 139 f. Tese. (Doutorado em Estudos Linguísticos) Instituto de Biociências, Letras e Ciências Exatas, Universidade Estadual Paulista, São José do Rio Preto, 2016.

VELOSO, J. **Verba Manent:** a palavra como unidade pertinente para a descrição linguística do português e de outras línguas flexionais. Porto Alegre: EDIPUCRS, 2016.

ZORZI, J. L. Alterações ortográficas nos transtornos de aprendizagem. In: MALUF, M. I. (Org.). **Aprendizagem:** tramas hipersegmentação de palavras do conhecimento, do saber e da subjetividade. Rio de Janeiro: Vozes; São Paulo, 2006. p. 144-162.

---

 ANEXOS
 

---

**ANEXO 1. Proposta – Internacionalização da Amazônia (9º ano - 2011)**


UNIVERSIDADE ESTADUAL PAULISTA  
 “JÚLIO DE MESQUITA FILHO”  
 Campus de São José do Rio Preto

Projeto de extensão: **Oficinas de Leitura, Interpretação e Produção Textual** – Coordenação  
 Luciani Tenani

E.E. Professora Zulmira da Silva Salles

**Nome:** \_\_\_\_\_

**Série/Turma:** 8ª \_\_\_\_\_

**Proposta 5**

**Data:** \_\_\_\_\_

Leia os textos abaixo:

**Texto 1**

Ao aumentar as manifestações em favor da internacionalização da Amazônia, o novo ministro do Meio Ambiente, Carlos Minc, respondeu que a aceitava – por que não? – desde que fossem internacionalizadas também Nova York e Paris. [...]

Com a idéia de internacionalizar também Paris e Nova York, Minc retomava um artigo do senador Cristovam Buarque, publicado em 2000 e de muito sucesso na internet. Cristovam Buarque conta que, questionado sobre a internacionalização, num debate nos Estados Unidos, disse que "como humanista" era a favor – assim como era a favor de internacionalizar as reservas de petróleo do mundo, libertando-as de países que arbitrariamente diminuem o capital financeiro global, sujeito a manobras dos especuladores. Era a favor igualmente da internacionalização dos museus, como o Louvre, "guardiães das mais belas peças produzidas pelo gênio humano". E também de Nova York, como sede da ONU, e de Paris, Veneza, Roma Londres e Rio de Janeiro, patrimônios da humanidade, sem se esquecer do arsenal nuclear americano, instrumento perigoso demais para estar sob controle de um só país.

Cristovam Buarque estava, claro, dando um chega pra lá nos interlocutores estrangeiros. Mas ao mesmo tempo desenhava um idílico mundo futuro, libertado das soberanias nacionais, em que tudo é de todos. Se tudo der certo no planeta (o que é discutível), quem sabe um dia, daqui a mil ou dois mil anos, chegemos lá. [...] A internacionalização só será aceitável quando se cumprirem duas premissas. Primeira: que desapareçam os estados nacionais. Segunda: que os grupos, ou comunidades, ou sociedades que restarem, mantenham entre si relações impecavelmente eqüitativas. Quem sabe, um dia...

*(Revista Veja, 28 de maio de 2008)*

**Texto 2**

A velha paranóia brasileira de que a soberania nacional na Amazônia está sob ameaça de potências estrangeiras e de ONGs ambientalistas acaba de ganhar um rosto. É o do milionário sueco Johan Eliasch, conselheiro do primeiro-ministro do Reino Unido, Gordon Brown. Em outubro de 2005, Eliasch comprou 160.000 hectares de terras de florestas na região norte

do país, uma área maior que a da cidade de São Paulo. Em seguida, criou a ONG *Cool Earth* (Esfrie a Terra) que tenta angariar doações para adquirir e conservar florestas na Amazônia. [...]

“O Papel do governo deve ser monitorar, fiscalizar e dirigir a ação desses grupos que querem investir, sejam brasileiros ou não”, diz Danilo Iglori, professor da Universidade Cambridge, na Inglaterra. “Mas o governo não tem dinheiro para investir na área e faz sentido que o mundo pague para preservar a floresta”.

(*Revista Época, 02 de junho de 2008*)

### **Texto 3**

“O jornal inglês *The Independent* publicou um editorial afirmando que a Amazônia ‘era importante demais para ser deixada aos brasileiros.’ O *New York Times* publicou artigo em que lembrava uma antiga frase de Al Gore, ex-vice-presidente americano, hoje santo protetor do meio ambiente global: ‘Ao contrário do que pensam os brasileiros, a Amazônia não é propriedade deles, e sim de todos nós.’”

