

INVESTIGANDO A APRENDIZAGEM DA ESCRITA EM UM CORPUS DIGITAL DE LINGUAGEM INFANTIL

INVESTIGATING LEARNING WRITING IN A CORPUS DIGITAL CHILD LANGUAGE

Maria Luisa Fickelcherer Moraes¹
Raquel Márcia Fontes Martins²

RESUMO:

Este artigo trata de um estudo sobre a aquisição da escrita infantil, em específico, sobre a aprendizagem escrita de encontros consonantais tautossilábicos (encontros CCV-consoante-consoante-vogal) por alunos dos anos iniciais do Ensino Fundamental (FONTES-MARTINS, OLIVEIRA-GUIMARÃES, 2009). Foram focalizados quatro encontros CCV: BR, CR (mais frequentes) e GR, FR (menos frequentes). Realiza-se aqui uma discussão sobre a alfabetização, o processo de aquisição da escrita infantil, considerando a interferência de fenômenos da fala na escrita, nas variações ou erros ortográficos (OLIVEIRA, 2005; MORAIS, ALBUQUERQUE, LEAL, 2005). Este estudo tem como fundamentação teórica, principalmente o Modelo de Redes (Bybee, 2001). Esse modelo considera a frequência (de ocorrência e de tipo) como fundamental na emergência da gramática e do léxico. Como um Modelo Baseado no Uso (FERRARI, 2011), ele contribui para a discussão sobre a organização da informação linguística na mente humana, considerando a experiência de uso da língua. Para realizar o presente estudo, foram utilizados dados do corpus digital de linguagem infantil do Projeto e-Labore (CRISTÓFARO-SILVA ET AL., 2007). Os resultados deste trabalho indicam que o fator frequência de ocorrência é relevante para a criança que se encontra no processo de aquisição da escrita dos encontros CCV analisados neste trabalho.

PALAVRAS-CHAVES: *Corpora*; Frequência; Aquisição da Escrita Infantil.

ABSTRACT:

This article deals with a study on the acquisition of children's writing, in particular, learning about writing tautossyllabic consonant clusters (CCV - consonant-consonant-vowel) by students in the early years of elementary school (FONTES-MARTINS, OLIVEIRA-GUIMARÃES, 2009). BR, CR (more frequency) and GR, FR (less frequent): Four clusters were focused CCV. We proceed with a discussion about literacy, the acquisition process of children's writing, considering the interference of phenomena of speech in writing, in variations or misspellings (OLIVEIRA, 2005; MORAIS, ALBUQUERQUE, LOYAL, 2005). This study has as theoretical foundation, especially the Network Model (BYBEE, 2001). This model considers the frequency (of occurrence and type) as essential in the emergence of grammar and lexicon. As a Model Based on Usage (FERRARI, 2011), it contributes to the discussion on the organization of linguistic information in the human mind, considering the experience of language use. To perform this study, we used data from the digital corpus of child language and Project-Labore (CRISTÓFARO -SILVA ET AL., 2007). The results of this study indicate that the frequency of occurrence factor is relevant to the child who is in the process of writing acquisition of CCV encounters analyzed in this work.

KEYWORDS: *Corpora*, Frequency, Acquisition of Children's Writing.

¹ Graduada em Letras pela Universidade Federal de Lavras. Currículo: <http://lattes.cnpq.br/3119129441094879>.

² Doutora e mestra em Estudos Linguísticos e graduada em Letras pela Universidade Federal de Minas Gerais. Professora da Universidade Federal de Lavras. Currículo: <http://lattes.cnpq.br/3971924930410275>.

01 – INTRODUÇÃO

Este artigo tem como objetivo estudar um fenômeno que ocorre na escrita infantil, analisando a interferência da fala na aquisição da escrita (OLIVEIRA, 2005; MORAIS, ALBUQUERQUE, LEAL, 2005). Especificamente, este trabalho visa investigar a aprendizagem da escrita de encontros consonantais tautossilábicos (FONTES-MARTINS, OLIVEIRA-GUIMARÃES, 2009), ou seja, a representação gráfica de uma sequência de duas consoantes que ocorrem na mesma sílaba (sílabas consoante-consoante-vogal, CCV) por crianças do 1º ao 6º ano de escolaridade.

É relevante destacar que a pesquisa aqui realizada considera o fator frequência (BYBEE, 2001) na aprendizagem escrita dos encontros consonantais CCV em análise. Assim, este estudo tem como fundamentação teórica, principalmente o Modelo de Redes (BYBEE, 2001) que, como um Modelo Baseado no Uso (FERRARI, 2011), contribui para a discussão sobre a organização da informação linguística na mente humana, considerando a experiência de uso da língua.

Para realizar o presente estudo, utilizamos dados do *corpus* e-Labore – Laboratório Eletrônico de Oralidade e Escrita (CRISTÓFARO-SILVA ET AL., 2007)³. Esse *corpus* pertence a um projeto realizado pela Universidade Federal de Minas Gerais – UFMG, em um trabalho com escolas da cidade, em que são realizadas coletas de redações de crianças de 6 a 12 anos de idade que cursam o Ensino Fundamental. Assim, a pesquisa deste artigo foi realizada, utilizando-se dados do *corpus* do e-Labore, tendo como foco as variações encontradas em encontros CCV.

Na próxima seção, além de tratar do encontro consonantal em foco, discutem-se o processo de aprendizagem da escrita e o fator frequência, bem como aborda-se a área da Linguística de *Corpus* (SARDINHA, 2000, 2004). Em seguida, a metodologia deste trabalho é detalhada, abordando-se também o *corpus* de linguagem infantil e-Labore (CRISTÓFARO-SILVA, OLIVEIRA-GUIMARÃES, FONTES-MARTINS, ALMEIDA, 2008), utilizado neste estudo.

