

HIPERSEGMENTAÇÃO DE PALAVRAS EM TEXTOS DO EFII: CARACTERÍSTICAS PROSÓDICAS GERAIS

Luciani Tenani¹

Roberta Pereira Fiel²

RESUMO: Este artigo trata de hipersegmentações de palavras, como “na quela”, a partir da hipótese de que os espaços em branco não-convencionais dessas grafias são pistas de características prosódicas dos enunciados falados nas quais os estudantes se ancoram, quando produzem seus textos. As hipersegmentações analisadas foram identificadas em uma amostra longitudinal de textos escritos por alunos de 11 a 14 anos em uma escola pública paulista. A análise de cunho fonológico, segundo o arcabouço da Fonologia Prosódica, permite afirmar que a configuração prosódica das palavras, especialmente a configuração de sílaba átona seguida de pé troqueu dissílabo, é fortemente relevante para explicar a colocação não-convencional do espaço em branco dentro de palavra pela maioria dos estudantes considerados. Essa relação entre características prosódicas da língua e espaço em branco da escrita é interpretada como evidência do modo pelo qual alunos do Ensino Fundamental estabelecem relação entre fala e escrita.

Palavras-chave: palavra; escrita; prosódia; oralidade; língua portuguesa.

ABSTRACT: This article deals with hypersegmented words, such as “na quela”, based on the hypothesis that unconventional white spaces in written words are clues for prosodic characteristics of utterances in which students anchor their writings. The analyzed hypersegmented words have been identified in a longitudinal sample of texts written by students from 11-14 years at a public school in São Paulo. Phonological analyses, supported by the Prosodic Phonology framework, suggest that the prosodic word configuration, especially the configuration of unstressed syllables followed by a disyllabic trochee foot, plays a major role in explaining the unconventional presence of white spaces in words as written by most students. The relation between language prosodic features and white spaces in writing is interpreted as an evidence of how Junior High School students establish relationships between speech and writing.

Keywords: word; writing; prosody; orality; Portuguese.

INTRODUÇÃO

Neste artigo, são analisadas as chamadas hipersegmentações de palavras, um tipo de segmentação não-convencional de palavra escrita, identificadas em textos de alunos do Ensino Fundamental II (doravante EFII). Os objetivos são: (i) descrever a frequência de hipersegmentações em textos do EFII para o conjunto de sujeitos investigados, e (ii) identificar características prosódicas das hipersegmentações. A partir de uma perspectiva linguística de análise desses dados de escrita, pretende-se contribuir com pesquisas sobre os chamados erros ortográficos, particularmente, sobre como se dá a relação fala e escrita para estudantes nos anos finais do Ensino Fundamental, caracterizando-os em relação aos dados de estudantes dos anos iniciais dessa etapa básica de ensino.

As segmentações não-convencionais de palavras escritas são aqui caracterizadas pela ausência e/ou presença não-convencional de fronteiras gráficas em locais previstos pela

¹ Docente do Departamento de Estudos Linguísticos e Literários, Doutor em Linguística, Instituto de Biociências, Letras e Ciências Exatas, UNESP, lutenani@ibilce.unesp.br

² Licenciada em Letras, Instituto de Biociências, Letras e Ciências Exatas, UNESP, roh_fiel@hotmail.com.

convenção ortográfica (TENANI, 2011b). As fronteiras gráficas são o espaço em branco entre palavras e o sinal gráfico hífen. Essa proposta se particulariza em relação às encontradas na literatura na medida em que não é baseada apenas no espaço em branco, mas acrescenta o hífen como caracterizador de fronteira de palavra, mesmo que esse tenha função ambígua de separar e unir verbo e clítico e bases em compostos. Também é inovadora essa definição de segmentação não-convencional de palavra por tomar a fronteira gráfica como informação verbo-visual que caracteriza o dado de escrita.

As segmentações não-convencionais são classificadas em dois tipos, segundo a literatura: (i) hipossegmentações: quando há a ausência não-convencional da fronteira de palavra, como em: “concerteza” (“com certeza”); (ii) hipersegmentações: quando há a presença não-convencional da fronteira, como em: “com migo” (“comigo”). Um terceiro tipo é denominado *mescla*, por Chacon (2006), e *híbrido*, por Cunha (2004), como “come cheia” (“comecei a”).³ Para este estudo, interessam hipersegmentações de palavras pela presença do espaço em branco, como “em bora”. Em outra oportunidade, analisaremos hipersegmentações pela presença do sinal gráfico hífen dentro de palavra, como em, “conversa-se” (“conversasse”). Essa seleção se faz necessária em função da complexidade linguística dos dados.

A análise dessas grafias não convencionais é feita com base na posição teórica defendida por Abaurre (1996), segundo a qual interessam à Linguística os dados de escrita infantil na medida em que podem trazer evidências sobre, por exemplo, a pertinência de constituintes fonológicos. Defendemos essa perspectiva também para dados de escrita infanto-juvenil, como procuraremos demonstrar neste artigo.

Ainda seguindo a perspectiva inaugurada por Abaurre (1996), as hipersegmentações de palavras são vistas como hipóteses – predominantemente conflitantes entre si em um mesmo texto – do que seja palavra. Explicitamos, neste artigo, que essas hipóteses conflitantes são ancoradas em informações (i) **prosódicas** (que dizem respeito à organização das sílabas átonas e tônicas em diferentes constituintes fonológicos, como pé métrico, palavra fonológica, etc), e (ii) **letradas** (que dizem respeito à colocação de espaços em branco e do hífen como indicadores de fronteira de palavra na escrita).

³ Chacon (2006) caracteriza *mesclas* quando há dois ou mais constituintes prosódicos envolvidos na segmentação de palavras, enquanto Cunha (2004) caracteriza *híbridos* por grafias em que há hipossegmentação seguida de hipersegmentação de palavras.

Dentre as várias análises possíveis, restringimo-nos, neste artigo, a responder a questão central: a que fronteiras prosódicas as hipersegmentações de palavra podem estar associadas? A hipótese assumida foi: há correlação entre os limites de palavra escrita (grafados não-convencionalmente) e os limites de constituintes prosódicos. A proposta dessa questão de investigação, bem como a configuração do cópuz, está ancorada nos resultados encontrados a partir do desenvolvimento da pesquisa intitulada “Segmentação de palavras escritas: evidências de relações entre enunciados falados e escritos” (CNPq 309872/2012-0; FAPESP 2013/14546-5), desenvolvida pelo primeiro autor deste artigo.

PRESSUPOSTOS TEÓRICOS

Neste trabalho, partilhamos a noção de Abaurre, Fiad, Mayrink-Sabinson (1997), sobre dados de escrita infantil, e Tenani (2010), sobre dados de escrita infanto-juvenil, de que a segmentação não-convencional de palavras escritas não é uma categoria de erro ortográfico resultante de supostas interferências do oral no escrito. Também para Abaurre (1998), é ingênuo supor que o aprendiz da escrita escreve e representa a escrita como uma simples transcrição da fala, pois há elementos reveladores da incorporação de aspectos convencionais, de escolhas de estruturas típicas da escrita. Essas autoras tomam as produções escritas infantis como dados linguísticos complexos resultantes de hipóteses conflitantes elaboradas pelos escreventes a partir da relação que esses estabelecem com a língua(gem). Segundo Tenani (2011a, b), se características dos enunciados falados podem ser encontradas em segmentações não-convencionais, então, é possível identificar características fonológicas, especialmente, as prosódicas relacionadas à delimitação de fronteiras de palavras. Sobre essa relação entre escrita e prosódia, autores como Abaurre (1989, 1991), Abaurre e Cagliari (1995), Abaurre de Silva (1993) entre outros, têm mostrado que a presença não-convencional de espaços em branco e fatos da natureza fonética-fonológica estão intimamente ligados.

