UNIVERSIDADE FEDERAL DE ALAGOAS - UFAL

CENTRO DE EDUCAÇÃO - CEDU

PROGRAMA DE PÓS-GRADUAÇÃO EM EDUCAÇÃO - PPGE

ELIAN DA SILVA SANTOS

MANUSCRITOS ESCOLARES DE ALUNOS RECÉM-ALFABETIZADOS: UM ESTUDO SOBRE A DENSIDADE E DIVERSIDADE LEXICAL

MACEIÓ

2020

ELIAN DA SILVA SANTOS

MANUSCRITOS ESCOLARES DE ALUNOS RECÉM-ALFABETIZADOS: UM ESTUDO SOBRE A DENSIDADE E DIVERSIDADE LEXICAL

Dissertação apresentada ao Programa de Pós-Graduação da Universidade Federal de Alagoas (UFAL), como requisito para a obtenção do título de Mestre em Educação.

Orientador: Prof. Dr. Eduardo Calil

MACEIÓ

2020

Catalogação na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico

Bibliotecário: Marcelino de Carvalho Freitas Neto – CRB-4 – 1767

S273m Santos, Elian da Silva.

Manuscritos escolares de alunos recém-alfabetizados : um estudo sobre a densidade e diversidade lexical / Elian da Silva Santos. – 2020.

116 f. : il. color.

Orientador: Eduardo Calil.

Dissertação (Mestrado em Educação) – Universidade Federal de Alagoas. Centro de Educação. Programa de Pós-Graduação em Educação. Maceió, 2020.

Bibliografia: f. 78-82.

Apêndices: f. 81-116.

1. Lexicografia. 2. Escrita colaborativa. 3. Indicadores linguísticos - Densidade e diversidade. 4. Neo-alfabetizado. I. Título.

CDU: 37.014.22:81'374

Universidade Federal de Alagoas
Centro de Educação
Programa de Pós-Graduação em Educação

MANUSCRITOS ESCOLARES DE ALUNOS RECÉM-ALFABETIZADOS: UM ESTUDO SOBRE A DENSIDADE E DIVERSIDADE LEXICAL

ELIAN DA SILVA SANTOS

Defesa de Dissertação de Mestrado submetida à banca examinadora, já referendada pelo Programa de Pós-Graduação em Educação da Universidade Federal de Alagoas e aprovada em 19 de Fevereiro de 2020.

Banca Examinadora:

Prof(a). Dr(a). EDUARDO CALIL DE OLIVEIRA (UFAL)
Orientador

Prof(a). Dr(a). ADRIANA CAVALCANTI DOS SANTOS (UFAL)
Examinador(a) Interno(a)

Prof(a). Dr(a). PEDRO DE LEMOS MENEZES (UNCISAL)
Examinador(a) Externo(a)

RESUMO

Ao longo das últimas décadas, diferentes abordagens vêm sendo propostas para explicar o conhecimento lexical e desenvolver métricas para a sua avaliação no contexto escolar. Um índice usado para mensurar esse conhecimento é a riqueza lexical, podendo ser representada através de indicadores linguísticos, como a densidade e diversidade lexical**.** O presente trabalho analisou a riqueza lexical em manuscritos escolares produzidos por díades de alunos brasileiros recém-alfabetizados, tomando como categorias de análise o número de Palavras Escritas (PE), Densidade Lexical (DeL) e Diversidade Lexical (DiL). Para isso, selecionamos 30 contos etiológicos escritos por 5 díades de alunos brasileiros recém-alfabetizados, coletados durante um projeto didático intitulado "Contos do como e do porquê" no ano de 2012 em uma escola privada de Maceió. Caracterizado como um estudo exploratório, adotamos uma abordagem metodológica de léxico-estatística textual para quantificação e análise de PE, DeL e DiL. Assim, para mensurar o número de PE utilizamos o *software* de análise lexical, *Lexicanalytics*. A DiL foi mensurada através da medida D, aplicada ao modelo VOCD e para a medição da DeL, dividimos o número de itens lexicais (substantivos, adjetivos, verbos e advérbios modais terminados em -mente) pelo total de palavras do texto. Posteriormente, os valores obtidos foram submetidos a uma análise estatística descritiva e inferencial. Nossos resultados indicaram que as díades escreveram seus textos com uma média de 90 palavras, com DeL de 48% e DiL de 49,40%. Na análise inferencial, não constatamos correlação linear entre as três categorias estudadas, sugerindo que são independentes e merecem uma avaliação individual no texto. Além disso, na comparação entre as médias obtidas pelas díades, não encontramos diferença significativa em nenhuma das categorias estudadas, ou seja, os alunos usaram valores aproximados de PE, DeL e DiL para produzirem seus textos. Em suma, os resultados deste estudo sugerem que a escrita colaborativa pode auxiliar nas escolhas lexicais dos alunos, favorecendo uma produção textual com vocabulário diversificado e densidade informacional.

Palavras-chave: Conhecimento lexical. Escrita colaborativa. Indicadores Linguísticos. Riqueza Lexical.

ABSTRACT

In the last decades, different approaches have been proposed to understand the vocabulary knowledge and to develop metrics for its analysis at scholar environments. An index used to measure this knowledge is the lexical richness, represented with linguistic components, e.g. lexical diversity and lexical density. This study analyzed the lexical richness in scholar manuscripts produced by newly literate Brazilian dyads, considering the number of Written Words (PE), Lexical Density (DeL), Lexical Diversity (DiL) as categories. Therefore. We selected 30 etiological stories written by 5 newly literate Brazilian dyads, gathered during the development of a didactic project named "Contos do como e do porquê" in the year of 2012 on a private school at Maceió. Focusing on an exploratory study, we adopted a textual lexico-statistical methodological approach to quantify and analyze the categories PE, DeL and DiL. Thus, to measure the number of written words, we used the lexical analysis software Lexicanylitcs. The DiL category was measured using the D value, applied to the VOCD model. To measure the DeL category, we divided the number of lexical items (subjects, adjectives, verbs and modal adverbs) by the total number of words. Then, the obtained values were submitted to a descriptive and inferential statistical analysis. Our results indicate that the dyads wrote their productions with an average of 90 words, scoring a DeL of 48% and a DiL of 49.40%. We did not find correlation between the three categories, suggesting that they are independents and deserve individual evaluation. Furthermore, we did not find relevant variation over the values obtained by the three categories, in other words, the students scored similar PE, DeL and DiL. In summary, the results suggest that the collaborative writing may aid in the student's lexical choices, supporting a textual production with diversified vocabulary and information density.

Keywords: Collaborative Writing. Linguistics Indexes. Lexical Richness. Lexical Knowledge.

A Deus, por seu imenso amor e por sempre guiar meus passos,
tornando possível a concretização desse sonho.

AGRADECIMENTOS

Aos meus pais, Maria e Manoel, pelo o amor, incentivo e apoio incondicional.

Aos meus familiares, em especial, meus irmãos Eliel e Anderson, meus tios Ivanildo e Lurdinha, pela torcida.

Ao meu amado Glauber, toda minha gratidão e amor. Você foi meu pilar durante essa caminhada. Sua presença, suas doces palavras, sua escuta e seu sorriso sempre me acalmaram nos momentos mais difíceis.

Aos meus sogros, Prof.ª Luciene e Prof. Geovane, pelas conversas, conselhos e por vibrarem comigo a cada conquista.

Ao meu orientador, Prof. Dr. Eduardo Calil, pela paciência, confiança, escuta e apoio constante. Sou extremamente grata por acreditar nesse trabalho e me guiar nessa caminhada.

À Prof.ª Drª Adriana Cavalcanti e Prof. Dr. Pedro Menezes, por se tornarem parte da história desse trabalho e pelas valiosas contribuições.

Aos meus queridos amigos Lys Calisto, Wilton e Vanessa, pela ajuda mútua, pelas palavras sinceras e por caminharem comigo durante essa jornada, pois como já dizia Fernando Pessoa, “Benditos sejam os amigos que acreditam na tua verdade ou te apontam a realidade. Porque amigo é a direção. Amigo é a base quando falta o chão.

Aos membros do LAME, em especial, à Salézia Magna e Mayara Cordeiro, pelo acolhimento, partilha e incentivo.

À Universidade Federal de Alagoas, minha segunda casa, por me proporcionar belos momentos de aprendizagem e alegria.

Ao Centro de Educação (CEDU) e ao Programa de Pós-Graduação em Educação (PPGE), por toda atenção e serviços prestados.

À CAPES, pelo apoio financeiro que contribuiu para a viabilização deste trabalho.

COMO SURGIRAM AS PALAVRAS?

Era uma vez duas meninas que eram irmãs.
Elas foram para uma casa e lá
haviam muitas letras e essas letras não
se encaixavam. A moça que morava lá
virou amiga das duas meninas e ela disse:
- Vocês querem tirar essas letras comigo?
E as meninas disseram:
- Sim
E todas tiraram as letras.
Aquela moça pensava que aquelas
letras não formavam palavras.
E elas formaram muitas palavras.

(Conto etiológico escrito por uma dupla de alunas brasileiras de 7 anos).