³ Maiores informações sobre o e-Labore podem ser obtidas em: <<http://www.projetoaspa.org/elabore/>>. Acesso em jun. 2014.

Na sequência, é realizada a análise dos dados do e-Labore aqui selecionados, bem como são discutidos os resultados deste estudo. Por fim, são feitas as considerações finais do trabalho.

02 – ENCONTROS CCV E SUA APRENDIZAGEM NA ESCRITA, CONSIDERANDO O FATOR FREQUÊNCIA E A LINGUÍSTICA DE *CORPUS*

Encontros consonantais tautossilábicos consistem de uma sequência de duas consoantes, uma obstruinte e uma líquida, que ocorrem em uma mesma sílaba, ou seja, trata-se de um encontro consonantal em uma sílaba consoante-consoante-vogal - CCV. Neste artigo chamaremos tal encontro de CCV.

A primeira consoante do encontro é a obstruinte que pode ser uma das seguintes consoantes [p, b, t, d, k, g, f, v]. Já a segunda consoante é líquida que pode ser a lateral [l] ou o tepe [r]. Segundo a autora Thais Cristófar (2000) no português brasileiro temos dois tipos de encontros CCV, o tautossilábico com as duas consoantes em uma mesma sílaba, como em “**B**rasil”, “**en**tro”. Outro caso de encontro CCV é chamado de heterossilábico (em sílabas diferentes), sendo que tal encontro é composto por uma consoante pós-vocálica adjacente à consoante da sílaba seguinte, como por exemplo, em “**gar**fo”, “**par**to”. O estudo apresentado neste artigo restringe-se em analisar a aquisição de encontros CCV em que a consoante líquida é o tepe [r]. Na metodologia, esclarecemos a opção por esse segmento sonoro.

Cabe aqui tratar da relevância deste estudo para a alfabetização de crianças, tendo como referência Oliveira (2005). Segundo Oliveira, o aprendiz é inteligente e naturalmente capaz de raciocinar, de fazer inferências e generalizações, sendo apto para criar regras e modelos próprios. O autor defende a importância de termos conhecimento de como a criança se comporta durante o processo de aprendizagem da escrita. Assim, esta pesquisa analisa possíveis equívocos cometidos por crianças durante o processo de aprendizagem da escrita, sem negligenciar o fato de que o aluno é inteligente e interage com o objeto de conhecimento, no caso a escrita (Oliveira, 2005, p. 4).

Revista Brasileira de Educação e Cultura – ISSN 2237-3098 Centro de Ensino Superior de São Gotardo http://periodicos.cesg.edu.br/index.php/educacaoecultura	Número XII Jul-Dez 2015 periodicoscesg@gmail.com	Trabalho 08 Páginas 125-142
--	--	--------------------------------

Desse modo, o presente estudo considera a psicogênese da língua escrita (FERREIRO E TEBEROSKY, 1999). Segundo Coutinho (2005), essa teoria tem caráter psicológico, abordando a maneira como os alunos se apropriam da escrita alfabética. A partir do momento em que tal teoria foi propagada no Brasil, ela possibilitou aos professores compreender os níveis de aquisição da escrita e, portanto, mudar a maneira de avaliar seus alunos (Coutinho, 2005).

Esta teoria é composta por quatro níveis: 1- pré-silábica: quando o aluno não distingue letra de desenho. 2 – silábica: quando o aluno entende que a palavra é escrita devido a quantidade de sílabas que esta é composta. 3 – Silábica-Alfabética: marca a transição da escrita do aluno, ocorre uma mistura das duas primeiras fases. 4 – Alfabética: quando o aluno possui domínio total da escrita. Para facilitar o entendimento desta teoria, podemos ilustrar, o nível silábico, através de um exemplo dado por Oliveira ao retratar erros comuns de alfabetizadores que, ao verem seu aluno escrever **pcr** para **procura**, entendem que a criança está "comendo" letra, quando, na verdade, está apenas escrevendo silabicamente. Como a palavra **procura** é composta por três sílabas, o aprendiz formula uma regra: representar a palavra também por três símbolos, ou melhor por três letras, fazendo uma associação o número de sílabas (e não o número de fonemas) e o número de letras (grafemas). Com o passar do tempo, este mesmo estudante irá alcançar a escrita alfabética, se tornando, enfim, capaz de compreender o princípio alfabético, segundo o qual, a cada fonema (e não a cada letra), escreve-se uma letra (OLIVEIRA, 2005). Dessa maneira, a palavra **procura** deve ser escrita com 7 letras por que é produzida com 7 fonemas.

É importante ressaltar que este trabalho dialoga com o estudo de Fontes Martins e Oliveira Guimarães (2009): "Efeitos de frequência na produção escrita de encontros consonantais". As autoras, nessa pesquisa, tendo por vista grande quantidade de dados, optaram por fazer um primeiro recorte, decidindo pela análise dos encontros consonantais tautossilábicos formados por consoante obstruinte e consoante líquida tepe [r]. Sendo necessário um segundo recorte de oito possibilidades de análise (tr, pr, br, cr, gr, fr, dr, vr), as autoras optaram pelos dois encontros mais frequentes (tr e pr) e pelos dois menos frequentes (dr e vr). No

Revista Brasileira de Educação e Cultura – ISSN 2237-3098 Centro de Ensino Superior de São Gotardo http://periodicos.cesg.edu.br/index.php/educacaoecultura	Número XII Jul-Dez 2015 periodicoscesg@gmail.com	Trabalho 08 Páginas 125-142
--	--	--------------------------------

presente estudo, é realizado semelhante recorte, contudo, serão analisados aqui os encontros consonantais tautossilábicos da consoante líquida tepe [r] não utilizados na pesquisa de Fontes-Martins e Oliveira-Guimarães (br, cr, gr, fr).