Autores como Chacon (2004, 2006), Capristano (2004, 2007) e Tenani (2010, 2011a, b) avançam no estudo de segmentação de palavras na medida em que propõem que a motivação dessas grafias não-convencionais da palavra escrita está relacionada a possibilidades de organização prosódica dos enunciados falados, tais como aqueles propostos por Nespor e Vogel (1986). Segundo esses autores, os enunciados falados são organizados hierarquicamente em

constituintes prosódicos que são: a sílaba, o pé, a palavra fonológica, o grupo clítico, a frase fonológica, a frase entoacional e o enunciado fonológico.

Neste trabalho, para o estudo de hipersegmentações, são relevantes informações dos constituintes prosódicos sílaba (s), pé métrico (Σ), palavra prosódica (ω) e grupo clítico (C). Iniciamos nossas considerações a partir da palavra prosódica, por ser a noção central para a análise de palavra escrita. Para o Português Brasileiro, Bisol (1996), seguindo o modelo de Nespor e Vogel (1986), assume que a palavra prosódica é o domínio em que se dá a interação entre os componentes fonológico e morfológico da gramática. A palavra prosódica caracteriza-se por ser portadora de um acento lexical ou acento primário. Um exemplo, segundo Tenani e Paranhos (2011), é a palavra morfológica “guarda-roupa” (hipossegmentada “guardaroupa”) que se constitui de duas palavras fonológicas [guarda] ω [roupa] ω , por haver dois acentos: um na primeira sílaba de “guarda” e outro na primeira sílaba de “roupa”. Na hierarquia prosódica de Nespor e Vogel (1986), a palavra prosódica é o constituinte imediatamente acima do pé métrico, podendo ser composta por um ou mais pés, e dominada pelo grupo clítico.

Acima da palavra na hierarquia prosódica, configura-se o grupo clítico que, de acordo com Bisol (2000), é formado por uma única palavra de conteúdo acompanhada de clíticos (palavras funcionais átonas, tais como artigos, preposições, conjunções). Exemplo da configuração de grupos clíticos é “ajude-me” (hipossegmentado “ajudime”), onde o clítico “me” é anexado ao hospedeiro “ajude” (palavra prosódica que têm acento).

Ainda segundo a hierarquia prosódica assumida, abaixo do constituinte palavra prosódica figura o pé métrico. O pé métrico constitui-se da relação de dominância/dominado entre duas ou mais sílabas. A depender da relação entre as sílabas, o pé métrico recebe denominação que segue a tradição das gramáticas greco-latinas: pés binários, com duas sílabas, em que o acento está à esquerda são chamados de *troqueus*, como em “casa”; pés binários com acento à direita são denominados *iambos*, como em “café”. Pés ternários, porque constituídos por três sílabas, são denominados dátilos, como em “árvore”.

Como constituinte basilar da hierarquia prosódica encontra-se a sílaba, definida por Bisol (1996), por conter um elemento cabeça (elemento forte), que, no caso do português, é sempre uma vogal e, em seu entorno, os elementos fracos, que são as consoantes e/ou glides, os quais são não obrigatórios em Português. Exemplos de sílabas do Português são: “é”, uma sílaba simples com apenas elemento nuclear; “fé”, uma sílaba canônica, com ataque e núcleo preenchidos,

respectivamente, por consoante e vogal; “mas”, uma sílaba complexa por apresentar ataque, núcleo e coda preenchidos; e “trens”, uma sílaba complexa por apresentar duas consoantes no ataque e duas consoantes na coda.

Além de características prosódicas, os pesquisadores do Grupo Estudos sobre a Linguagem (CNPq) têm defendido que somente essa natureza fonética–fonológica identificável nos dados não é suficiente para explicar todos os casos de segmentações não–convencionais de palavra. Chacon (2004) defende que as segmentações não–convencionais são marcas privilegiadas de observação da heterogeneidade constitutiva da escrita, noção defendida por Corrêa (2001). Para Corrêa (2001), o produto escrito (no nosso caso, as hipersegmentações) é o resultado da circulação do escrevente por práticas orais/faladas e letradas/escritas. Chacon (2004) salienta de Corrêa (2001) que não existem textos e/ou discursos que são essencialmente orais ou essencialmente escritos, todos são, na verdade, o resultado de um modo heterogêneo de constituição da escrita. Ainda segundo Chacon (2004), fatores que indicam o trânsito do escrevente por essas práticas orais e letradas são a ação da língua sobre o sujeito e dele sobre a língua, no que se refere ao papel do componente prosódico da linguagem na escrita, e a essa mesma ação recíproca língua/sujeito no que diz respeito à aquisição das convenções ortográficas do Português.

Dessa maneira, defendemos, neste trabalho, que a cada momento em que é chamado a produzir um texto, o escrevente estabelece relação entre características prosódicas e letradas dos enunciados. Essa relação não é dicotômica que identifica interferências da fala na escrita. Antes, essa relação, que buscamos explicitar, é de constituição, ou seja, as características dos enunciados falados constituem os enunciados escritos e vice-versa. Partilham conosco dessa perspectiva autores como Tenani (2009a, b, 2010, 2011a, b), Silva (2014), Chacon (2004, 2005), Capristano (2007a,b; 2004). Para esses autores, a complexidade dos dados de segmentação de palavras se dá em razão de trazerem quer características da oralidade, quer características de um processo de letramento, evidenciando a multifacetada relação do sujeito com a língua(gem).

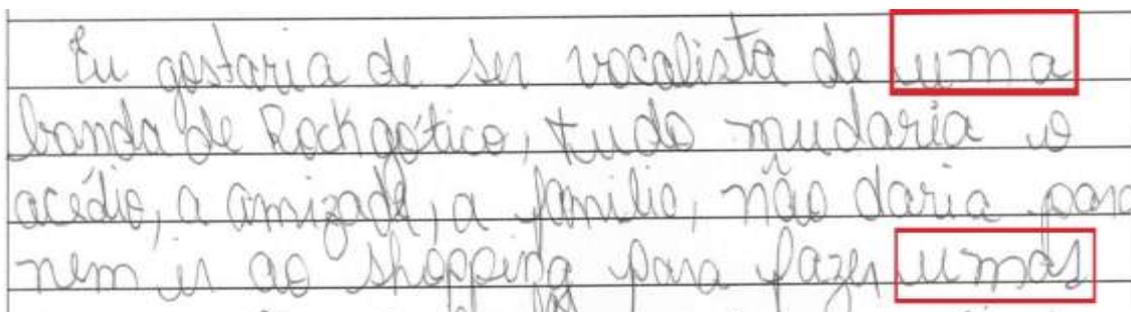
MATERIAL E QUESTÕES METODOLÓGICAS

A pesquisa foi desenvolvida com base em um corpúsculo de hipersegmentações de palavras constituído a partir de uma amostra longitudinal de textos que pertencem ao Banco de Dados de Escrita do EF II, IBILCE/UNESP (gratuitamente disponível em: <http://www.convenios.grupogbd.com/redacoes/Login>). A amostra longitudinal no referido banco

é composta de 2.495 textos produzidos por 291 alunos, dos quais 116 produziram textos em todos os anos letivos. Dentre esses 116, 19 alunos produziram entre 100% e 95% dos textos coletados em cada ano letivo (equivalente a 25 a 26 textos/aluno), sendo que 5 são do sexo masculino e 14 do sexo feminino.⁴ Essa subamostra é composta de 484 textos⁵, cuja distribuição ao longo dos anos letivos permitiu um estudo longitudinal das hipersegmentações que serão apresentadas a seguir.