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE QUADROS

LISTA DE GRÁFICOS

LISTA DE ABREVIATURAS E SIGLAS

AL	Alagoas
ANOVA	ANalysis Of VAriance
BDTD	Biblioteca Digital Brasileira de Teses e Dissertações
CL	Conhecimento Lexical
CNPQ	Conselho Nacional de Pesquisa
CL	Conhecimento de Lexical
CRTTR	Corrected Type-Token Ratio
DeL	Densidade Lexical
DiL	Diversidade Lexical
LAME	Laboratório do Manuscrito Escolar
NE	Número de Erros
PIBIC	Programa Institucional de Bolsas de Iniciação Científica
RL	Riqueza Lexical
RTTR	Root Type-Token Ratio
SCIELO	*Scientific Electronic Library Online*
SL	Sofisticação Lexical
PE	Palavras Escritas
TTR	Type-Token Ratio
UFAL	Universidade Federal de Alagoas

SUMÁRIO

1 INTRODUÇÃO

Nas últimas décadas, o grupo de pesquisa Escritura, Texto & Criação (ET&C) e o Laboratório do Manuscrito Escolar (LAME)[1] vem encadeando uma ampla investigação sobre o processo de escritura de alunos recém-alfabetizados, elegendo como objeto de pesquisa manuscritos escolares produzidos a partir de uma proposta metodológica de escrita colaborativa a dois (díade).

Vários aspectos do processo de escrita textual têm sido analisados, dentre eles, a ortografia, a rasura, a intertextualidade e os comentários (CALIL e PEREIRA, 2018; CORDEIRO e CALIL, 2018; CALIL et al., 2017; CALIL, 2016). Diante disso, observamos que o conhecimento lexical dos alunos, tomando como métrica a riqueza lexical de seus manuscritos (MENARD, 1983; READ, 2000) também é um objeto de estudo valioso. Em relação à riqueza lexical, ela é definida na literatura como uma característica multidimensional da escrita, representada através de indicadores linguísticos, como a densidade e diversidade lexical (READ, 2000).

Compreendendo a importância desses dois indicadores para um maior aprofundamento acerca dos aspectos lexicais da escrita escolar, iniciamos uma pesquisa em meados de 2016, focalizando na escrita textual de alunos recém-alfabetizados. Vale ressaltar que esse trabalho foi iniciado na graduação em Pedagogia, através do Programa Institucional de Bolsas de Iniciação Científica (PIBIC)[2].

Essa pesquisa inicial foi realizada durante o período de dois anos (2016 a 2018). No primeiro ano, concentramo-nos na revisão bibliográfica por meio de buscas em algumas bases de dados nacionais e internacionais como Periódicos CAPES, Scientific *Electronic Library Online* (SCIELO), *SciVerse Scopus*, Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), Associação Nacional de Pesquisa e Pós-Graduação em Educação (ANPEd*)* e Repositório Institucional da Universidade Federal de Alagoas (UFAL).

No segundo ano, através de uma colaboração internacional com a universidade de Aveiro/Portugal, por vias do projeto *InterWriting*[3], examinamos a densidade e

[1] O Laboratório do Manuscrito Escolar (LAME), criado em junho de 2010 e reúne pesquisadores do Programa de Pós-Graduação em Letras e Linguística (PPGLL) e do Programa de Pós-Graduação em Educação (PPGE) da Universidade Federal de Alagoas (UFAL).

[2] Estudo financiado pelo Conselho Nacional de Pesquisa (CNPQ).

[3] Um projeto colaborativo internacional entre pesquisadores do Brasil (LAME), França e Portugal. O projeto visa o intercâmbio entre pesquisadores, coleta de dados nos três países, e a constituição de um corpus comum de processos de escritura de manuscritos escolares, coletados nas mesmas condições de

diversidade lexical em um *corpus* composto por textos de alunos portugueses, com faixa etária de 7 a 8 anos, a partir da análise de suas produções individuais e colaborativas (díade).

Os resultados dessa pesquisa nos possibilitaram uma ampliação de informações sobre a escrita desses alunos, como a média total de palavras usadas em cada texto, as palavras com maior e menor ocorrência, além de destacar as mudanças nas taxas de densidade e de acordo com cada proposta de produção individual e em díades.

Ao longo desse trabalho, colhemos bons frutos. Em 2017 recebemos o prêmio de Excelência Acadêmica pelo PIBIC-UFAL e indicação para o Prêmio Destaque de Iniciação Científica do CNPq. Na sequência, em 2018, essa pesquisa se transformou no meu Trabalho de Conclusão de Curso, sendo no mesmo ano publicado na revista científica Calidoscópio.

É nesse cenário, visando a continuidade de nossas investigações, que surge o presente trabalho de mestrado, tendo como proposta a análise da densidade e diversidade lexical em um *corpus* composto por contos etiológicos produzidos por díades de alunos recém-alfabetizados, com faixa etária de 7 a 8 anos, matriculados no 2º ano do Ensino Fundamental de uma Escola da rede privada de Maceió/AL. Destacamos que esse *corpus* foi coletado em 2012 por pesquisadores do LAME e armazenado em seu banco de dados.

Dentre as principais motivações que nos impulsionaram nesta pesquisa, está a tentativa de preencher uma lacuna sobre essa temática em estudos brasileiros, observada durante a revisão bibliográfica do nosso estudo inicial (2016-018), quando constatamos apenas três trabalhos. O primeiro trabalho encontrado é uma dissertação de Costa (1997) que usou a densidade lexical para avaliar a complexidade gramatical em redações de alunos do ensino fundamental (4º e 8º ano) e do ensino médio (2º ano). Os outros dois trabalhos são artigos que trazem resultados de uma pesquisa desenvolvida por Scherer (2002, 2011) quando analisou a diversidade lexical em falas espontâneas de crianças de 1 a 5 anos.

Entretanto, não encontramos estudos que se voltassem para a investigação do conhecimento lexical em textos escolares de alunos brasileiros, principalmente, recém-alfabetizados, tomando como medida os dois indicadores linguísticos, densidade e diversidade lexical.

---

produção e seguindo o mesmo protocolo de coletas de dados, o que possibilitará o desenvolvimento de futuros estudos comparativos acerca das práticas de ensino da escrita nos três países participantes.

Em virtude disso, a partir de um estudo exploratório, essa dissertação visa contribuir para produção de conhecimento nessa área de investigação, como também, apresentar novas possibilidades metodológicas para avaliação dos aspectos lexicais presentes na escrita textual dos alunos. Para tanto, buscamos responder às seguintes questões norteadoras desta pesquisa:

(i) Qual a taxa de palavras escritas, de densidade e diversidade lexical em manuscritos escolares produzidos por díades de alunos, quando escrevem seus contos etiológicos?

(ii) Existe correlação entre taxas de palavras escritas, de densidade e diversidade lexical desses contos?

(iii) Há diferença significativa dessas taxas entre as díades de uma mesma sala de aula?

Considerando que os alunos participantes deste estudo possuíam a mesma faixa etária, cursavam o mesmo ano escolar (2º ano do Ensino Fundamental), estudavam na mesma sala de aula, produziram o mesmo gênero textual (contos etiológicos) e seguiram o mesmo protocolo de coleta, também buscamos verificar nossas duas hipóteses sobre suas produções textuais:

- **Hipótese Principal**: Não há diferença significativa entre os textos escritos pelas 5 díades de alunos recém-alfabetizados, em termos de Densidade e Diversidade Lexical.
- **Hipótese Alternativa**: Há diferença significativa entre textos escritos por dupla de alunos recém-alfabetizados, em termos de Densidade e Diversidade lexical.

Para elucidar as questões da nossa pesquisa e verificar as hipóteses levantadas, traçamos os seguintes objetivos de estudo:

- **Objetivo Geral**: Mensurar a número de palavras escritas, a Densidade e Diversidade Lexical em contos etiológicos produzidos por alunos recém-alfabetizados, de uma escola da rede privada de Maceió/AL, a partir de uma proposta metodológica de escrita colaborativa (a dois).

- **Objetivos Específicos**:

1. Mensurar a média de palavras que as díades de alunos recém-alfabetizados usaram para escrever suas produções.
2. Medir a Densidade e Diversidade Lexical das produções textuais escritas pelas díades.
3. Quantificar o número de ocorrências dos substantivos, verbos, adjetivos e advérbios modais (terminados em -mente) nas produções das díades.
4. Verificar o modelo de distribuição que as palavras escritas, a densidade e a diversidade seguem.
5. Verificar se há diferença significativa entre os textos escritos pelas duplas, em termos de número de palavras, densidade e diversidade lexical
6. Analisar se há correlação entre o número de palavras escritas, densidade e diversidade lexical.

Para esse propósito, estruturamos a presente dissertação em seis seções, incluindo a introdução. Na segunda seção trazemos uma explanação da literatura e formalizamos alguns conceitos que constituem a nossa base teórica como conhecimento lexical, competência lexical, riqueza lexical, densidade e diversidade lexical. Em seguida, apontamos as principais medidas usadas para mensurar os indicadores lexicais.

Na terceira seção descrevemos alguns trabalhos que usaram esses dois indicadores para avaliar a escrita textual de alunos da educação básica. Ademais, trazemos uma discussão acerca dos fatores que podem contribuir para a densidade e diversidade lexical de um texto.

Descrevemos nossos procedimentos metodológicos na quarta seção, assim como a apresentação do *corpus*, das categorias investigadas a das métricas utilizadas para sua mensuração. Para concluir, destacamos todas as técnicas aplicadas para a avaliação das nossas categorias, o que envolve técnicas de estatística descritiva e inferencial.