(*Revista Veja, 28 de maio de 2008*)



## ANEXO 2. Proposta – Massacre em Realengo (9º ano – 2011)



UNIVERSIDADE ESTADUAL PAULISTA  
"JÚLIO DE MESQUITA FILHO"  
Campus de São José do Rio Preto

Projeto de extensão: **Oficinas de Leitura, Interpretação e Produção Textual** – Coordenação  
Luciani Tenani

E.E. Professora Zulmira da Silva Salles

**Nome:** \_\_\_\_\_

**Série/Turma:** 8ª \_\_\_\_\_

**Proposta 3**

**Data:** \_\_\_\_\_

- A carta transcrita abaixo foi postada no fórum de um site de relacionamento. Leia-a atentamente.

Barcelona, 17 de

abril de 2011. Caro

brasileiro,

Sou espanhol, moro atualmente em Barcelona, mas morei durante cinco anos em São Paulo, onde estudei e trabalhei como professor de espanhol. Durante o tempo que fiquei no Brasil, além de aperfeiçoar meu português, fiquei admirado pelo país, pela cultura popular e, em particular, pelo povo brasileiro.

Voltei a morar na Espanha há quase seis meses, mas ainda me sinto ligado ao Brasil. É pela consideração que tenho por esse país que escrevo esta carta. Vi nos noticiários internacionais a tragédia que aconteceu em uma escola da cidade do Rio de Janeiro, onde várias crianças foram assassinadas, e fiquei sensibilizado. Gostaria de ter mais informações sobre o caso: como aconteceu, quantas crianças foram mortas e qual a idade média dessas crianças.

A partir do que consegui saber, fiquei pensando sobre a gravidade de uma tragédia como essa para os envolvidos e para a população brasileira em geral. O crime atingiu, ao mesmo tempo, crianças que não tem vínculo nenhum com os problemas pessoais e mentais do assassino e uma escola, um lugar que, a princípio, deve garantir segurança para proporcionar boa qualidade de educação e situações de bom convívio social.

Além de obter informações sobre o que de fato aconteceu, gostaria de saber sua opinião sobre essa tragédia. Você acha que as crianças, mesmo as que não vivenciaram a situação mais de perto, podem desenvolver traumas e acabar tendo medo do ambiente escolar? Se sim, qual seria a melhor forma de proceder para amenizar essas consequências? Você acha que os pais devem conversar com as crianças sobre o ocorrido? E, nas escolas, seria interessante proporcionar discussões sobre esse e outros casos de violência? Em sua opinião, ações como essas surtiriam algum efeito? Se sim, quais?

Termino aqui esta carta e aguardo um retorno.

Coloco-me à disposição para ajudar naquilo que for possível mesmo estando longe.

Como retribuição a tudo o que o povo brasileiro representa para mim, compartilho a dor desse momento.

Atenciosamente,

Francisco  
González

**Massacre de Realengo** - Origem: Wikipédia, a enciclopédia livre

Massacre de Realengo refere-se ao assassinato em massa ocorrido em 7 de abril de 2011, por volta das 8h30min da manhã (UTC-3), na Escola Municipal Tasso da Silveira, localizada no bairro de Realengo, na cidade do Rio de Janeiro. Wellington Menezes de Oliveira, de 23 anos, invadiu a escola armado com dois revólveres e começou a disparar contra os alunos presentes, matando doze deles, com idade entre 12 e 14 anos. Oliveira foi interceptado por policiais, cometendo suicídio.

A motivação do crime Figura incerta, porém a nota de suicídio de Wellington e o testemunho público de sua irmã adotiva e o de um colega próximo apontam que o atirador era reservado, sofria *bullying* e pesquisava muito sobre assuntos ligados a atentados terroristas e a grupos religiosos fundamentalistas. O crime causou comoção no país e teve ampla repercussão em noticiários internacionais. A presidente



## TERMO DE REPRODUÇÃO XEROGRÁFICA

Autorizo a reprodução xerográfica do presente Trabalho de Conclusão, na íntegra ou em partes, para fins de pesquisa.

São José do Rio Preto, 22/03/2018



Assinatura do autor