Há um grau de dificuldade na aquisição da escrita infantil de encontros CCV que podem ser considerados um desafio para o aluno que se encontra nesta fase. Existem estudos que correlacionam esta dificuldade à questão da variação de tais encontros na fala. Oliveira (2005), por exemplo, aborda três concepções de aprendizagem da escrita, sendo a terceira interessante porque

prevê que esta interação com a escrita seja intermediada pela oralidade, ou seja, por aquilo que o aprendiz já conhece sobre sua língua (NB: ele *falaa* língua) quando inicia seu processo de construção da escrita, em outras palavras, o conhecimento sobre a língua falada controla o processo de aprendizado da língua escrita (OLIVEIRA, 2005, p. 9).

Partindo dessa premissa, podemos afirmar que, no início da aprendizagem da escrita, a interferência da oralidade se relaciona ao erro ortográfico infantil, já que a criança apresenta um domínio sólido da fala quando chega à escola para aprender a escrita (Cagliari, 1998). A linguagem falada é a primeira forma de socialização da criança. A socialização através da fala pode ocorrer também de forma implícita, por meio de participação em interações verbais que têm marcações sutis de papéis e status (ELY & GLEASON, 1996). No entanto, Oliveira acrescenta em seu estudo que, em algum momento de nossas vidas, somos capazes de tratar a escrita de maneira autônoma e separada da oralidade, chegando o tempo em que não mais escrevemos de acordo com a nossa fala (OLIVEIRA, 2005, p. 10).

Kato (1986) faz uma importante observação sobre a relação fala e escrita na aquisição da escrita: inicialmente, a criança apresenta uma fala, a Fala 1, que interfere na sua primeira escrita, Escrita 1. Esse seria o caso de uma criança que escreve a palavra "tomate" como "tumatchi", porque fala desta forma. No entanto, também a escrita pode interferir na fala, de modo que, depois, a criança apresenta uma Escrita 2 que interfere na Fala 2. Este seria o caso de uma criança que, ao aprender a escrita de uma palavra como "menino", passa a querer pronunciar tal palavra tal como escreve, mesmo que a pronúncia dessa palavra no seu dialeto seja

diferente (por exemplo, "mininu"). Neste trabalho, nos detemos na relação Fala 1 - Escrita 1, pois a relevância para este trabalho é a escrita da criança e não a fala.

Passamos a tratar do fator frequência e o Modelo de Redes, de Bybee (2001). O fator frequência é importante para este estudo, porque, dentro do Modelo de Redes, Bybee propõe que a frequência das palavras na língua pode afetar a fala dos usuários da língua. Mais do que isso, a frequência é fundamental na emergência da gramática e do léxico. Como um Modelo Baseado no Uso (FERRARI, 2011), o Modelo de Redes contribui para a discussão sobre a organização da informação linguística na mente humana a partir da experiência de uso da língua.

Bybee (2001) afirma que a frequência influencia na representação mental das palavras em cada pessoa, de modo que, por exemplo, uma palavra muito utilizada e frequente na língua, com forte representação mental, tende mais a sofrer fenômenos de redução sonora. Segundo Bybee, as palavras são estocadas na mente em forma de rede. Como afirma Huback,

[...] as palavras estão listadas no léxico mental em forma de redes, sendo agrupadas de acordo com identidade e similaridade fonológica ou semântica. A estrutura morfológica emerge dessas conexões, fazendo com que uma determinada classe paradigmática seja mais ou menos forte, dependendo da quantidade de itens que se flexionam através dela (HUBACK, 2007, p.121).

Para que fique mais clara essa discussão do Modelo de Redes, apresentamos aqui uma representação ilustrativa de como a informação linguística se estrutura na mente de uma pessoa (com as palavras em rede), considerando a estrutura linguística em foco neste estudo: o encontro CCV.