Na quinta seção trazemos os resultados referentes às nossas categorias: Palavras Escritas (PE), Diversidade Lexical (DiL) e Densidade Lexical (DeL), subdivididos em duas análises: descritiva e inferencial. Nos resultados da análise descritiva, expusemos as médias de PE, DeL e DiL obtidas por cada dupla. Enquanto que na análise inferencial, trazemos os resultados dos testes de distribuição, variância e correlação. E por fim, na sexta seção ressaltamos nossas considerações finais, retomando os pontos principais da dissertação e apontando novas possibilidades para o desenvolvimento de trabalhos futuros.

2 INVESTIGAÇÃO: ANTECEDENTES E CONCEITOS

Nesta seção formalizamos alguns conceitos que constituem a nossa base teórica como conhecimento lexical, competência lexical, riqueza lexical, Em seguida, apontamos as principais medidas desenvolvidas ao longo dos anos para mensurar a densidade e diversidade lexical.

2.1 Conhecimento Lexical

O Conhecimento Lexical (CL) é essencial para a aprendizagem da língua materna e para o desenvolvimento das competências lexicais na escrita textual dos alunos. A competência lexical, por sua vez, é definida como a capacidade de compreender as palavras, na sua estrutura fonológica, morfossintática e nas suas relações de sentido com outros itens constitutivos da língua (FERRAZ, 2016, p.6).

Ao longo dos anos, estudos (HENRIKSEN, 1999; NATION, 2001; DALLER, 2007) vêm introduzindo diferentes abordagens com o propósito de proporcionar uma melhor compreensão acerca do CL, como também, desenvolver metodologias que possibilitem seu acompanhamento e avaliação no contexto escolar. No tocante dessas diferentes abordagens, encontramos em Henriksen (1999) o CL definido a partir de três níveis**:**

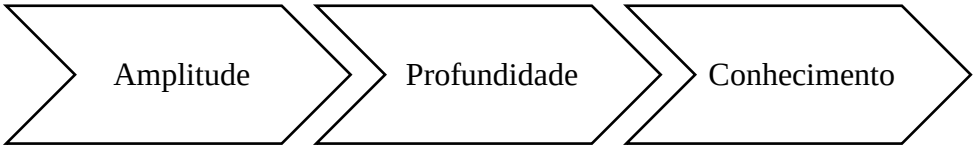


Figura 1- Níveis do Conhecimento Lexical (Henriksen, 1999).

A amplitude é definida como a quantidade de palavras que usamos durante nossas produções textuais, enquanto a profundidade refere-se à compreensão dos significados das palavras. O terceiro nível é o conhecimento, sendo este, classificado como receptivo ou produtivo. O conhecimento receptivo está relacionado com a capacidade de reconhecer e associar uma palavra ao seu significado durante um contexto de diálogo ou de leitura, por outro lado, o conhecimento produtivo está associado com a capacidade de representar

e associar corretamente a forma gráfica da palavra ao seu significado durante uma produção escrita.

Uma segunda abordagem acerca CL é proposta por Nation (2001). De acordo com esse autor, o conhecimento lexical é constituído por três níveis: formas das palavras, significado e uso. Sendo cada um deles, formados por três subníveis detalhados no quadro a seguir:

Quadro 1 - Níveis do Conhecimento Lexical (NATION, 2001)

Formas da palavra	Escrita	Como a palavra é escrita?
	Fala	Como a palavra é pronunciada?
	Partes da Palavra	Quais partes da palavra são necessárias para expressar tal significado?
Significado	Forma e significados	Que forma a palavra pode ser usada para expressar esse significado?
	Conceitos e referentes	A quais itens o conceito pode se referir?
	Associações	Que outras palavras podem ser usadas em vez desta palavra?
Uso	Funções gramaticais	Em que padrões deve-se usar essa palavra?
	Colocações	Que palavras ou tipos de palavras devem ser usadas?
	Restrições em uso	Onde, quando e com que frequência essa palavra pode ser usada?

Em uma definição mais recente o linguista Daller (2007), na introdução de sua obra "*Modelling and Assessing Vocabulary Knowledge*"[5], descreve o CL como um espaço lexical tridimensional no qual cada dimensão representa um determinado conhecimento da palavra, como mostra a figura a seguir:

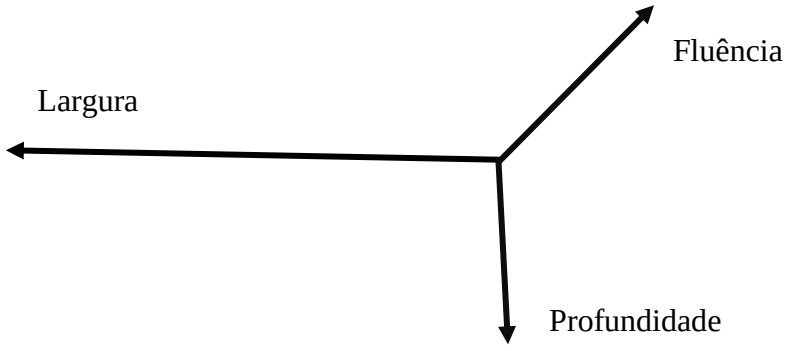


Figura 2 - Espaço lexical: dimensões do conhecimento

[5] Tradução: Modelagem e Avaliação do Conhecimento de vocabulário.

Nesse espaço lexical, a largura representa a quantidade de palavras que o aluno conhece, incluindo seu significado e representação grafo-fônica. A dimensão profundidade está relacionada com o conhecimento do significado da palavra, no que se refere ao seu conceito e funções gramaticais. E, por fim, a fluência que reflete a facilidade e a velocidade de acessar as palavras e usá-las nas diversas práticas comunicativas.

Como pode ser observado, apesar desses três estudos serem desenvolvidos em momentos e contextos diferentes, há uma concordância teórica sobre a relevância de se considerar, ao avaliar um determinado vocabulário, os diferentes níveis/dimensões de conhecimento. Logo, essa concordância proporcionou novos direcionamentos para o desenvolvimento de parâmetros que possibilitam a avaliação e acompanhamento do crescimento do vocabulário dos alunos no contexto escolar.

Todavia, é importante salientar que a mensuração do CL não é um desafio recente, pelo contrário, ele perpassa nas últimas oito décadas, trazendo uma problemática que se repete estudo após estudo: que método deve ser adotado para medir o conhecimento de vocabulário dos alunos, de modo que seja possível compará-lo e classificá-lo?

Na tentativa de desenvolver uma métrica abrangente e confiável para resolver essa problemática, diferentes campos de conhecimentos como a Linguística, a Matemática, Psicologia e Estatística se uniram e criaram um índice denominado de Riqueza Lexical. A seguir serão apresentadas as principais características desse índice e suas modificações ao decorrer do tempo.

2.2 Riqueza Lexical

Como dissemos, a Riqueza Lexical (RL) é um dos mais importantes índices para a avaliação do conhecimento de vocabulário. Para compreendermos como ele se apresenta hoje, é fundamental conhecermos suas diferentes abordagens e os componentes propostos para sua aplicação.

De acordo com Jarvis (2011), o termo RL foi introduzido por Yule (1944) para se referir ao número de palavras que um autor possui em seu léxico mental. Anos depois, Giraud (1954) define a RL como a variedade de palavras de um vocabulário, ou como sinônimo de diversidade lexical, assim, para medi-la apenas uma categoria de análise era usada: palavras ortograficamente diferentes.

No entanto, ao avaliar a RL em 100 ensaios escritos por alunos do 1º ano do curso de inglês da universidade de Lyons na França a partir de um tema sugerido, sob o título "O que você sugere para melhorar a educação secundária na França[7]?", Aranud (1984) concluiu que a avaliação baseada apenas na diversidade das palavras não conseguia capturar todas as características do texto. Sendo assim, ele propôs a inserção de uma nova categoria de análise: as palavras raras, isto é, que não se apresentam com frequência nos textos, palavras de um vocabulário mais específico e de contexto mais técnico. Logo, esse autor conceitua a RL como sinônimo de raridade lexical.

Na busca de uma avaliação mais abrangente, Linnarud (1973) sugere que para analisar a RL é necessário observar as diferentes características do texto, como a densidade lexical (proporção de palavras lexicais[8]), variação lexical (palavras ortograficamente diferentes) e sofisticação lexical (palavras que normalmente não são esperadas em um certo nível de instrução). Vale ressaltar que essas categorias já existiam, contudo, não eram consideradas na avaliação da RL.