Para a identificação das hipersegmentações, fizeram-se necessários procedimentos metodológicos que diz respeito à categorização do espaço em branco. Com base em Tenani (2011a), foram estabelecidos dois critérios para a identificação de segmentação não-convencional de palavra, a saber: (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. Abaixo, seguem exemplos de textos que têm diferentes tipos de grafias não-convencionais que geram dúvidas quanto à identificação de dados.

Figura 1. Problema de categorização da fronteira gráfica: “uma”



Fonte: Banco de Dados de Escrita do Ensino Fundamental II (Z10_7A_31M_02)

Na Figura 1, a grafia de “uma” pode ser, em princípio, um dado de hipersegmentação por haver um espaço em branco entre “um” e “a”. No entanto, ao observar as grafias das palavras em todo o texto, notamos que o aluno procura escrever até o limite da margem direita, ocupando o espaço disponível da linha para escrever. A partir dessa observação, supõe-se que tenha havido uma espacialização do “a” em relação à “um” da palavra “uma” para preencher toda a linha e para evitar a necessidade de escrever outra palavra no espaço que sobrasse, exigindo, talvez, uma

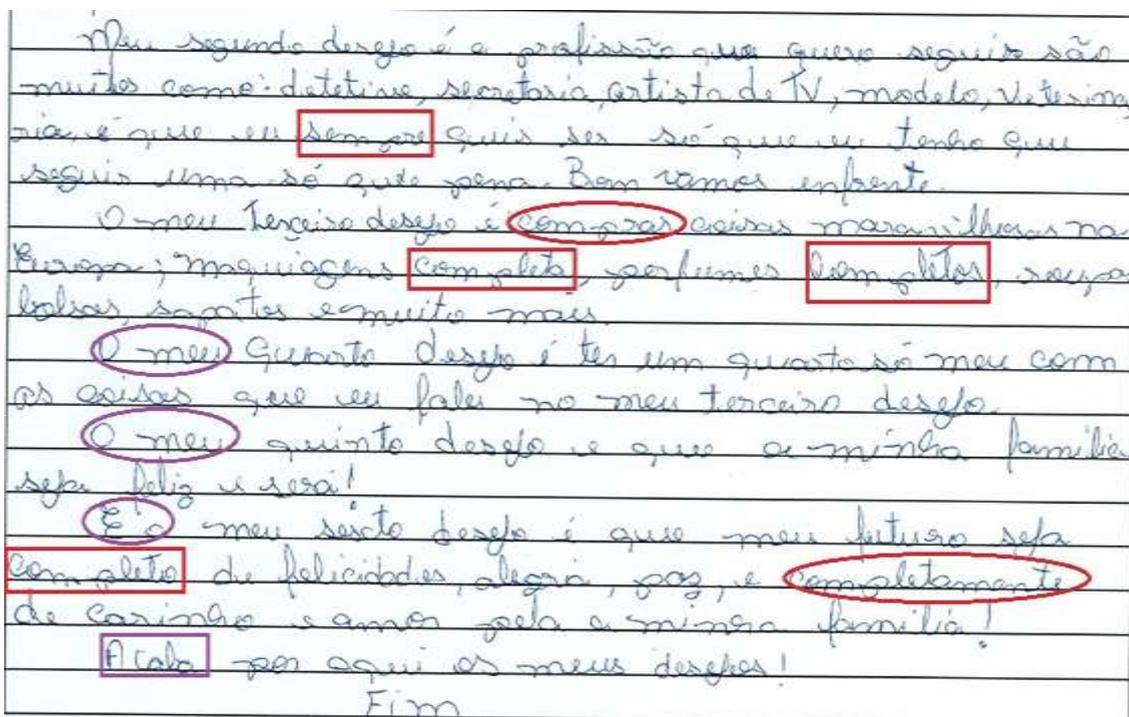
⁴ Neste artigo, desprezamos a variável gênero/sexo como fator de análise dos resultados.

⁵ Esse total de textos foi obtido a partir da soma do produto de textos X alunos. Definidos em 25 e 26 os números de textos, identificamos, no banco, o número de alunos que produziram esse total de textos e realizamos a seguinte operação: $(25 \times 10) + (26 \times 9) = 484$.

translineação. Notamos também que na palavra “umas”, logo abaixo da primeira grafia, é possível observar que a grafia das letras <m> e <a> não possuem ligaduras, mesmo quando escrita sem o espaço em branco. Concluímos, então, que este caso trata-se de caligrafia do sujeito e não um dado de segmentação não-convencional de palavra.

Na Figura 2, exemplificamos como proceder diante de grafias duvidosas como as das palavras “sem pre”, “com pleta”, “com pletos” e “com pleto” (grifadas com um retângulo), que foram classificadas, nesta pesquisa, como dados de hipersegmentações de palavras. Para esta classificação, tomamos como critério a comparação dos traçados das palavras que possuíam a mesma sequência <M + P> das palavras em questão ao longo do texto e a distribuição do espaço em branco entre essas palavras. Destacamos, também, as palavras “completamente” e “compras” (grifadas com um círculo). Nessas palavras, é possível notar que o espaço entre a letra <M> e a letra <P> é bem menor que o espaço entre as outras palavras destacadas por meio de um retângulo. Em função dessa comparação visual, consideramos apenas “sem pre”, “com pleta”, “com pletos” e “com pleto” como dados de hipersegmentações de palavras escritas.

Figura 2. Problema na categorização da fronteira gráfica: “A caba”



Fonte: Banco de Dados de Escrita do Ensino Fundamental II (Z08_5C_39F_03)

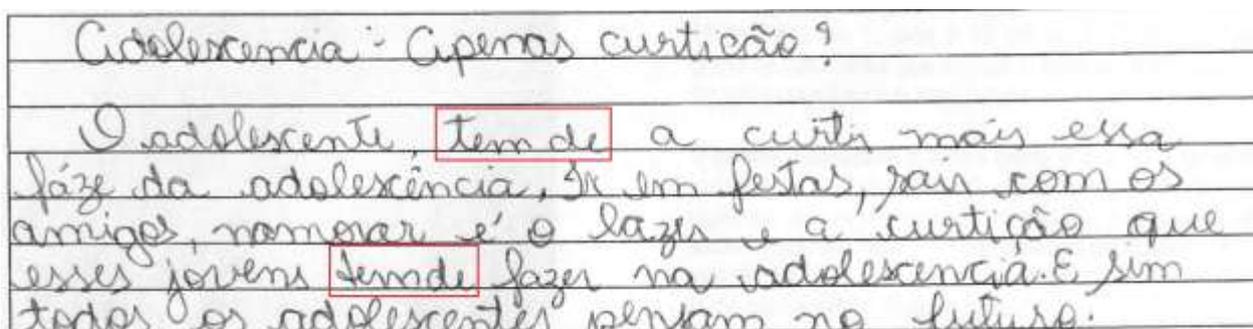
Ainda na Figura 2, a palavra “A caba” (destacada por um retângulo) que inicia um parágrafo é passível de ser um dado de segmentação não-convencional de palavra. Para definirmos se seria Linguagem, São Carlos, v. 26 (2): 2016.