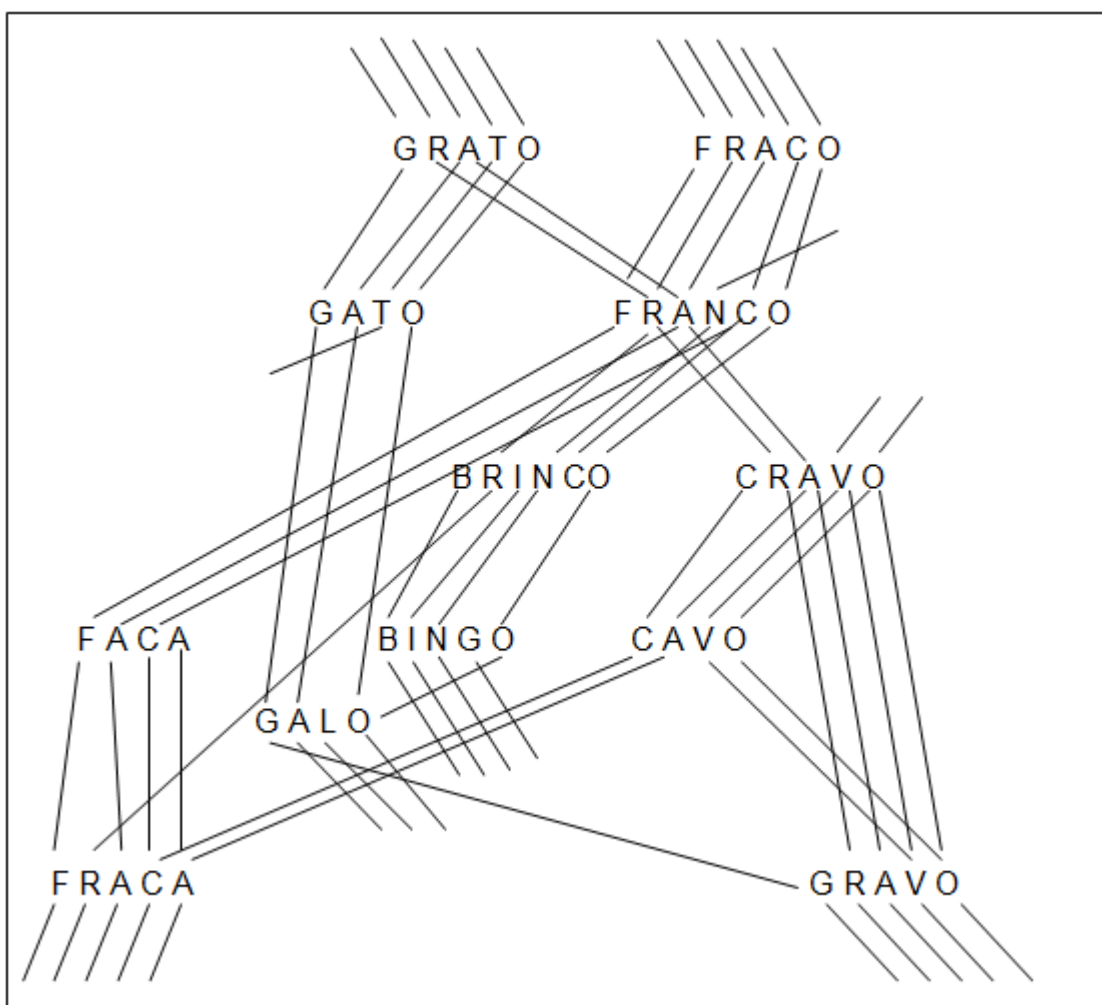


Figura 1: Rede de conexões lexicais – fonéticas, semânticas e morfológicas. Baseado em Bybee (2001).

Por meio da figura 1, podemos observar como as palavras se estruturam na mente humana, tal como proposto pelo Modelo de Redes. A informação linguística se organiza em nossa mente com as palavras em rede, de modo que a gramática e o léxico emergem dessa forma de organização.

Pode-se notar, na figura, a ligação das palavras por semelhança fonológica ou semântica, ou fonética e semântica que resulta em relação morfológica. Nessa figura, notamos relações fonéticas, semânticas e morfológicas, por exemplo, entre as palavras “fraco” e “frango”. Tais itens léxicos se relacionam apenas foneticamente no encontro consonantal FR. Também se relacionam semântica e foneticamente (e, por consequência, morfológicamente) por meio do sufixo indicativo de gênero ‘o’. É necessário esclarecer que, na Figura 1, os traços

que aparecem nas extremidades indicam que as palavras podem se conectar a outras palavras que se encontram além da rede focalizada na figura. Essa ilustração permite, ainda, perceber que as palavras não são estocadas na mente de forma aleatória, mas sim de maneira organizada e sistematizada.

Bybee (2001) propõe dois tipos de frequência: *type frequency* (frequência de tipo) e *token frequency* (frequência de ocorrência). Bybee descreve a primeira (*type frequency*) como sendo a frequência de um determinado padrão na língua, como por exemplo, o plural –eis que é frequente no português do Brasil (painéis, quartéis, pastéis), em contrapartida ao plural –eus já não é tão frequente (troféus, degraus). Assim, como o primeiro plural é mais usado, sua representação mental torna-se mais forte, de modo que, quando um falante realizar a pronúncia de uma palavra de menor frequência de que não conhece a sua forma plural, tenderá a fazer o uso do plural de maior frequência.

A frequência de ocorrência (*token frequency*) seria a quantidade de vezes que uma forma, como uma palavra, ocorre, por exemplo, em um *corpus* da língua. Assim, para se saber a frequência de ocorrência de uma palavra como “cravo” em um *corpus*, deve-se contar o número de vezes que tal palavra ocorreu nesse *corpus*. Também a frequência de ocorrência tem impacto nas representações mentais, segundo o Modelo de Redes, de modo que formas mais frequentes também apresentam representação mais forte em nossa mente. Este estudo irá avaliar se há algum efeito de frequência (de tipo ou de ocorrência) no fenômeno aqui observado.

É importante ressaltar que, para avaliar aqui o fator frequência na aquisição da escrita, recorreremos à área de Linguística de *Corpus* (SARDINHA, 2000; SARDINHA, 2004) que passa a ser tratada a partir do próximo parágrafo.

Para alguns estudiosos da língua, é importante apoiar-se em usos reais da linguagem. Essa é uma premissa importante para a linguística de *corpus* que tornou possíveis estudos de uso, para generalizar ou esboçar teorias sobre o funcionamento linguístico (SARDINHA, 2000).

Atualmente, uma ferramenta muito utilizada para realizar tais estudos é o computador, visto que os *corpora* são eletrônicos. Portanto, podemos afirmar que a linguística de *corpus* contemporânea é caracterizada por ter sua coleta e análise de

corpora eletrônicos utilizando o auxílio de ferramentas eletrônicas (SARDINHA, 2000).