Assim, por muitas décadas buscou-se uma definição consistente, que permitisse uma avaliação mais holística. Entretanto, a tentativa de um consenso teórico só teve início a partir dos anos 2000 (MALVERN, 2004), fomentada pela nova explicação de RL sugerida por Read (2000) em que a define como uma característica multidimensional da escrita composta por quatro indicadores linguísticos:

- **Densidade Lexical (DeL):** refere-se à proporção de palavras com valores lexicais, os chamados "itens lexicais" de um texto (URE, 1971; HALLIDAY, 1985).
- **Diversidade Lexical (DiL):** descrita como a variedade do vocabulário empregada em um texto, oral ou escrito (MCCARTHY e JARVIS, 2007).
- **Sofisticação Lexical (SL):** refere-se ao uso de palavras pouco frequentes no idioma (ARNAUD, 1984; LINNARUD, 1973). As ocorrências dessas palavras também estão associadas à maturidade linguística do escrevente, à medida que ele se torna mais experiente, capaz de usar novas estratégias linguísticas, como por exemplo, o conhecimento de um vocabulário mais específico (MARZANO e PICKERING, 2005) ou de termos técnicos (HARMON, 2009; REHDER et

[7] Em francês: *Que faire pour un concert avec une école secondaire en france*?
[8] Palavras adicionam maior informação ao texto.

al.,1998). Portanto, presume-se que a SL pode refletir um nível de vocabulário mais avançado.

- **Número de Erros (NE):** abrange os problemas ortográficos, erros lexicais e erros de morfologia derivacional (READ, 2000).

Apesar da importância desses quatro indicadores para a avaliação da riqueza lexical, a densidade e diversidade lexical têm se destacado como os mais frequentes nos trabalhos sobre o desenvolvimento linguístico. O uso desses dois indicadores pode ser observado nos estudos que avaliam os estágios do desenvolvimento linguístico de crianças e adolescentes (RODRIGUES, 2008); em investigações sobre a progressão escolar (JOHANSSON, 2009; MARTINS, 2016); em estudos comparativos entre textos orais e escritos (URE, 1971; HALLIDAY, 1985; STROMQVIST et al., 2002; STEGEN, 2007); em trabalhos que avaliam a proficiência da segunda língua (L2) (LEKI e CARSON, 1994; RAIMES, 1985; UZAWA e CUMMINGS, 1989), além de estudos que avaliam a correlação entre a diversidade lexical e fator idade (BERMAN e VERHOEVEN, 2002).

Essa prevalência, na maior parte dos estudos, pode ser justificada pelo fato de serem considerados indicadores linguísticos confiáveis, sobretudo para os estudos que investigam o conhecimento lexical (READ, 2000). Outro fator importante é sua natureza quantitativa que permite a realização de uma análise comparativa entre textos de diferentes comprimentos e gêneros textuais (MARTINS, 2017).

Desse modo, compreendendo as diversas aplicabilidades da densidade e diversidade lexical, sua consolidação da literatura como indicadores confiáveis, assim como esses teóricos, o presente estudo também elegeu esses indicadores para a avaliação da riqueza lexical de manuscritos escolares de alunos recém-alfabetizados.

2.3 Densidade Lexical

A Densidade Lexical (DeL) é um indicador linguístico que representa a proporção de itens lexicais de um texto. Historicamente, sua análise é fundamentada na distinção da função que cada palavra exerce no texto.

De acordo com O'Loughlin (2001), a definição de densidade lexical foi originalmente proposta por Ure (1971) para oferecer uma medida de proporção entre as palavras com propriedades lexicais (substantivos, adjetivos e verbos) e palavras com

propriedades gramaticais (as demais classes gramaticais). Apesar de sua grande relevância para as pesquisas sobre essa temática, o estudo de Ure não aborda sobre os critérios estabelecidos para a diferenciação entre palavras lexicais e gramaticais (O'LOUGHLIN, 2001, p. 101).

Alguns anos depois, o conceito de densidade é reformulado por Halliday (1985) e as palavras ganham novas nomenclaturas “itens lexicais” e “itens gramaticais”. Para ele, o termo item contempla uma gama maior das características lexicais, visto que os itens lexicais podem ser constituídos por mais de uma palavra.

Desta maneira, entendem-se como itens lexicais as palavras que adicionam maior informação ao texto, isto é, durante a produção textual o escrevente precisa fornecer o maior número de informações sobre o tema abordado, e para isso acontecer de maneira explícita, é necessário o uso dos itens lexicais representados pelos substantivos, adjetivos, verbos e os advérbios de modais terminados com o sufixo “mente”. Halliday (1985) ainda acrescenta que esses itens fazem parte de um grupo aberto e infinitamente extensível, em que novas palavras são constantemente adicionadas.

Por outro lado, os “itens gramaticais” são classificados como aquelas palavras que exercem apenas a função de conectar um item lexical ao outro, ou uma sentença a outra, assim, as propriedades dessas palavras estão mais próximas da função gramatical. Logo, são compreendidos como itens gramaticais: artigos, pronomes, numerais, preposições, conjunções e interjeições. Diferentes dos itens lexicais, os itens gramaticais são considerados como membros de um grupo fechado, em que raramente novos componentes são adicionados.

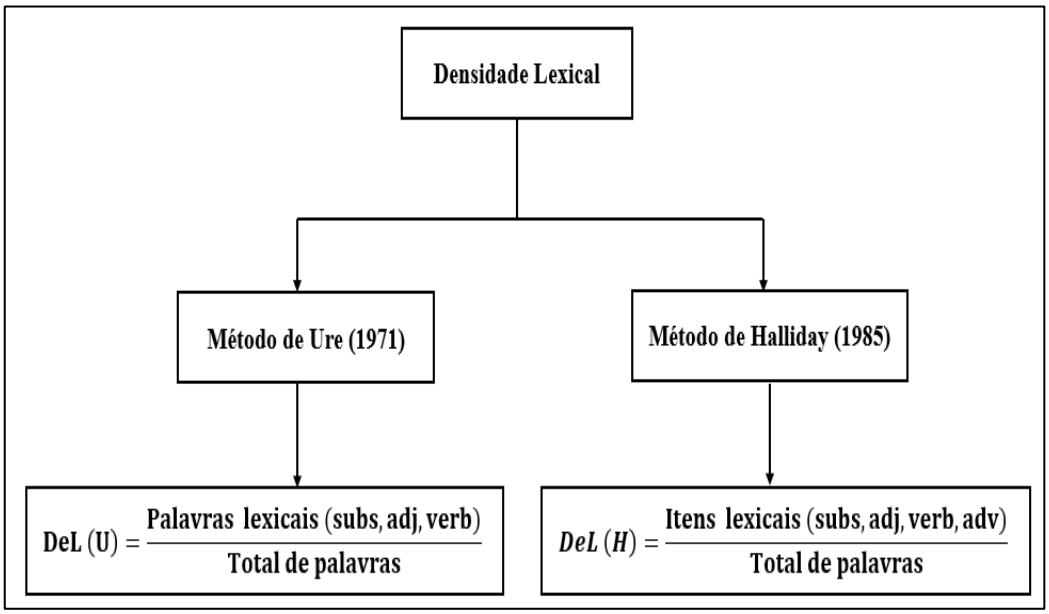
Ainda sobre essa distinção entre itens gramaticais e lexicais, é necessário destacar que Halliday (1985, p.63) defende que há um contínuo do léxico para a gramática, sendo assim, alguns itens não apresentam uma distinção clara, não se encaixam em sua totalidade como lexical, nem como gramatical. Esse é o caso dos advérbios. No entanto, alguns advérbios não seguem esse mesmo comportamento, isto é, os advérbios modais terminados em -mente. Por isso, eles são definidos por esse teórico como um item lexical.

Por conseguinte, a densidade passa a ser definida por Halliday (1985) como a proporção de itens lexicais (substantivos, adjetivos, verbos e advérbios modais terminados em mente) de um texto, sendo mensurada pela divisão dos itens lexicais pelo total de palavras escritas. Sua medição é normalmente composta por três etapas: i) classificação e seleção dos itens lexicais, levando-se em conta todas as suas ocorrências, ou seja, se forem escritas cinco palavras “livro”, elas serão contabilizadas como cinco

itens lexicais; ii) quantificação dos itens lexicais e total de palavras escritas no texto; iii) divisão entre itens lexicais e total de palavras. Os resultados obtidos por essa operação podem assumir valores entre 0 a 1, ou em porcentagem de 0 a 100%.

Em síntese, os principais métodos para avaliação da densidade lexical de um texto foram propostos por Ure (1971) e Halliday (1985), representados na figura abaixo.

Figura 3 - Métodos de mensuração da Densidade Lexical



Fonte: Elaborado pela autora com base em Ure (1971) e Halliday (1985).

Independentemente dos parâmetros adotados para dimensionar a Densidade Lexical em uma produção textual, tanto o estudo de Ure (1971), quantos os de Halliday (1985, 1989) destacam que, em geral, os textos escritos possuem maior densidade que os textos orais.

Ure (1971) mediu a densidade em diferentes registros a partir de um *corpus* composto por 30 textos orais extraídos de diferentes situações comunicacionais como: discussão entre alunos, conversas informais, entrevistas de rádio e 30 textos escritos de diferentes gêneros textuais, como narrativas ficcionais, contos, relatórios, receitas, manuais, matérias de jornais e trabalhos científicos.

Os resultados desse estudo indicaram que a maioria dos textos orais obtiveram taxas de densidade menores que 40%, enquanto os textos escritos obtiveram taxas maiores que 40%, sugerindo que os textos escritos possuem uma maior frequência de itens lexicais. Logo, esse autor sugere que a densidade lexical pode ser afetada pelo

planejamento da atividade, isto é, o texto escrito pode ser preparado, corrigido e revisado ao longo de sua produção.