A palavra “encherg ando” também é caso semelhante, uma vez que, após observar o traçado da letra <A>, como em “uma” (destacada por um retângulo), verificou-se que essa letra é traçada muitas vezes, pelo escrevente, sem a ligadura. Trata-se também de uma característica da escrita desse sujeito em questão.

Além do cuidado metodológico na identificação do dado, são adotados dois critérios para a classificação das hipersegmentações de palavras: (1) critério gráfico – que distingue a presença não-convencional do branco (“em bora”) ou do hífen (“chama-da”) nas hipersegmentações –, e (2) critério morfossemântico – que considera o fato de haver homônimas, como “em fim” (quando previsto “enfim”), cujas grafias levam à identificação de palavra e/ou locução. Esses critérios são importantes para tomar decisões metodológicas na identificação e classificação do dado de escrita.

No caso da identificação de palavras homônimas, consideramos todo o enunciado em que a grafia ocorreu para verificarmos qual significado da palavra é privilegiado pelo escrevente. Na Figura 5 abaixo, a ocorrência “tem de” pode ser, em um primeiro momento, considerada grafia convencional, já que parece se tratar da forma verbal “tem” seguida da preposição “de”. Porém, ao analisarmos o enunciado, identificamos que o escrevente quis grafar a forma “tende” do verbo “tender” para explicar, de acordo com a proposta da redação, que os adolescentes *tendem* a curtir mais essa fase da adolescência. Dessa forma, essa grafia foi considerada um caso de hipersegmentação de sequência homófona a “tem de”.

Figura 5: caracterização das homônimas “tende” e “tem de”



Fonte: Banco de Dados de Escrita do Ensino Fundamental II (Z11_8C_31M_07)

A descrição dessas decisões metodológicas adotadas se baseia em Tenani (2011b) e exemplifica o tipo de dificuldade encontrada na categorização do espaço em branco entre palavras, mostrando o tipo de olhar necessário para a identificação dos dados.

Quanto à análise prosódica das hipersegmentações, foi tomada a presença não-convencional do branco ou do hífen como marca gráfica motivada pela presença de fronteira de constituinte prosódico, como exemplificado no quadro abaixo. A identificação do constituinte prosódico, como palavra fonológica (ω), pé métrico (Σ), sílaba (σ), abaixo indicados, é feita com base nos algoritmos de formação desses constituintes, consideradas as análises feitas para o Português Brasileiro por BISOL (2005, 2004).

Quadro 1. Exemplos de análise em estruturas prosódicas de hipersegmentações

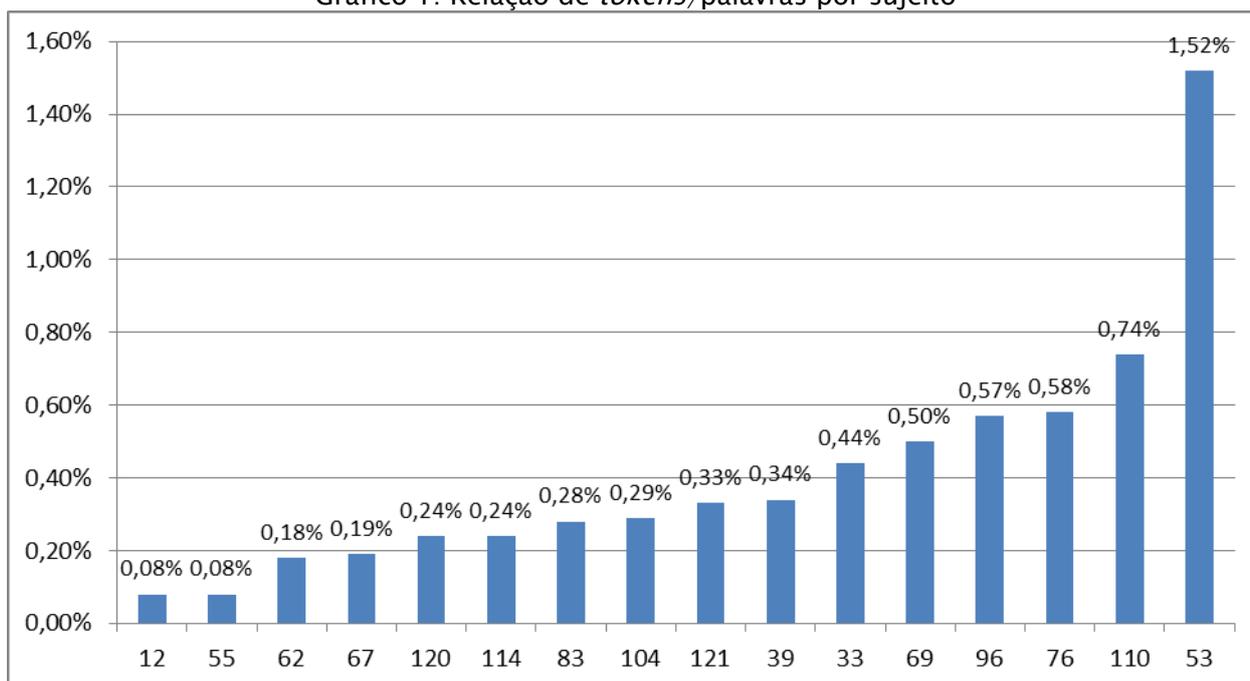
Estruturas prosódicas	Hipersegmentações
$\omega > \Sigma + \sigma$	“gostava” > “gosta va”
$\omega > \sigma + \Sigma$	“importa” > “im porta”
$\omega + \omega > \Sigma + \Sigma$	“antifumo” > “anti fumo”
$\omega > \Sigma + \Sigma$	“simplesmente” > “simples mente”

Por meio desses passos metodológicos, visamos a obter evidências de como se dá a relação do escrevente do EF-II com a língua(gem). Mais especificamente, neste trabalho, delimitou-se a busca por evidências de características dos enunciados falados que se mostram mobilizadas pelos escreventes em seus textos escritos.

RESULTADOS

Os 19 sujeitos selecionados para esta pesquisa produziram um total de 235 dados, incluindo ocorrências não-convencionais de branco, como “com migo”, e de hífen, como “arranja-se”, além de ocorrências que envolvem ausência do hífen, como “ex genro”, que deveria ser grafado com hífen, de acordo com as convenções ortográficas vigentes à época da produção escrita. Abaixo, segue gráfico por meio do qual se visualiza o resultado da razão entre o total de dado por sujeito pelo total de palavras que cada aluno escreveu durante os quatro anos do EFII.

Gráfico 1: Relação de *tokens*/palavras por sujeito



Legenda: No eixo horizontal, sujeitos identificados conforme o banco de dados. No eixo vertical, taxa de segmentação não-convencional de palavra escrita.