Um dos principais *corpora* digitais de linguagem infantil no mundo é o Sistema de Intercâmbio de Dados de Linguagem Infantil (CHILDES)⁴, criado em 1984 por Brian MacWhinney e Catherine Snow. O principal objetivo desse *corpus* é tornar acessível dados sobre a aquisição da fala da primeira língua. As primeiras transcrições são datadas dos anos 60, atualmente é possível pesquisar no *corpus* transcrições, áudios e vídeos em mais de 20 idiomas dentro de 130 *corpora* diferentes disponíveis ao público de todo mundo. Foi acrescentado ao CHILDES recentemente um componente chamado *TalkBank*, um *corpus* de maior número de dados que inclui linguagem de afásicos, aquisição da segunda língua, análise de conversação e aprendizagem de línguas em sala de aula.

O presente estudo adota o já citado *corpus* do e-Labore, uma grande referência hoje no Brasil em dados de aquisição da escrita. A próxima seção aborda o e-Labore em detalhes.

03 – METODOLOGIA

Como mencionado, na análise aqui proposta, foi utilizado um *corpus* digital de escrita infantil o e-Labore: Laboratório Eletrônico de Oralidade e Escrita (CRISTÓFARO-SILVA et al., 2007)⁵. O e-Labore apresenta redações de alunos de 6 a 12 anos (do 1º ao 7º ano Ensino Fundamental) de escolas públicas e particulares da cidade de Belo Horizonte. Os temas das redações foram escolhidos por cada professor de maneira aleatória, de modo a garantir diferentes tipos de palavras. O *corpus* apresenta os textos dos alunos digitalizados e digitados no original, com as formas incorretas ou erros ortográficos entre colchetes e as formas corretas correspondentes a esses erros entre chaves. Esse procedimento metodológico viabiliza a pesquisa no *corpus* de erros ortográficos.

⁴ Disponível em: <http://childes.psy.cmu.edu/>. Acesso em jun. 2014.

⁵ Maiores informações sobre o *corpus* e-Labore podem ser obtidas em: <http://www.projetoaspa.org/elabore/>. Acesso em jun. 2014. Agradecemos a Profa. Thaís Cristófaros-Silva (UFMG), coordenadora geral desse projeto, o acesso ao *corpus* do e-Labore.

Revista Brasileira de Educação e Cultura – ISSN 2237-3098 Centro de Ensino Superior de São Gotardo	Número XII Jul-Dez 2015	Trabalho 08 Páginas 125-142
http://periodicos.cesg.edu.br/index.php/educacaoecultura	periodicoscesg@gmail.com	

Para realizar o presente estudo no *corpus*, tivemos de estabelecer um recorte⁶, tomando como referência o trabalho de Fontes-Martins e Oliveira Guimarães (2010). As autoras realizaram uma pesquisa com os dois encontros tautossilábicos com a líquida tepe [r] mais frequentes (TR e PR) e os dois menos frequentes (DR e VR) no *corpus* do e-Labore também. Assim, aqui optamos por analisar os quatro encontros CCV que não foram objeto de pesquisa de Fontes-Martins e Oliveira-Guimarães: os encontros “BR e CR” (mais frequentes) e os encontros “GR e FR” (menos frequentes).

Apresentamos os resultados de frequência encontrados pelas autoras no *corpus* do e-Labore na Tabela 1 a seguir. Os encontros CCV considerados neste trabalho encontram-se em destaque de negrito na tabela.

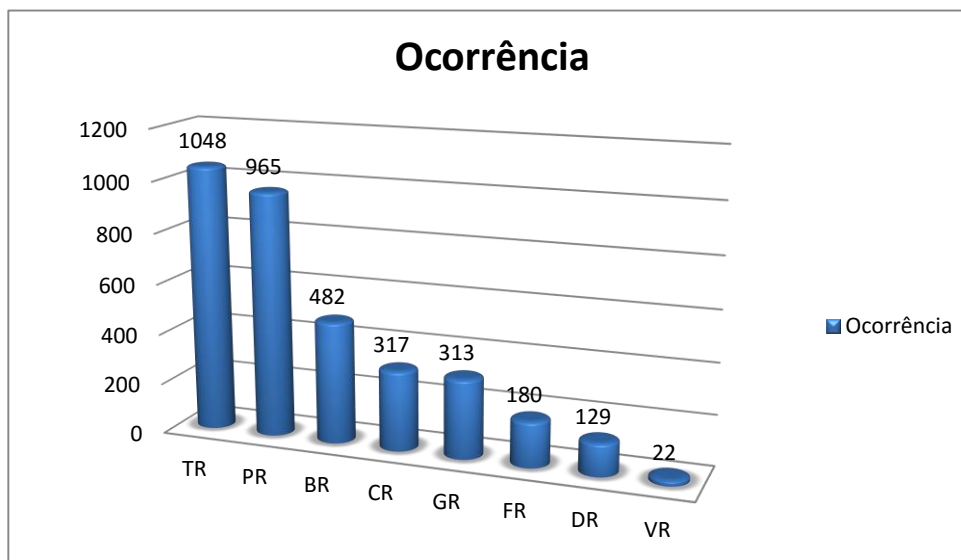
Tabela 1: Número de ocorrências dos 8 tipos de encontros CCV com “r” no *corpus* do e-Labore

Tipo	Número de ocorrências
TR	1048
PR	965
BR	482
CR	317
GR	313
FR	180
DR	129
VR	22
Total	3456

Fonte: Fontes-Martins e Oliveira-Guimarães (2010)

Para exemplificar de outra maneira a informação dessa tabela, apresentamos o gráfico a seguir:

⁶ Os recortes realizados neste estudo foram importantes em razão da limitação deste trabalho, fruto de um Trabalho de Conclusão de Curso – TCC.

Gráfico 1: Número de ocorrências dos tipos de encontros CCV com “r” no *corpus* e- Labore.