Sobre essa diferença entre as modalidades oral e escrita, Martins (2016, p.47) defende que "um texto com baixa densidade lexical se configura como mais dêitico, mais situado no aqui-agora do discurso, sendo comumente um texto com características da modalidade falada. Enquanto que o texto que apresenta taxas mais altas de densidade lexical estará mais próximo da modalidade escrita". Sendo assim, devido a sua alta concentração em textos escritos, a densidade vem sendo considerada um dos grandes diferenciadores entre essas duas modalidades.

2.4 Diversidade Lexical

As investigações voltadas para a quantificação das palavras a partir da análise de texto, inicialmente se concentravam apenas em determinar a taxa de repetitividade das palavras (THOMPSON, 1915; ZIPF, 1935). Anos depois, essas investigações ganham outros direcionamentos e uma nova abordagem é proposta pelo psicólogo Carroll (1938), baseando-se na análise de duas categorias: número de palavras diferentes (considerando a primeira ocorrência de cada palavra) e o número de palavras repetidas (a partir de sua segunda ocorrência).

A Diversidade Lexical (DiL) é definida por Carroll (1938, p. 379) como "a quantidade relativa do vocabulário encontrada no uso da linguagem produtiva de um falante ou escritor" medida pela razão entre número de palavras diferentes (*types*) e total de palavras escritas no texto (*tokens*). De acordo com esse autor, a DiL pode ser considerada como um forte parâmetro para estimar a extensão do vocabulário de um escrevente/falante a partir de suas produções textuais.

A medição da DiL se fundamenta no princípio da diferenciação entre duas unidades de análise, *types* e *tokens.* Sendo *types*, as palavras ortograficamente diferentes contidas no texto, considerando-se apenas sua primeira ocorrência. Os *tokens* são todas as palavras escritas no texto, porém, considerando-se todas as suas ocorrências. Por exemplo, na frase "Eu tenho um livro, **um livro** de contos" temos um total de oito palavras escritas (*tokens*), dessas oito, seis são ortograficamente diferentes (*types*) e duas são repetidas (em negrito).

Nesse sentido, o número de *types* pode ser considerado como o representativo do tamanho do vocabulário ativo usado durante a produção textual. Assim, uma alta taxa de

types reflete em um maior alcance lexical, enquanto que uma baixa taxa pode indicar um repertório de palavras limitado ou, até mesmo, uma certa dificuldade na linguagem (MILLER, 1981).

2.5 Medidas para mensuração da DiL

De acordo com Jarvis (2013), a primeira proposta para estimar o tamanho do vocabulário, baseado em padrões de escrita, foi desenvolvida pelo linguista Zipf (1935), através da criação da "lei de Zipf". Essa lei defende que a frequência das palavras de um texto não é aleatória, mas determinada por um princípio de proporcionalidade representado pela seguinte fórmula:

$$P(n) \cong \frac{1}{n^a} \qquad (1)$$

Em que "P(n)" representa a proporção da palavra que ocupa a n-ésima posição na tabela de distribuição de frequências /ocorrências de palavras. O expoente "a" é um fator com valor aproximadamente unitário. Nessa fórmula, o número de ocorrência das palavras é representado pela proporção 1, 1/2, 1/3, 1/4 e assim sucessivamente.

Para uma melhor compreensão sobre essa aplicação, suponha que se queira conhecer o ranking das quatro palavras mais frequentes de um determinado texto escrito com um total de 200 palavras. Ao analisar as palavras, identifica-se que a maior ocorrência foi a palavra "livro", sendo escritas 20 vezes no texto. Logo, essa palavra ficaria em primeiro lugar no ranking.

Esse primeiro número de ocorrência (20) auxiliará na identificação das próximas palavras do ranking. Assim, para determinar a segunda palavra de maior ocorrência, bastaria aplicar o princípio de proporcionalidade de 1, 1/2, 1/3, 1/4 em que:

- A palavra mais frequente do texto = 20 ocorrências
- A segunda palavra mais frequente equivale a 1/2 do número de ocorrências da palavra mais frequente no texto, ou seja, 1/2 de 20, que representa aproximadamente 10 ocorrências.
- A terceira palavra mais frequente do texto equivale a 1/3 de 20, sendo aproximadamente 7 ocorrências.
- A quarta palavra mais frequente do texto equivale a 1/4 de 20, correspondendo aproximadamente a 5 palavras.

Portanto, de acordo com seu número de ocorrências, as palavras seriam ranqueadas do seguinte modo:

Tabela 1- Ocorrências de palavras

Ranking	Número de Ocorrências
1°	20
2°	10
3º	7
4º	5

Zipf (1935) também sugere que o número de ocorrência no texto pode ser influenciado por dois fatores: i) pela importância que a palavra possui, dito de outro modo, quanto mais influente for a palavra para a construção do sentido do texto, maior será o seu número de ocorrência; ii) pelo comprimento, quanto menor a palavras[11], maior ocorrência ela terá.

Baseado em Zipf (1935), o psicólogo Carroll (1938) propõe que a diversidade lexical de um texto seja mensurada através da proporção do número de palavras diferentes pelo total de palavras escritas. No entanto, ao analisar essa proporção em um conjunto de textos, Carroll observou que os resultados demonstravam uma certa dependência em relação ao comprimento dos textos.

Essas medidas inicialmente propostas por Zipf (1935) e Carroll (1938) trouxeram um novo olhar para as pesquisas sobre vocabulário, especialmente no que se refere a sua diversidade lexical, como também para a expansão nos estudos quantitativos com base na análise do texto, que por sua vez passa a ser considerado um instrumento de grande potencialidade para a comparação e avaliação do vocabulário em seus mais diversos contextos de uso.

Esses estudos inaugurais ajudaram a promover novos resultados cada vez mais confiáveis sobre a DiL. A seguir serão apresentadas as principais medidas para mensuração da diversidade de acordo com sua ordem cronológica.

2.5.1 *Type-Token Ratio* (TTR)

[11] Referindo-se aos itens gramaticais.

Type-Token Ratio (TTR) desenvolvida por Johnson (1944) é considerada uma das medidas mais influentes e recorrente para a avaliação da diversidade lexical, sendo definida como:

> Uma medida da "flexibilidade" ou variabilidade do vocabulário projetada para indicar certos aspectos da adequação da linguagem, expressa pela proporção de diferentes palavras (types) para o total de palavras (tokens) em uma determinada amostra de idioma (JOHNSON, 1944, p.1).[12]

Em síntese, a TTR é representada pela seguinte fórmula:

$$TTR = \frac{Types}{Tokens} x100 \qquad (2)$$

Os resultados dessa operação matemática são demonstrados no intervalo de 0 a 100%. No entanto, notou-se que o problema inerente ao tamanho do texto também afetava os valores obtidos pela TTR, provocando uma limitação para o uso desta. Esse fenômeno é detalhadamente explicado por Jarvis e Mccarthy (2010, p.10):

> À medida que um texto se torna mais longo (ou seja, adicionando novas palavras a ele), o número de palavras gerais no texto (ou tokens) aumenta. Mas, embora o aumento de *token* seja linear (uma nova palavra, um novo token), a taxa de aumento do número de palavras diferentes no texto (ou *types*) diminui constantemente. A desaceleração do aumento de *type* ocorre porque, com cada nova instância de um *token* há uma diminuição correspondente na probabilidade de um novo *type.*[13]

Logo, o distanciamento entre o número de *types* e *tokens* à medida que o texto aumenta, ocasiona uma queda nos valores da TTR. Esse fenômeno provocou algumas implicações para o desenvolvimento de análises comparativas entre textos de diferentes extensões. Para superar essa deficiência, Johnson (1944) sugere uma pequena alteração em sua aplicação, através do uso de uma TTR média segmentar:

> As TTRs para amostras de diferentes magnitudes podem ser comparadas dividindo cada amostra em segmentos do mesmo tamanho, digamos, 100 palavras cada, calculando a TTR de cada segmento e calculando a média das TTRs segmentadas de cada amostra. Pode-se presumir com segurança que esses TTRs segmentados são diretamente comparáveis, desde que representem

[12] Versão original: Type-Token Ratio (TTR) is a measure of vocabulary "flexibility" or variability, designed to indicate certain aspects of language adequacy. It expresses the ratio of different words (types) to total words (tokens) in a given language sample.

[13] Versão original:Text becomes longer (that is, adding new words to it), the number of general words in the text (or tokens) increases. But while the token increase is linear (a new word, a new token), the rate of increase in the number of different words in the text (or types) decreases steadily. The increase in type deceleration occurs because each new instance of a token has a corresponding decrease in the likelihood of a new type.

> segmentos de tamanho igual, e que os meios desses TTRs segmentados também sejam diretamente comparáveis (JOHNSON, 1944, p. 2).[14]

Nesse sentido, ao analisar um conjunto de textos de diferentes comprimentos, a alternativa seria dividi-los em amostras com tamanhos similares, por exemplo: como comparar a TTR de dois textos, sendo o primeiro com 200 palavras e o segundo com 150 palavras? Nesse caso, bastaria dividir os dois textos em amostras de 50 palavras, e posteriormente extrair a TTR de cada uma e, por fim, calcular a média final da TTR e compará-las. Porém, esse tipo de procedimento exigiria uma grande demanda tempo, visto que a amostra a ser segmentada poderia variar de acordo com o número de palavras escritas.