De modo geral, a média desse grupo de alunos é 0,32% de segmentações não-convencionais, havendo apenas um aluno que não tem hipersegmentação em seus textos. Dentre os 19 alunos que têm dados, dois alunos apresentam taxa entre 0,74% e 1,52%, portanto, quase quatro vezes mais ocorrências do que a média do grupo. Excluídos do cálculo esses dois alunos, a média passa a 0,29% de ocorrências, a qual tomamos como parâmetro para identificar grupos de sujeitos em relação às taxas identificadas. Observamos que há:

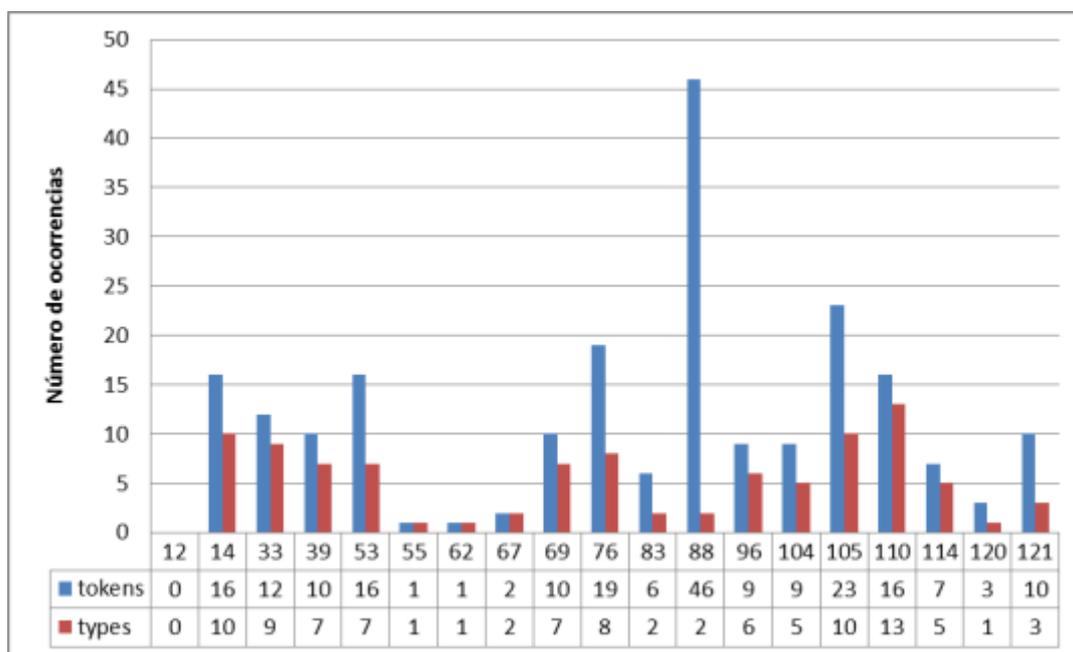
1. 05/17 (29,4%) sujeitos com taxas que variam de 0,43% a 0,59%; portanto, apresentam de 1,5 a 2 vezes dados acima da média do grupo;
2. 08/17 (47%) sujeitos com taxas que variam de 0,17% a 0,37%, estando próximos da média do grupo, seja para mais ou para menos;
3. 04/17 (23,5%) sujeitos com taxas que variam de 0,04% a 0,10%, estando 0,34 vezes abaixo da média do grupo.

Portanto, na amostra que compõe o material pesquisado, predominam sujeitos (70,5%) com taxas relativamente baixas (0,04% a 0,37%) de segmentações não-convencionais de palavras ao longo do Ensino Fundamental II. Para a melhor caracterização desses dados, optamos por organizá-los em função do tipo de registro gráfico da fronteira de palavra, a saber: (i) 216 (91,9%)

dados com presença não-convencional do branco; (ii) 11 (4,66%) dados com presença não-convencional do hífen; (iii) 08 (3,38%) dados com ausência não-convencional do hífen envolvendo palavras compostas. Neste artigo, restringimos a análise às segmentações não-convencionais por branco, por essas serem predominantes no conjunto de dados e porque a complexidade de todos os dados demandaria uma análise que a dimensão deste artigo.

Dentre o conjunto de segmentações não-convencionais de palavras pela presença do branco, faz-se necessário observar a alta frequência de certos itens gramaticais hipersegmentados, a saber: “porque”, “senão”, “demais” e “naquele(a, s)”, “aquele(a,s)”, em relação aos dados identificados.⁶ A alta frequência de ocorrências (*tokens*) desse conjunto de dados envia os resultados quantitativos dos sujeitos analisados, como pode ser visto no Gráfico 2. A relação entre *types* (palavra morfológica) e *tokens* (ocorrências da mesma palavra morfológica), apresentada no Gráfico 2, permite observar que há vários sujeitos que têm muitas ocorrências de uma mesma palavra morfológica grafadas não convencionalmente (muitos *tokens* de um mesmo *type*), como é o caso do sujeito com 88, com 47 ocorrências de *porque* hipersegmentado.

Gráfico 2. Relação *types/tokens* para hipersegmentações por branco



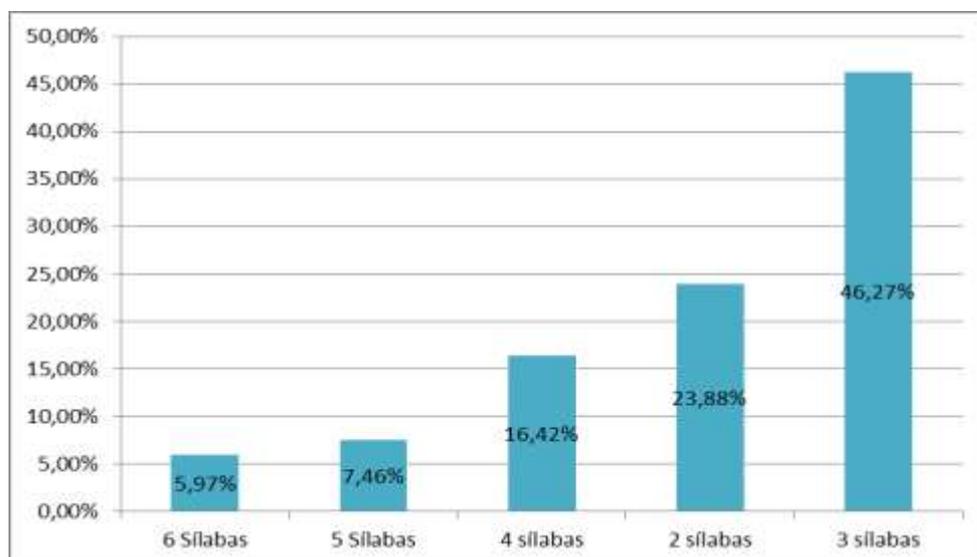
⁶ Em relação às grafias de “senão”, Tenani e Longhin (2015) descrevem a complexidade linguística que o particulariza em relação aos demais dados de hipersegmentação. As grafias não-convencionais de *senão* trazem, segundo as autoras, marcas da constituição histórica da palavra *senão* (uma junção das palavras “se” e “não”) e também mobilizam uma rede polissêmica diversificada, com sentidos de condição, contraste e alternância. Além disso, os dados de *senão* mobilizam uma análise prosódica que vai além da palavra envolvendo a prosodização de orações em enunciados. Em relação à “porque” hipersegmentado, Soncin e Longin (2014) também descrevem a complexidade linguística que o particulariza e discutem sobre a complexidade morfosintática mobilizada nos enunciados escritos pelos alunos do Ensino Fundamental. Estão em curso estudos específicos sobre as ocorrências de “demais” hipersegmentado por Tenani e Oliveira (inédito) e sobre “naquele(a,s)” e “aquele(a,s)” hipersegmentados por Tenani (inédito).