A análise do Gráfico 1 permite perceber que os dois encontros CCV com "r" mais frequentes já analisados pelas autoras citadas (TR e PR) se destacam em termos de ocorrência no *corpus* e-Labore. O gráfico possibilita também a visualização da pequena diferença no número de ocorrências dos 4 encontros CCV ("BR"/ "CR" e "GR"/"FR") aqui considerados. A maior diferença ocorre entre BR (482 ocorrências) e FR (180 ocorrências), sendo menor a diferença entre os outros dois encontros (CR com 317 ocorrências e GR com 313). Na análise aqui realizada, isso deverá ser ponderado, a fim de verificar se essas diferenças maior ou menor entre tais encontros será significativa nos resultados encontrados. Possivelmente, tais fenômenos ocorre devido a um efeito de frequência (BYBEE, 2001)

Veremos agora os resultados referentes à análise realizada, no *corpus* digital e-Labore, com os 4 encontros CCV com "r" aqui considerados: BR, CR, GR e FR. Como Fontes-Martins e Oliveira-Guimarães (2010), fizemos um novo recorte, em que propusemos a análise das cinco palavras mais frequentes de cada encontro CCV citado e cinco palavras de baixa frequência desses 4 encontros, o que totalizou 40 palavras a serem aqui analisadas. Tal recorte foi necessário por causa da grande quantidade de dados presentes no *corpus* e-Labore. A seguir, apresentamos os resultados deste estudo e sua respectiva análise.

04 – ANÁLISE DOS ENCONTROS CCV EM FOCO, NO *CORPUS* DO E-LABORE

Neste momento, abordamos os resultados da análise realizada. Primeiramente, são apresentados os resultados referentes às cinco palavras mais frequentes e cinco palavras de baixa frequência⁷ com encontro consonantal do tipo BR nas tabelas 2 e 3 a seguir.

É importante esclarecer que, neste trabalho, são considerados erros ortográficos que recaem somente nos encontros consonantais tautossilábicos em foco. As demais partes das palavras não foram consideradas. Esse dado deve ser considerado na análise de todas as tabelas a seguir.

Tabela 2: Palavras de maior frequência de ocorrência com encontro CCV BR

Palavra	Frequência absoluta no <i>corpus</i>	Frequência relativa no <i>corpus</i>	N e % de erros ortográficos envolvendo o encontro CCV
Brasil	4471	0,10%	34 (0,7%)
Sobre	735	0,24%	8 (1%)
Brincar	629	0,21%	10 (1,6%)
Brasileiro	241	2,48%	2 (0,8%)
Brasileira	220	3,36%	1 (0,4%)

Tabela 3: Palavras de baixa frequência de ocorrência com encontro CCV BR

Palavra	Frequência absoluta no <i>corpus</i>	Frequência relativa no <i>corpus</i>	Número de erros ortográficos envolvendo o encontro CCV
Sobrevivente	10	0%	0
Sobreviveria	10	0%	0
Sobreviveríamos	10	0%	0
Sóbrio	10	0%	0
Zambrota ⁸	10	0%	0

Pela Tabela 2, pode-se observar que as palavras de maior frequência com encontro do tipo BR no *corpus* e-Labore apresentam ocorrências com erro ortográfico. Nota-se, na última coluna, que os percentuais de erros são baixos, mas há ocorrência de erro ortográfico, o que não ocorre nos dados da Tabela 3. Nesta

⁷ Seguimos o procedimento metodológico indicado por Fontes-Martins e Oliveira-Guimarães (2010) de não selecionar palavras de menor frequência em um *corpus*. Isso porque em *corpora* de língua, é muito comum que haja um número abundante de tipos palavras com apenas 1 ocorrência. Muitas vezes, quase metade das palavras de um *corpus* têm apenas uma ocorrência. Uma frequência de ocorrência tão baixa dificulta a análise dos resultados, a possibilidade de se fazerem afirmações mais acuradas sobre um fenômeno estudado. Dessa maneira, adotamos palavras de baixa ocorrência no *corpus* do e-Labore, no caso, palavras na faixa de 10 ocorrências.

⁸ A palavra coletada “Zambrota” corresponde ao nome de um jogador italiano.

tabela, nota-se que as palavras de baixa frequência não apresentam sequer uma ocorrência de erro ortográfico. Esse dado parece indicar que palavras com maior frequência apresentam maior probabilidade de serem acometidas por erros ortográficos, um achado semelhante ao de Fontes-Martins e Oliveira-Guimarães (2010). Cabe prosseguirmos com as análises, para investigar se esse comportamento se repete nos demais encontros aqui focalizados: CR, GR e FR.

A seguir as tabelas 4 e 5 mostrarão à análise das palavras mais frequentes e de baixa frequência com encontro consonantal do tipo CR:

Tabela 4: Palavras de alta frequência de ocorrência com encontro CCV CR

Palavra	Frequência absoluta no <i>corpus</i>	Frequência relativa no <i>corpus</i>	N e % de erros ortográficos envolvendo o encontro CCV
Crianças	819	0,69%	2 (0,2%)
Escravos	373	1,34%	1 (0,2%)
Croácia ⁹	283	0,57%	7 (2,4%)
Criança	212	0,61%	2 (0,9%)
Escreve	126	0,79%	1 (0,7%)

Tabela 5: Palavras de baixa frequência de ocorrência com encontro CCV CR

Palavra	Frequência absoluta no <i>corpus</i>	Frequência relativa no <i>corpus</i>	N e % de erros ortográficos envolvendo o encontro CCV
Escreves	10	0%	0
Escrever	10	0%	0
Lucros	10	0,1%	1 (10%)
Micro	10	0%	0
Incrivelmente	10	0%	0