Outra solução para a TTR seria estabelecer um ponto de corte de palavras. Sendo assim, em uma comparação entre um texto de 300 palavras e um de 250 palavras, para deixá-los com tamanhos similares seria necessário cortar 50 palavras do texto maior, assim ambos ficariam com 250 palavras. No entanto, essa alternativa também poderia gerar alguns problemas para a validade dos resultados, devido à adoção de diferentes números de cortes para cada texto.

Além disso, outra dificuldade estaria em como estabelecer os critérios para justificar a parte cortada: Quais palavras seriam cortadas? O recorte seria no início do texto, no meio ou nos três últimos parágrafos? Ou seria de maneira aleatória? Como garantir que a parte recortada não possui a maior diversidade lexical? Essa metodologia, assim como a anterior, também não garantiria a realização de uma análise confiável. Logo, como também aponta Mccarthy (2005), a TTR segmentar não seria uma boa opção metodológica para medir a DiL.

2.5.2 RTTR e CTTR

Cientes de que o grande problema da TTR era o crescimento acelerado do número de tokens, outras abordagens foram propostas na tentativa de modelá-lo. A primeira opção seria a raiz da razão entre *types* e *tokens (Root Type-Token Ratio* - RTTR),

[14] Versão original: TTR's for Samples of different magnitudes can be made comparable by dividing each sample into like-sized segments of, say, 100 words each, computing the TTR for each segment and then averaging the segmental TTR's for each sample. It can' be safely assumed that such segmental TTR's are directly comparable, so long as they represent segments of equal size, and that means of such segmental TTR's are also directly comparable.

conhecida como “índice de Guiraud” (GUIRAUD, 1954), representada pela seguinte fórmula:

$$RTTR = \frac{types}{\sqrt{tokens}} \quad (3)$$

Seguindo esse enfoque Carroll (1964) propõe a C*orrected Type-Token Ratio* (CTTR) ou “razão de *types* e *tokens* corrigida” representada pela fórmula:

$$CTTR = \frac{types}{\sqrt{2\ tokens.}} \quad (4)$$

Como se pode notar, as duas propostas recorrem ao uso da raiz quadrada como tentativa de desacelerar o aumento constante de *tokens* no texto. Porém, ao avaliar essas duas medidas Malver (2004) observou que os valores de *tokens* permaneciam aumentando ao longo das primeiras centenas de palavras. Ele também identificou que os textos com alto número de palavras possuíam uma tendência de queda no valor da diversidade lexical. Entretanto, por se tratar de uma queda lenta, e muitas vezes quase imperceptível, essas medidas acabavam gerando uma falsa ilusão na consistência de seus valores.

Na literatura não identificamos a quantidade exata do número de palavras que ocasiona essa queda. Partindo do princípio da necessidade de um índice que permita uma comparação confiável entre textos, independentemente de seu tamanho, a RTTR e o CTTR acabaram por manter a mesma deficiência da TTR. Esse aspecto torna essas medidas pouco confiáveis para esse tipo de análise (COVINGTON e MCFALL, 2010; MALVERN, 2004). Apesar dessas restrições, é possível observar que alguns trabalhos ainda têm aplicado TTR para examinar a DiL (TEMPLIN,1957; SCHERER e SOUZA, 2011; RIFFO, 2019).

2.5.3 Medida D e VOCD

As problemáticas envolvendo essas medidas para mensuração da DiL, fomentou o desenvolvimento de novos trabalhos, desta vez, embasados em uma modelagem matemática e computacional mais eficiente (MALVERN e RICHARD, 1997; MCCARTHY e JARVIS, 2007, 2010).

Nesse cenário, surge a “medida D” desenvolvida por Malvern e Richard (1997) baseada em um modelo matemático mais avançado, ajustando a queda da curva da TTR, isto é, a curva que indica a diminuição de *types* à medida que o tamanho do texto aumenta. Assim, sua fórmula é determinada pela seguinte equação:

$$\mathrm{TTR} = \frac{\mathrm{D}}{\mathrm{N}}\left[\left(1 + 2\frac{\mathrm{N}}{\mathrm{D}}\right)^{\frac{1}{2}} - 1\right] \qquad (5)$$

Sendo:

TTR representa a relação entre *types e tokens*

N é o total e número de *tokens*

D é o parâmetro de ajuste da TTR

Diante de sua complexidade, a medida D logo sofre algumas reformulações e passa a ser aplicada com o auxílio do programa VOCD desenvolvido por McKee (2000). Sendo assim, para medir a DiL, o VOCD retira do texto transcrito algumas amostras aleatórias de palavras, começando com amostras de 35 tokens, 36, 37 e, assim sucessivamente, até chegar em 50 tokens. Para cada uma dessas amostras, a TTR é calculada. Ao final, o programa calcula uma média geral de todas as TTRs. Essa média geral passa por um ajuste através da medida D, impedindo que o resultado final seja afetado pelo tamanho da amostra.

Entretanto, umas das principais exigências do VOCD é a quantidade mínima de palavras do texto. Assim, para obter resultados mais precisos sobre sua taxa de diversidade, o texto deve possuir um mínimo de 50 palavras (MCKEE, 2000).

A fim de avaliar o nível de precisão de todas essas medidas apresentadas, Jarvis (2002) desenvolveu um estudo em que as aplicou para mensurar a diversidade lexical em narrativas escritas por 140 alunos de língua finlandesa, 70 de língua sueca, e 66 com língua nativa inglês estudantes do 5º, 7º, 9º ano do ensino básico, com idade de 10, 12 e 14 anos. Seus resultados indicaram que dentre as medidas usadas, a D foi quem forneceu os valores mais precisos das curvas de TTR em mais de 90% dos textos, oferecendo uma maior resistência em relação ao comprimento dos textos.

Tendo isso em vista, diversos teóricos têm adotado o VOCD para avaliar a diversidade lexical de textos de alunos de diferentes nacionalidades: holandês, islandês, espanhol, hebraicos (BERMAN e VERHOEVEN, 2002); sueco (JOHANSSON, 2009);

chinês (WANG, 2014); iraniano (SADEGHI e DILMAGHANI, 2013); francês (TREFFERS-DALLER , 2013) e português europeu (MARTINS, 2016).

Esses estudos, apoiados nos resultados de Jarvis e McCarthy, defendem a mensuração da diversidade baseada em modelagem matemática como uma alternativa confiável para aplicação em investigações que visem comparar textos de diferentes tamanhos.

Em consonância com esses autores, e compreendendo a validade da medida D, aplicada ao modelo (VOCD) para obtenção das informações sobre as características lexicais de um vocabulário, nossa investigação também elegeu esse método para analisar a DiL em produções textuais de alunos brasileiros. Para concluir esta seção, apresentamos no quadro abaixo uma síntese geral das métricas discutidas.

Quadro 2 - Medidas para avaliação do vocabulário

Autor	**Nome**	**Ano**	**Fórmula**
Zipf	Frequência de palavras	1935	$P(n) \cong \frac{1}{n^a}$
Johnson	TTR	1944	$TTR = \frac{Types}{Tokens} x100$
Guiraud	RTTR	1954	$RTTR = \frac{Types}{\sqrt{Tokens}}$
Carroll	CTTR	1964	$CTTR = \frac{types}{\sqrt{2\ tokens.}}$
Malvern e Richards	Medida D	1997	$\mathrm{TTR} = \frac{\mathrm{D}}{\mathrm{N}}\left[\left(1+2\frac{\mathrm{N}}{\mathrm{D}}\right)^{\frac{1}{2}}-1\right]$
McKee Malvern, Richards	VOCD	2000	Ajuste de curva e modelagem

3 AVALIAÇÃO DA DENSIDADE E DIVERSIDADE LEXICAL EM TEXTOS ESCOLARES

Nesta seção descrevemos alguns trabalhos que usaram a Densidade e Diversidade Lexical para avaliar a produção textual de alunos da educação básica. Em seguida, trazemos uma discussão acerca desses estudos e dos fatores que podem contribuir para a densidade e diversidade lexical de um texto.

3 .1 Densidade e Diversidade Lexical em produções textuais de alunos de diferentes nacionalidades

A escrita textual é uma atividade que exige dos escreventes conhecimentos de ortografia, de gramática e conhecimento lexical, adquiridos através de sua participação nas mais diversas práticas comunicativas, seja no âmbito escolar, de forma sistematizada ou nas práticas comunicativas informais do cotidiano (KOCH e ELIAS, 2010). Com a exigência de tais conhecimentos, a produção de um texto torna-se uma atividade desafiadora.

É de salientar que a produção textual requer, entre outros itens, uma seleção lexical adequada ao assunto exposto, ao gênero textual e ao seu grau de formalidade (BEZERRA, 2000). Assim sendo, podemos considerar o conhecimento lexical como um fator essencial para a composição de um texto. Partindo dessa premissa, alguns estudos vêm investigando o conhecimento lexical de alunos de diferentes nacionalidades, idades, e níveis escolares, a partir da avaliação de duas medidas lexicais: densidade e diversidade lexical, tomando como objeto de investigação suas produções textuais (JOHANSSON, 2009; BERMAN, 2010; WOERFEL e YILMAZ, 2011, RODRIGUES, 2010; MARTINS, 2016, SANTOS e CALIL, 2018).