A partir desse levantamento, obtivemos os seguintes resultados gerais: 99 *types* em 216 *tokens* para 18 sujeitos. Esses dados permitem estabelecer algumas relações gerais, a saber: a razão entre *tokens*/sujeitos é de 27,7; enquanto que a razão entre *types*/sujeitos é de 5,5. Esses resultados sinalizam que, de modo geral, os alunos do Ensino Fundamental II grafam não convencionalmente muitas vezes as mesmas palavras, como já observado anteriormente para o sujeito 88: era um sujeito que estava muito acima da média, em relação aos seus pares, quanto à razão entre número de segmentações não-convencionais e número de palavras escritas, mas a razão entre *types* e *tokens* permite constatar que se trata de um dos sujeitos que tem menos hipersegmentações de diferentes palavras, não chegando a dois o total de palavras morfológicas hipersegmentadas. A característica desse sujeito contrasta com a do sujeito 110: é o sujeito com o maior número de palavras hipersegmentadas (13 diferentes palavras morfológicas) no conjunto de sujeitos, apresentando, também, alta taxa de segmentação não-convencional no total de palavras que escreveu. Essa observação da relação de taxa de *types* e *tokens* leva-nos a chamar a atenção para o fato de ser relevante considerar não apenas número de ocorrências de palavras grafadas não convencionalmente, mas principalmente informação sobre quais são essas palavras.

Feitas essas considerações mais gerais sobre *types* e *tokens*, passamos à análise de características prosódicas de 67 dados de hipersegmentações com a presença não-convencional do espaço em branco. Lembramos que esse conjunto de dados foi definido a partir da exclusão das ocorrências “porque”, “senão”, “demais” e “naquele(a, s)”, “aquele(a,s)”.

A seguir trazemos um gráfico no qual podemos vislumbrar que palavras, de acordo com o seu número de sílaba, são hipersegmentadas no cópuz estudado.

Gráfico 3. Porcentagem de hipersegmentações em função do número de sílabas das palavras



A consideração do número de sílabas de palavras permite observar em que medida o tamanho da palavra é um fator relevante para descrever característica das hipersegmentações no EF. De forma geral, observamos, do gráfico acima, que há mais hipersegmentação de palavras dissílabas e trissílabas, uma vez que esse conjunto soma 70,15% (47/67), ou seja, mais da metade dos dados. Para os dissílabos, que perfazem 23,8% (16/67) dos dados, há uma predominância em hipersegmentar palavras constituídas de pés iâmbicos, como “em fim”. Para os trissílabos, que possuem um total de 46,2 % (31/67), há tendência em hipersegmentar a primeira sílaba pretônica, como “com tigo”. Já para os polissílabos, que somam menos da metade dos dados, 29,8% (20/67), há tendência em hipersegmentar as duas primeiras sílabas que formam um pé métrico, como em “simples mente”.

Abaixo, segue quadro com exemplos organizados por tipos (coluna 1), de acordo com configurações possíveis de fronteiras prosódicas das grafias hipersegmentadas (coluna 3), tendo como ponto de partida a configuração métrica da palavra convencional (coluna 2). Dois dados, a saber “ante penúltimo” e “viti mas”, são excluídos desse quadro por se tratarem as únicas palavras com pés dátilos. Portanto, no Quadro 2, apresentamos as hipersegmentações que envolvem palavras com pés iâmbico e troqueu. Em seguida, números e porcentagens de hipersegmentação são apresentados na Tabela 3, tendo em consideração os anos letivos em que ocorrem.

Quadro 2. Exemplos de hipersegmentações e estruturas prosódicas possíveis

Tipos	Estrutura rítmica da palavra	Fronteiras prosódicas possíveis	Exemplos
1	(● *)	(●)σ (*)σ	<i>em fim</i>
2	(* ●)	(*)σ (●)σ	<i>tem de</i>
3.1	(● * ●)	(●)σ (* ●)Σ	<i>com migo</i>
3.2		(● *)Σ (●)σ	<i>abri se</i>
4.1	(* ● *)	(●)σ (● *)Σ	<i>a final</i>
5.1	(* ● * ●)	(* ●)Σ (* ●)Σ	<i>simples mente</i>
5.2		(*)σ (● * ●)Σ	<i>ver dadeira</i>
5.3		(* ● *)Σ (●)σ	<i>começa mos</i>
6.1	(* ● ● *) ou (● * ● *)	(*)σ (● ● *)Σ ou (●)σ (* ● *)Σ	<i>de coração</i>
6.2		(*●)Σ (● *)Σ ou (●*)Σ (● *)Σ	<i>acon teceu</i>

7.1	(● * ● * ●) ou (* ● ● * ●)	(● *)Σ (●*●)Σ ou (*●)Σ (●*●)Σ	<i>deses perado</i>
7.2		(●)σ (* ● * ●)Σ ou (*)σ (● ● * ●)Σ	<i>com pletamente</i>
7.3		(● * ●)Σ (* ●)Σ	<i>espaço nave</i>
8.1	(* ● * ● * ●) ou (● * ● ● * ●)	(* ● * ●)Σ (* ●)Σ ou (● * ● ●)Σ (* ●)Σ	<i>absoluta mente</i>
8.2		(* ●)Σ (* ●* ●)Σ ou (● *)Σ (● ● * ●)Σ	<i>tele transportado</i>

Onde: * para sílaba tônica; ● para sílaba átona; σ para palavra prosódica; Σ para pé métrico; parênteses para fronteira prosódica.

Tabela 03. Número e porcentagem de hipersegmentações por ano letivo

Tipos	5ª série	6ª série	7ª série	8ª série	Nº totais	%
1.1	5	3	1	1	10	14,93%
2.1	5	1	0	1	6	8,96%
3.1	11	6	7	6	26	38,81%
3.2	0	0	1	1	2	2,99%
4.1	1	1	0	1	3	4,48%
5.1	0	1	2	3	6	8,96%
5.2	0	0	0	1	1	1,49%
5.3	0	1	0	0	1	1,49%
6.1	1	1	0	0	2	2,99%
6.2	0	0	1	0	1	1,49%
7.1	0	0	0	1	1	1,49%
7.2	1	1	0	1	3	4,48%
7.3	1	0	0	0	1	1,49%
8.1	0	0	1	0	1	1,49%
8.2	1	0	0	0	1	1,49%
Outro	1	0	0	1	2	2,99%
Total	27	15	13	17	67	100,00%

De acordo com a Tabela 3, no cópuz analisado, trissílabos são os mais hipersegmentados. Em sua maioria, a primeira sílaba dessas hipersegmentações coincide com uma palavra da língua, como uma preposição em “com migo”. Dessa forma, nossos resultados apontam para o fato de o aluno interpretar a sílaba pretônica dessas palavras trissílabas como sendo um clítico monossilábico. Alguns exemplos da tendência observada são as hipersegmentações: “com tigo” (contigo), “a miga” (amiga), “de pende” (depende), “em quanto” (enquanto). É raro encontrar palavras trissílabas com a última sílaba hipersegmentada, como “abri se”. Nos dois tipos de dados, porém, a sílaba que fica isolada entre espaços em branco coincide com possíveis itens gramaticais

átomos e monossilábicos. Nos dados do tipo 3.1, a hipersegmentação resulta em sílaba seguida de pé métrico; no tipo 3.2, a hipersegmentação resulta em pé métrico seguido de sílaba. Constatase, portanto, que o branco é inserido quer por motivos morfossintáticos, na medida em que predomina segmentação de sílaba como se fossem possíveis itens gramaticais da língua, quer por motivos prosódicos, na medida em que a fronteira de pé métrico é respeitada em todos os casos.