A análise feita nas tabelas 2 e 3 pode ser aplicada, praticamente na íntegra, para a análise feita para o encontro CR nas tabelas 4 e 5. No entanto, aqui é destaque um maior índice de erros ortográficos em uma palavra das mais frequentes, “Croácia”, com 2,4% de erros. Esse percentual pode ser considerado ainda baixo, mas indica que as palavras mais frequentes podem apresentar algum comportamento diferente entre si, o que indica um efeito lexical (BYBEE, 2001). Também ressalta aqui o fato de que uma palavra de baixa frequência, “lucros”, apresentou 1 ocorrência de erro ortográfico, o que representa 10% no total de 10 ocorrências. No entanto, no ponto de vista da estatística, em termos de

⁹ Assim como a palavra coletada “Zambrotá”, citada acima, também a palavra “Croácia”, ilustra como as provas do e-labore são realizadas, sendo o tema escolhido pelo professor da classe. Desse modo é possível inferir o possível tema para essas redações, como “Copa do Mundo”, e este seria um fator para a coleta em maior número destas palavras.

representatividade de dados, esse dado percentual precisa ser relativizado, tendo em vista que o significado de 1 ocorrência em 10 é muito inferior ao significado de 1 ocorrência em 373 (na palavra “escravos” na tabela 4).

Até aqui, tratamos dos dois encontros CCV mais frequentes aqui analisados, BR e CR. A partir de agora, tratamos dos encontros CCV menos frequentes a serem aqui avaliados: GR e FR.

Passamos então às tabelas 6 e 7 que trazem em seu conteúdo as palavras mais frequentes e de baixa frequência do encontro consonantal menos frequente GR.

Tabela 6: Palavras de alta frequência de ocorrência com encontro CCV GR

Palavra	Frequência absoluta no <i>corpus</i>	Frequência relativa no <i>corpus</i>	N e % de erros ortográficos envolvendo o encontro CCV
Grande	907	0,77%	12 (1%)
Negros	383	0,31%	4 (1%)
Alegria	309	0,51%	3 (1%)
Negro	224	0,53%	1 (0,4%)
Alegre	194	0,51%	2 (1%)

Tabela 7: Palavras de baixa frequência de ocorrência com encontro CCV GR

Palavra	Frequência absoluta no <i>corpus</i>	Frequência relativa no <i>corpus</i>	N e % de erros ortográficos envolvendo o encontro CCV
Ingrata	10	0%	0
Íngreme	10	0%	0
Ingressar	10	0%	0
Negrão ¹⁰	10	0,1%	1 (10%)
Tigre	10	0%	0

Notamos também, nas tabelas 6 e 7, comportamento semelhante aos das palavras analisadas anteriormente: maior ocorrência de erros nas palavras mais frequentes. Na verdade, as palavras com encontro GR (tabelas 6 e 7) se comportaram de modo mais similar ainda às palavras com encontro CR (tabelas 4 e 5). Isso também se repete, de modo geral, na análise do último encontro consonantal a ser aqui analisado FR, o que pode ser verificado nas tabelas 8 e 9 a seguir.

¹⁰ A variação encontrada na palavra “negrão” vem a explicar toda a discussão realizada no artigo sobre a possível influência do léxico da criança durante sua produção escrita. Foi encontrada a variante “negão”, portanto podemos entender que esta criança em seu meio ouve e reproduz esta palavra da maneira como escreveu, assim como Oliveira afirmou que as crianças no período de aquisição da escrita tende a escrever como falam.

Tabela 8: Palavras de alta frequência de ocorrência com encontro CCV FR

Palavra	Frequência absoluta no <i>corpus</i>	Frequência relativa no <i>corpus</i>	N e % de erros ortográficos envolvendo o encontro CCV
França	663	0,25%	11 (1,6%)
Frente	214	0,06	2 (1%)
Fred	170	0,31%	4 (2%)
África	104	0,68%	0
Frio	70	0,05	0

Tabela 9: Palavras de baixa frequência de ocorrência com encontro CCV FR

Palavra	Frequência absoluta no <i>corpus</i>	Frequência relativa no <i>corpus</i>	Número de erros ortográficos envolvendo o encontro CCV
Frequente	10	0%	0
Frequento	10	0%	0
Frigideira	10	0%	0
Sofredor	10	0%	0
Sofríamos	10	0%	0

Um dado diferente notado na Tabela 8 é que, pela primeira vez, há duas palavras mais frequentes (“África” e “Frio”) que não apresentam qualquer erro ortográfico. Pondera-se sobre a possibilidade de o próprio encontro FR (presente em “África” e “Frio”), que é um tipo de encontro CCV menos frequente, interferir nesse caso, considerando, como os dados têm apontado, que a menor frequência parece tender a desfavorecer erros ortográficos.

De modo geral, as análises das tabelas de 2 a 9 permitem notar que a tendência é que quanta mais baixa a frequência de ocorrência menor é o número de erros ortográficos (na maioria das palavras chegando a zero). Esse resultado reforça os achados de Fontes-Martins e Oliveira-Guimarães (2010).

05 – CONSIDERAÇÕES FINAIS

Esta pesquisa realizou uma análise de sobre escrita infantil, utilizando dados coletados do *corpus* e-Labore. O foco deste estudo foi a grafia de encontros tautossilábicos mais e menos frequentes (BR, CR, GR, FR), tendo como referência a pesquisa de Fontes-Martins e Oliveira-Guimarães (2010).