Em um estudo de grande abrangência, envolvendo alunos de diferentes idiomas: inglês, espanhol, hebraico, francês, islandês, sueco e holandês, Berman (2010) avaliou como a diversidade lexical se apresentava em textos narrativos e expositivos em duas modalidades diferentes (fala e escrita). Nesse estudo, foram analisados textos produzidos por alunos de quatro níveis escolares: anos iniciais do ensino fundamental, anos finais do ensino fundamental, ensino médio e ensino superior.

Os resultados para a diversidade lexical, calculados através da medida D, indicaram uma diferença significativa entre os gêneros em todos os idiomas. Os textos do

gênero expositivo obtiveram maior diversidade que os textos dos gêneros narrativos. Também foi constatado que os níveis de escolaridade influenciavam na diversidade lexical, uma vez que o vocabulário dos textos dos alunos do ensino médio foi mais diversificado que dos alunos do ensino fundamental. Todavia, em relação à modalidade (oral ou escrita), não foi identificado nenhum impacto sobre a diversidade.

Os resultados de Berman foram posteriormente apoiados pelos resultados do estudo de Woerfel e Yılmaz (2011) quando analisou produções escritas de textos narrativos e expositivos de alunos alemães do ensino médio e identificou que tanto o cumprimento dos textos, quanto a medida da diversidade lexical diferem de acordo com a idade e o gênero textual.

Seguindo a mesma temática, Johansson (2009) examinou o desenvolvimento lexical de alunos suecos através da densidade e diversidade lexical de seus textos. Assim, foram avaliados 316 produções textuais escritas e orais, dos gêneros narrativos e expositivos de alunos de quatro grupos etários: 10 anos (4ª ano), 13 anos (7ª ano), 17 anos (2º ano do ensino médio) e adultos (universitários).

Os resultados da densidade lexical, mensurada através do método de Halliday (1985), indicaram que se tratando dos textos narrativos orais, a taxa da densidade foi semelhante em todas as faixas etárias analisadas, ou seja, os adultos usaram a mesma proporção de itens lexicais que os alunos de 10 anos. Na modalidade escrita, nos textos narrativos foi identificada uma diferença apenas entre os alunos de 10 e 17 anos, e entre os alunos de 13 anos e os adultos universitários. Nos textos expositivos orais, os resultados apontaram para uma densidade maior para os adultos. Segundo Johansson (2009), nos textos expositivos orais os adultos usaram mais itens lexicais do que todas as outras faixas etárias. Já nos textos expositivos escritos, não foram encontradas diferenças significativas entre adultos universitários e jovens de 17 anos, indicando uma semelhança na densidade entre essas faixas etárias.

Logo, podemos considerar que uma alta taxa de densidade está diretamente relacionada ao conhecimento de vocabulário que o escrevente/falante possui, pois quanto maior for o seu domínio, menor será o uso de palavras repetidas e palavras sem valores lexicais (artigos, pronomes, preposição etc.). Em outras palavras, o domínio lexical possibilita um maior uso de palavras lexicais permitindo a expansão das orações presentes no texto de ambas as modalidades, contribuindo eventualmente para um texto mais informativo.

Todavia, diferentemente dos resultados de Berman (2010), no estudo de Johansson (2009) não foi detectada diferença significativa entre os textos narrativos e expositivos, em ambas as modalidades. Entretanto, um ponto incomum entre esses dois estudos é a indicação da idade e do nível escolar como fatores que influenciam na diversidade do texto.

Com o propósito de investigar a diversidade do vocabulário de alunos americanos em textos de diferentes gêneros, Olinghouse e Wilson (2012) desenvolveram uma pesquisa com 105 alunos de seis turmas do 5º ano de uma escola pública de ensino fundamental, localizadas em um distrito urbano no sudeste dos Estados Unidos. Esses autores avaliaram a diversidade em textos narrativos, persuasivos e informativos, desenvolvidos a partir do tema "espaço sideral".

Os valores da diversidade extraídos desses textos demonstraram uma média de 61,67% para os narrativos, 61,01% para os persuasivos e 52,76% para os descritivos. Assim, esses resultados mostram que dos três gêneros textuais produzidos pelos alunos americanos, os textos narrativos foram os que obtiveram maior extensão de vocabulário.

Seguindo essa mesma perspectiva a partir de diferentes gêneros textuais, Sadeghi e Dilmaghani (2013) avaliaram a sensibilidade da diversidade em relação aos gêneros argumentativos, narrativos e comparativos, a partir da análise de 90 produções textuais escritas por 30 alunos iranianos do ensino médio. Os resultados evidenciaram que os textos argumentativos obtiveram maior diversidade lexical (87,16), seguido pelos textos comparativos (82,80) e textos argumentativos (80,23).

Sadeghi e Dilmaghani (2013) finalizaram esse estudo trazendo algumas indicações importantes. Segundo eles, o domínio do tema e do gênero textual a ser produzido são fatores que podem influenciar no uso de um vocabulário mais diversificado, logo um ensino mais orientado para o vocabulário, contribui para o aumento da diversidade lexical e, consequentemente, evita o uso excessivo de repetições de palavras.

Se aproximando do nosso idioma, português brasileiro, podemos citar os trabalhos de Rodrigues (2010), Martins (2016), Santos e Calil (2018) que analisaram esses indicadores no idioma do português europeu. Nesses três estudos, a densidade e diversidade foram avaliadas a partir de produções escritas de alunos portugueses da educação básica.

Rodrigues (2008) examinou o desenvolvimento das capacidades de composição de alunos portugueses através do total de palavras escritas, diversidade lexical e densidade

lexical de seus textos. Participaram desse estudo 227 alunos, de ambos os sexos e com faixa etária de 6 a 11 anos. Cada aluno escreveu 2 narrativas a partir de temas previamente definidos (1ª O Carro que Queria Ser Táxi; 2ª Uma Aventura na Quinta) sem restrição de tempo, nem de número de páginas ou palavras.

Os resultados mostraram que em termos de produtividade, os alunos escreveram seus textos com uma média de 101 palavras (1º ano), 108 palavras (2º ano), 214 palavras (3º ano) e 201 palavras (4º ano). Como podemos observar, há uma aproximação entre o número de palavras dos alunos do 1º e 2º ano, e dos alunos 3º e 4º. Porém, quando comparamos os alunos do 1º ano com os do 3º e 4º ano, é notável um aumento do número de palavras.

Em relação à diversidade, os alunos obtiveram as seguintes médias: 55% (1º ano), 68% (2º ano), 57% (3º ano) e 64% (4º ano) e na densidade, eles escreveram seus textos com média de 55% (1º ano), 52% (2º ano), 50% (3º ano) e 51% (4º ano). Os resultados desse estudo indicaram uma diferença significativa entre as médias da diversidade em todos os anos escolares, exceto entre o 1º e 3º. Entretanto, na densidade não foi identificada diferença significativa entre os textos dos alunos dos quatro níveis de escolaridade analisados.

A correlação entre os níveis escolares, densidade e diversidade lexical, também foi tema de uma investigação realizada por Martins (2016). Esse autor examinou produções textuais de 122 alunos monolíngues de português europeu matriculados no 5º, no 7º e no 10º ano, representando cada um dos três ciclos da escolaridade básica em Portugal. Cada aluno produziu dois textos, sendo um registro narrativo (narrando um fato marcante real ou imaginado) e um registro argumentativo (As redes sociais são importantes?), resultando em um *corpus* de 244 textos.

Os resultados indicaram as seguintes médias de densidade para os textos narrativos: 44,47% para os alunos do 5º ano, 43,93% para os alunos do 7º ano e 43,94% para os alunos do 10º ano. Quanto à diversidade, as médias obtidas pelos alunos dos três anos escolares foram superiores a 78%. Nos textos argumentativos as médias de densidade foram maiores: 45,16% para os alunos do 5º ano, 47% para os alunos do 7º ano e 46,85% para os alunos do 10º ano. Em relação à diversidade lexical, as médias obtidas pelos alunos dos três anos escolares foram superiores a 65%.

Segundo os resultados de Martins (2016), o teste de correlação entre progressão escolar, densidade e diversidade indicaram uma correlação positiva apenas entre a diversidade e a progressão escolar, em ambos os registros para os alunos do 5º ao 10º,

sugerindo que o vocabulário dos alunos aumenta de acordo com sua progressão escolar. Entretanto, esse autor sugere que esse crescimento não é constante entre os três anos analisados, visto que não foi revelada diferença significativa na diversidade lexical entre o 5º e o 7º ano, nem no registro narrativo, nem no registro argumentativo. Em relação à densidade, não foi detectada qualquer correlação com a progressão escolar, sugerindo que o avanço dos alunos nos anos escolares não traz implicações no uso das palavras de valores lexicais.

Comparando os resultados dos estudos de Rodrigues (2010) e Martins (2016) podemos encontrar alguns pontos incomuns como o crescimento da diversidade lexical das produções textuais dos alunos ao decorrer da progressão escolar. Em contrapartida, ambos autores destacam que esse crescimento é melhor observado quando analisado em textos de alunos com idades e níveis escolares distantes. Outro ponto observado é a diferença entre os valores das médias de diversidade e densidade. No geral, as médias de diversidade foram mais altas que as médias da densidade, para 5 dos 6 níveis escolares analisados, indicando não haver uma similaridade entre a proporção de densidade e diversidade dos textos.