Essa tendência à configuração métrica de sílaba seguida de pé métrico para as hipersegmentações entre palavras com três ou mais sílabas foi atestada em 52,24% dos dados e recorrente em todos os anos do EF.

Um conjunto de 16,42% das grafias configura sequência de dois ou mais pés métricos, como em “acon teceu”. Ainda nesse conjunto menor de dados, há grafias, como “simples mente”, nas quais informações morfossintáticas também se mostram como potencialmente motivadoras (como relação entre a base “simples” e o sufixo “mente”), ao lado de fronteira de pés métricos, a saber: (acon)Σ (teceu)Σ e (simples)Σ (mente)Σ.

Essa forte confluência de fatores prosódicos e morfossintáticos na motivação para hipersegmentações também é identificada nos dissílabos hipersegmentados (que chegam a 23,81% dos dados): quer dissílabos trocaicos (dados do tipo 1), quer dissílabos iâmbicos (dados do tipo 2) apresentam uma sílaba átona que coincide com possíveis monossílabos gramaticais. Os exemplos “em fim” (enfim) e “tem de” (tende), apresentados no Quadro 2, permitem constatar que o reconhecimento de possíveis palavras da língua gera a dúvida ortográfica: “enfim” ou “em fim”? E ainda “tende” ou “tem de”? Verifica-se, nesses registros não-convencionais, que o estudante conhece possibilidades de representações ortográficas de mesma sequência fônica, mas desconhece qual das grafias deve empregar segundo a convenção para as sequências que produz. A ancoragem somente em características prosódicas dessas sequências não fornece pistas de qual grafia o estudante deve empregar, pois a prosodização dos clíticos torna semelhantes palavras e sequência de palavras. A definição de qual grafia escolher passa, na maioria dos casos identificados – exemplificados no Quadro 2 –, por compreender os aspectos morfossemânticos e morfossintáticos envolvidos nos usos de palavras e expressões da língua.

CONSIDERAÇÕES FINAIS

Resumidamente, demonstramos que estudantes que cursaram os quatro últimos anos do EF (do sexto ano ao nono ano) ainda apresentam dificuldades na identificação do que seja uma

palavra ortográfica e por consequência na colocação dos espaços em branco nos textos que escrevem. Mostramos também que a colocação não-convencional desses espaços segue uma tendência, pois, em sua maioria, esses espaços em branco delimitam estruturas prosódicas, como sílaba e pé métrico ou pé métrico e sílaba e, ainda, sequência de pés métricos. Foi possível observar também, em termos quantitativos, que a palavra mais hipersegmentada é aquela composta com por três sílabas.

Em trabalho semelhante a este, porém com crianças em início de aquisição de escrita, Chacon (2005) afirma que é possível enxergar fatos prosódicos mais interessantes em trissílabos do que em dissílabos e polissílabos hipersegmentados. O autor encontrou poucos dados de polissílabos, como ocorreu no cópulus do EF II que analisamos. Quanto aos dissílabos hipersegmentados, esses foram quantitativamente mais frequentes do que os trissílabos, mas os dissílabos, no que se refere à ação da prosódia, mostram variação pouca significativa, dada a presença do limite gráfico ocorrer entre as duas sílabas das palavras, mostrando que apenas a sílaba foi o constituinte prosódico relevante nessas ocorrências.

Segundo o autor, os trissílabos são interessantes, porque qualquer ponto pode ser contexto para uma possível separação não-convencional da palavra. Porém, segundo o autor, a colocação do espaço em branco não foi aleatória no cópulus de EF I; pelo contrário: o modo como os espaços em branco foram inseridos mostra que princípios da oralidade foram seguidos, como, “é o caso do *ritmo*, que, já estruturado na língua, marca-se na oralidade, por exemplo, por meio de contrastes entre sílabas acentuadas e não-acentuadas” (CHACON, 2005).

Chacon (2005) explica que esses contrastes se marcam na língua por uma relação do tipo dominante/dominado que, como já explicamos, define o constituinte prosódico pé métrico. Nos dados do autor, a característica rítmica é mais evidente nos trissílabos, já que 81,62% dos dados que analisou apresentaram a ruptura entre os limites de sílabas e de pés. Essa mesma tendência é reportada nos resultados ora descritos para o EFII, o que sugere que alunos do EF I e EF II se ancoram fortemente em características rítmicas da oralidade quando produzem seus textos escritos.

Dentre o conjunto de trissílabos hipersegmentados, Chacon (2005) relata que: (i) 82,88% deles se tratavam da combinação de uma sílaba e um pé métrico, como em “e Rita” (*irrita*); e (ii) 17,12% geraram a combinação contrária, a de um pé métrico e uma sílaba, como em “colo que” (*coloque*).

Nos dados do EF II ora analisados, encontramos uma configuração parecida com a descrita para o EF I por Chacon (2005), embora com percentuais distintos. Em 100% (35/35) dos trissílabos identificados, o ponto de corte pelo branco foi entre limites de sílabas e pés métricos da palavra, sendo que: (i) 94,2% (33/35) se tratam de uma sílaba e um pé, como em “em bora” (embora); (ii) 5,71% (2/35) se tratam da combinação contrária, um pé e uma sílaba, como em “gosta va” (gostava). A tendência comum entre alunos dos anos iniciais e finais do EF para hipersegmentar trissílabos é, pois, a combinação sílaba e pé métrico e são aqui interpretadas como evidência da característica prosódica predominante do português em que monossílabos átonos são prosodizados à direita de seu hospedeiro, resultando na sequência de três sílabas: $\sigma'\sigma$.

Como já explicitado anteriormente, para Chacon (2004, 2005), os dados de hipersegmentação – com como os de hipossegmentações – são efeitos não somente de relações prosódicas entre sílabas, mas da complexa relação do sujeito com a linguagem, evidenciando a circulação do sujeito por práticas letradas/escritas. Dessa forma, visando não deixar de contemplar essa complexidade, Chacon passa a analisar, em um segundo momento do seu trabalho, as características dos dados que indiciam essa circulação. O autor mostra que quando hipersegmentam os trissílabos, em qualquer um dos pontos descritos por ele (pé + sílaba ou sílaba + pé), os escreventes dão indícios de que uma parte da palavra pode ter sido reconhecida como uma possível palavra da língua, as quais, na maioria das vezes, são correspondentes a categorias gramaticais, como em “**com** migo”. A estrutura “com” pode ter sido interpretada como uma preposição da língua, uma pista é a grafia “com” com “m” ao invés da letra “n”, como previsto na convenção da palavra “comigo” (cf. discussão detalhada sobre essa temática em Tenani, 2010).

Segundo o autor, quando não é identificado um monossílabo, o reconhecimento pode ocorrer com o dissílabo, como se observa nos exemplos do EFII: as possíveis palavras “gosta” e “abri” a partir da hipersegmentações “gosta va” (gostava) e “abri se” (abrisse), respectivamente.

Baseados nos resultados e discussões feitos neste artigo, podemos afirmar que dados de hipersegmentação do EF II, tal como os de EF I, trazem tanto características da oralidade quanto características de um processo de letramento, evidenciando a multifacetada relação do sujeito com a linguagem.