O Modelo de Redes (BYBEE, 2001) foi utilizado como principal embasamento teórico para este trabalho, fundamentando a questão da influência da frequência na aquisição da escrita infantil. Foi preciso também discutir teorias sobre

aquisição, considerando, inclusive, que a criança, durante o processo de aquisição da escrita, tende a escrever da maneira que fala (OLIVEIRA, 2005).

A partir das análises feitas, podemos observar que há efeitos de frequência no processo da aquisição dos quatro encontros CCV aqui avaliados. Palavras com maior frequência de ocorrências apresentaram mais erros, enquanto que as de baixa frequência apresentaram pouquíssimas ocorrências de erro.

Além disso, como os tipos CCV mais frequentes e menos frequentes não apresentaram comportamento diferente, isso nos permite afirmar que a frequência de ocorrência é que se tornou relevante para que ocorressem os erros ou variações nas palavras analisadas (FONTES-MARTINS E OLIVEIRA-GUIMARÃES 2010).

Como o tipo FR (o menos frequente dos avaliados) apresentou menos ocorrências de erros ortográficos nas palavras mais frequentes, é possível que há algum efeito de frequência de tipo, em que a menor frequência de tipo parece tender a desfavorecer erros ou variações ortográficas. Contudo, esse dado precisa ser investigado em um estudo mais amplo.

06 – REFERÊNCIAS

BYBEE, Joan. *Phonology and Language Use*. Cambridge: Cambridge University Press, 2001.

_____, Joan; HOPPER, Paul (Ed.). *Frequency and the emergence of linguistic structure*. Amsterdam; Philadelphia, PA: John Benjamins, 2001. p. 337 - 361.

CAGLIARI. A respeito de alguns fatos do ensino e da aprendizagem pelas crianças na alfabetização. In: ROJO, Roxane (Org.). *Alfabetização e letramento: perspectivas lingüísticas*. Campinas, SP: Mercado de letras, 1998.

CANTONI, M. M. *Máximo e próximo: um estudo sobre opacidade lexical no Português brasileiro segundo modelos baseados no uso*. Disponível em: <http://www.fflch.usp.br/dlcv/lport/pdf/post/13.pdf>. Acesso em jun. 2014.

Revista Brasileira de Educação e Cultura – ISSN 2237-3098 Centro de Ensino Superior de São Gotardo http://periodicos.cesg.edu.br/index.php/educacaoecultura	Número XII Jul-Dez 2015 periodicoscesg@gmail.com	Trabalho 08 Páginas 125-142
--	--	--------------------------------

COUTINHO, Marília de Lucena. Psicogênese da língua escrita: o que é? Como intervir em cada um das hipóteses? Uma conversa entre professores. In: MORAIS, Arthur Gomes de; ALBUQUERQUE, Eliana Borges Coreia de; LEAL, Telma Feraz (ORGS). *Alfabetização: apropriação do sistema de escrita alfabeto*. Belo Horizonte: Autêntica, 2005.

FERREIRO, Emilia e TEBEROSKY, Ana. *Psicogênese da Língua Escrita*. Porto Alegre: Artmed Editora, 1999.

FONTES-MARTINS, Raquel; OLIVEIRA-GUIMARÃES, Daniela. Efeitos de frequência na produção escrita de encontros consonantais. *Revista de Estudos Linguísticos*, 39 (2), 18-35, 2010.

HUBACK, A. P. *Efeito de frequência nas Representações Mentais*. 2007. Tese (Doutorado) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2007. P.114 - 147. Disponível em: <http://www.projetoaspa.org/cristofaro/orientacao/phd/concluida/huback2007.pdf>. Acesso em jun. 2014.

KATO, Mary. *No mundo da escrita: uma perspectiva psicolinguística*. São Paulo: Ática, 1986.

MIRANDA, I. C. C. e CRISTÓFARO SILVA, T. Aquisição de encontros consonantais tautossilábicos: uma abordagem multirrepresentacional. *Revista Linguística*. Disponível em: <http://www.letras.ufrj.br/poslinguistica/revistalinguistica/index.php/teoria-fonologica/aquisicao-de-encontros-consonantais-tautossilabicos-uma-abordagem-multirrepresentacional-2/>. Acesso em: 19/05/2014.

MORAIS, Artur Gomes de; ALBUQUERQUE, Eliana B. C. de; LEAL, Telma Ferraz. (Orgs.). *Alfabetização: apropriação do sistema de escrita alfabética*. Belo Horizonte: Autêntica, 2005, p. 47-70.

FERRARI, Lílian. *Introdução à Linguística Cognitiva*. São Paulo: Contexto, 2011.

MORAES, Maria Luisa Fickelcherer; MARTINS, Raquel Márcia Fontes. Investigando a Aprendizagem da Escrita em um *Corpus* Digital de Linguagem Infantil.

OLIVEIRA, M. A. *Conhecimento Linguístico e Apropriação do Sistema de Escrita*. Disponível em: http://www.pucminas.br/imagedb/mestradodoutorado/publicacoes/PUA_ARQ_ARQUI20121017141358.pdf. Acesso em: 22/05/2014.

PROJETO E-LABORE (Laboratório eletrônico de Oralidade e Escrita). *Corpus de escrita e fala infantil*. Disponível em: <www.projetoaspa.org/elabore>. Acesso em: jun 2009.

Sistema de Intercâmbio de Dados de Linguagem Infantil – CHILDES. Disponível em: <http://childes.psy.cmu.edu/>. Acesso em: 06/06/2014.

Revista Brasileira de Educação e Cultura – ISSN 2237-3098 Centro de Ensino Superior de São Gotardo	Número XII Jul-Dez 2015	Trabalho 08 Páginas 125-142
http://periodicos.cesg.edu.br/index.php/educacaoecultura	periodicoscesg@gmail.com	