Concluimos enfatizando que este trabalho busca contribuir tanto para os estudos sobre segmentação de palavra, quanto para os estudos sobre ensino da escrita. Ao considerarmos textos do segundo ciclo do EF II, ampliamos a reflexão sobre a aquisição da escrita para um conjunto de

alunos pouco estudados, já que os estudos sobre segmentações se concentram em dados de anos iniciais do EF, e demonstramos que não se constata o que é esperado do escrevente nesse período de escolarização, pois em sua produção textual não se observa “domínio *da separação* em palavras, estabilidade de palavras de ortografia regular e de irregulares mais frequentes na escrita e utilização de recursos do sistema de pontuação para dividir o texto em frases” (PCN, 1997, p. 80).

Dessa forma, por meio da análise dos dados do EF II e da comparação com dados do EF I, mostramos que escreventes que já concluíram, em termos formais, o período dos anos iniciais da alfabetização em ambiente escolar continuam a ter dificuldades semelhantes às de alunos de anos iniciais de alfabetização, no que diz respeito à colocação do branco entre as palavras.

AGRADECIMENTOS: ao CNPq, pela bolsa de Produtividade Pesquisa (CNPq 309872/2012-0), concedida ao primeiro autor, e pela bolsa PIBIC (Proc. 2014/29329), concedida ao segundo autor do artigo; à FAPESP pela concessão de auxílio à pesquisa (Proc. 2013/14546-5) ao primeiro autor.

REFERÊNCIAS

ABAURRE, M. B. M. Hipóteses iniciais de escrita: evidências da percepção, pela criança, de unidades rítmico/ entonacionais na fala. In: IV Encontro Nacional da ANPOLL, 1989, Recife. *Anais do IV Encontro Nacional da ANPOLL*. Recife: UFPE, 1989. p. 751-764

_____. A relevância dos critérios prosódicos e semânticos na elaboração de hipóteses sobre segmentação na escrita inicial. *Boletim da Abralin*, Campinas, v. 11, p. 203-17, 1991.

_____. Os Estudos Linguísticos e A Aquisição da Escrita. In: M. F. C. P.de CASTRO. (Org.). *O método e o dado no estudo da linguagem*. 1ª ed. Campinas, SP: Editora da Unicamp, 1996, v. 1, p. 111-163.

_____. 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*. Campinas: Pontes Editores, 1998, p.135-142.

ABAURRE, M. B. M.; CAGLIARI, L. C. Textos espontâneos na primeira série: evidências da utilização, pela criança, de sua percepção fonética da fala para segmentar a escrita. *Jornal da Alfabetizadora*, v. 9, 1995.

ABAURRE, M. B. M.; FIAD, R. S.; MAYRINK-SABINSON, M. L. T. *Cenas de Aquisição da Escrita: O Trabalho do Sujeito Com O Texto*. 1ª. ed. Campinas, SP: Mercado de Letras, 1997.

ABAURRE, M. B. M.; SILVA, A. O desenvolvimento de critérios de segmentação na escrita. *Temas em psicologia*. São Paulo, v. 1, 1993, p. 89–102.

BISOL, L. Constituintes prosódicos. In: _____. *Introdução a estudos de fonologia do português brasileiro*. Porto Alegre: EDIPUCRS, 1996, p. 247–261.

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

_____. *Mattoso Câmara Jr. e a Palavra Prosódica*. DELTA. Documentação de Estudos em Lingüística Teórica e Aplicada, São Paulo, v. 20, n. especial, p. 59–70, 2004.

_____. *O clítico e seu hospedeiro*. Letras de Hoje, EDIPUCRS, v. 40, n.3, p. 163–184, 2005.

BRASIL. Secretaria de Educação Fundamental. *Parâmetros Curriculares Nacionais: língua portuguesa*. Brasília: 1997.

CAPRISTANO, C. C. *Mudanças na trajetória da criança em direção à palavra escrita*. 2007. 263 f. Tese (Doutorado em Linguística Aplicada) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Campinas, 2007a.

_____. *Aspectos de segmentação na escrita infantil*. São Paulo: Martins Fontes, 2007b.

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

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

_____. Hipersegmentações na escrita infantil: entrelaçamento de práticas de oralidade e de letramento. *Estudos Linguísticos*, Campinas, v. 34, 2005, p. 77–86.

_____. 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.

CORREA, M. L. G. *O modo heterogêneo de constituição da escrita*. São Paulo: Martins Fontes, 2001.

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

CUNHA, A. P. N. ; A Influência dos Constituintes Prosódicos na Aquisição da Língua Escrita. *Caderno de Letras (UFPEL)*, Pelotas, v. 1, n.10, p. 14–23, 2004.

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

SILVA, L. M. *Um estudo longitudinal sobre as hipersegmentações de palavras escritas nos anos finais do ensino fundamental*, São José do Rio Preto, 2014. Dissertação (Mestrado em Estudos Linguagem, São Carlos, v. 26 (2): 2016.

Linguísticos) – Universidade Estadual Paulista.

SONCIN, G.; LONGHIN, S. R. A causalidade de porque em textos escolares: domínios de atuação, gêneros de produção. *Revista Linguagem & Ensino (Online)*, v. 17, p. 525–549, 2014.

TENANI, L. Entre o grupo clítico e a palavra fonológica: os erros de segmentação não-convencional de palavras. *VI Congresso Internacional da ABRALIN*, 2009a.

_____. A segmentação não-convencional de palavras: evidências de representações de relações entre enunciados falados e escritos. *Anais do II SIMELP: língua portuguesa: ultrapassar fronteiras, juntar culturas*, 2009b, p. 107–127.

_____. Sobre a relação entre constituintes prosódicos e a ortografia. In: IV Conferência Linguística e Cognição, 2007, Belo Horizonte. *Caderno de Programação e Resumos*, 2007. v. 01. p. 13–13.

_____. A segmentação não-convencional de palavras e a aquisição da escrita. *Palestra apresentada no II Ciclo de Palestras sobre aquisição da linguagem e escrita*. Faculdade de Educação, Universidade Federal de Pelotas, Pelotas. 2010a.

_____. A grafia dos erros de segmentação não-convencional de palavras. *Cadernos de Educação (UFPEl)*, v. 35, p. 247–269, 2010b.

_____. Letramento e segmentações não-convencionais de palavras. In: TFOUNI, Leda. *Letramento, escrita e leitura: questões contemporâneas*. Campinas: Mercado das Letras, 2011a.

_____. A segmentação não-convencional de palavras em textos do ciclo II do Ensino Fundamental. *Revista da ABRALIN*, v.10, n.2, jul./dez. 2011b, p. 91–119.

_____. Hipersegmentação de “aquele(a,s)” e “naquele(a,s)”: evidências de relações entre prosódia e escrita. (inédito).

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

TENANI, L; LONGHIN – THOMAZI, S. Flutuação gráfica entre 'senão?' e 'se não': considerações sobre polissemia, constituição morfossintática e prosódica. *Gragoatá (UFF)*, v. 20, p. 138–206, 2015.

TENANI, L; OLIVEIRA, G. Hipersegmentação de 'demais' em textos do Ensino Fundamental II. (inédito).

TENANI, L; PARANHOS, F. C. Análise Prosódica de segmentações não-convencionais de palavras em textos do sexto ano do EF. *Filologia e Linguística Portuguesa*, v. 13, p. 477–504, 2011.

Recebido em: 30/03/2016. Aceito em 20/07/2016.