Adotando uma abordagem metodológica diferente, Santos e Calil (2018) analisaram a densidade e diversidade lexical em produções textuais de alunos portugueses do 2º ano do ensino básico a partir de uma proposta metodológica de escrita individual e colaborativa. Compreende-se, em nosso estudo, como escrita colaborativa a situação didática proposta por Calil (2009), quando dois alunos dialogam e escrevem um único texto.

Para isso, foram avaliadas 60 narrativas ficcionais escritas por 12 alunos portugueses, sendo 36 narrativas escritas a partir de 6 produções colaborativas em dupla em que os alunos combinavam e escreviam um único texto, e 24 narrativas escritas a partir de 2 produções individuais. Em síntese, cada aluno escreveu 8 produções, sendo a 1ª (pré-teste) e a 8ª (pós-teste) individuais e da 2ª a 7ª em dupla.

Na comparação das médias gerais entre textos individuais e colaborativos, foi encontrada uma média de densidade de 55,5% para os textos individuais, e 54,4% para os textos colaborativos. Na diversidade, os textos individuais obtiveram média de 54,20% e os textos colaborativos 53,9%. Além disso, também foi observado que os textos escritos em díades apresentaram um equilíbrio maior, ou seja, foram escritos com uma proporção de densidade e diversidade próxima, o que pode indicar uma maior modalização das palavras usadas nesses textos.

Comparando os resultados da densidade e diversidade no pré-teste (1ª produção individual) e pós-teste (8ª produção individual), foi constatado um aumento apenas para a densidade lexical, sugerindo que após serem submetidos às produções colaborativas, o número de itens lexicais usado pelos alunos em suas produções textuais individuais aumentou. Por outro lado, na diversidade foi observado um comportamento contrário, ou melhor, uma diminuição.

Voltando nossa atenção para o contexto brasileiro, percebemos uma escassez de trabalhos que investigam a densidade e diversidade lexical em textos escolares, principalmente, de alunos dos anos iniciais do ensino fundamental. Analisando algumas bases de dados (Periódicos CAPES, SCIELO, SciVerse Scopus, BDTD, *ANPEd)*, encontramos apenas três estudos: Costa (1997) e Scherer (2002, 2011).

Costa (1997), em sua pesquisa de mestrado, analisou a densidade lexical em 116 textos dissertativos escritos por alunos do ensino fundamental (4º e 8 anos) e do ensino médio (2º ano) de uma escola pública estadual de Fortaleza. Usando o método de Halliday (1989) para mensurar a densidade, essa autora identificou que não houve diferença significativa entre os textos dos alunos do 4º e 8º ano ensino fundamental, concluindo que "o trabalho da escola tem contribuído muito pouco para ampliar o acervo lexical desses alunos do ensino fundamental, pois, em termos de densidade lexical, o progresso que se verificou foi muito pequeno, principalmente, considerando que a diferenças entre as duas turmas analisadas não foi apenas de um ano escolar, mas de quatros anos"(COSTA,1997, p.75).

Todavia, na comparação entre os textos dos alunos do ensino fundamental e médio, os resultados apontaram para uma diferença significativa, tanto entre os alunos do 4º e 2º ano (médio), como entre os alunos do 8º e 2º ano (médio), ou seja, os textos do ensino médio foram escritos com um número maior de itens lexicais.

Seguindo uma perspectiva de investigação focalizada nos textos orais, Scherer (2002, 2011) examinou a diversidade lexical a partir de registros orais da fala espontânea de crianças brasileiras com faixa etária de 1 a 5 anos. Em seu estudo inicial, Scherer (2002) buscou traçar um perfil evolutivo da diversidade lexical do vocabulário de crianças de 3 a 5 anos através da relação entre *type/token*. Para tanto, foram analisadas gravações de fala espontânea de 60 crianças brasileiras em creches e/ou na residência das crianças.

A coleta foi realizada pela pesquisadora e pelas mães das crianças durante atividades lúdicas, com duração de 45 minutos, resultando em um c*orpus* de 50 enunciados. Para analisar a diversidade, os enunciados foram transcritos e,

posteriormente, aplicaram a medida TTR para mensurar a diversidade. Os resultados mostraram que a TTR das crianças de 4 anos (55%) foi superior à das crianças de 3 anos (48%) e de 5 anos (44%).

Dando continuidade a essa investigação inicial, Scherer (2011) desenvolveu um novo estudo visando analisar a fala espontânea de crianças de uma faixa etária menor, assim, participaram desse trabalho 60 crianças com idade de 1 ano e 6 meses, 2 anos e 2 anos e 6 meses. Aplicando a mesma metodologia, a autora calculou a relação de *type/tokens*, o que resultou em uma TTR de 44% paras enunciados das crianças de 1 ano e 6 meses, 46% de para as crianças de 2 anos e 46% para as de 2 anos e 6 meses.

Em síntese, os resultados dos estudos de Scherer (2002, 2011) demonstraram que dentre as faixas etárias analisadas, o vocabulário das crianças de 4 anos foi o mais diversificado, enquanto as crianças de 5 anos obtiveram medidas de TTR similares as crianças de 1 ano e meio. Entretanto, a autora destaca que foi observado uma sensibilidade da TTR em relação ao comprimento do enunciado, pois à medida que os enunciados dos alunos aumentavam, a TTR diminuía.

Todavia, quando voltamos nosso olhar para os textos de alunos brasileiros, muitas lacunas ainda se encontram abertas acerca da densidade e diversidade. Por exemplo: com quantas palavras esses alunos escrevem seus textos? Qual é a média de densidade e diversidade lexical usada em suas produções? Há diferença significativa entre os alunos, em relação a esses indicadores? É em busca dessas respostas que a presente dissertação foi desenvolvida. Para contribuir para uma maior compreensão acerca dessas características lexicais e para possibilitar novos caminhos metodológicos para o ensino da produção textual em sala de aula.

3.3 Fatores que podem contribuir para a densidade e diversidade lexical de um texto

Através dos estudos apresentados nesta seção, é possível compreender a relevância da densidade e diversidade lexical como facilitadores para o acesso aos conhecimentos lexicais dos alunos, através da análise das palavras contidas em suas produções textuais. Desse modo, podemos pontuar alguns fatores que podem contribuir para o enriquecimento do vocabulário dos alunos como gênero textual, nível escolar, idade

O gênero textual é considerado como um fator de grande impacto na diversidade e densidade lexical (OLINGHOUSE E WILSON, 2012; SADEGHI e DILMAGHANI, 2013). Isso sugere que os valores obtidos através das medidas de densidade e de diversidade podem auxiliar os professores a diagnosticar em quais gêneros textuais os alunos sentem mais dificuldades em suas composições escritas, implicando assim, na possibilidade de um ensino de produção textual mais direcionado.

Outro fator de grande impacto, principalmente para a diversidade lexical, é o nível escolar e a idade dos alunos, como indicam os trabalhos de Johansson (2008), Berman (2010), Rodrigues (2010) e Martins (2016). Esses quatro estudos destacaram que à medida que os alunos se tornam mais proficientes em sua língua materna, mais diversificado seu vocabulário se torna.

Entretanto, um ponto crucial a se entender sobre a densidade e diversidade durante essa progressão escolar é que seu crescimento nem sempre é linear, como foi apontado pelos resultados de Rodrigues (2010) e Martins (2016) que demonstraram que os textos de alunos de idades e anos escolares próximos não apresentam diferença significativa em termos de densidade e diversidade. Logo, o impacto da idade e do nível escolar na diversidade e densidade lexical é melhor visualizado quando comparamos textos de alunos com idades e níveis escolares diferentes (JOHANSSON, 2008).

Como demonstraram os resultados de Santos e Calil (2018), a metodologia de escrita colaborativa também pode favorecer a riqueza lexical dos alunos, proporcionando um equilíbrio entre as taxas de densidade e diversidade lexical de seus textos. Assim, compreendendo a potencialidade desta metodologia, nesta dissertação defendemos que escrever um texto a dois pode oferecer momentos de troca de conhecimento acerca do vocabulário, das estruturas dos gêneros textuais, além de contribuir para a modalização das palavras que serão escritas.

Nesse contexto, a escrita a dois promove momentos privilegiados de aprendizagem entre os alunos, pois, como destaca Felipeto (2019, p. 135) “em situações de escrita colaborativa, a presença do outro pode suscitar um processo de reflexão sobre a linguagem, de forma ainda mais intensa que quando a escrita é realizada individualmente”.

4 METODOLOGIA

Nesta seção, descrevemos a metodologia empregada em nossa investigação, assim como a composição do nosso corpus e os procedimentos adotados para a coleta de dados. Em seguida, apresentamos as categorias investigadas, as métricas utilizadas para sua mensuração e as técnicas aplicadas para sua análise, o que envolve uma análise de estatística descritiva e inferencial.

4.1 Natureza da Pesquisa

Por se tratar de uma investigação que se concentra na análise da densidade e diversidade lexical em produções textuais escritas por duplas de alunos brasileiros recém-alfabetizados, visando contribuir para a produção de conhecimento sobre essa temática, o presente estudo se caracteriza como uma pesquisa exploratória (GIL, 2008). As pesquisas exploratórias têm como principal finalidade explorar um determinado fato para proporcionar uma visão geral sobre suas características (GIL, 2008, p.27).

Para a análise do *corpus,* adotamos uma abordagem metodológica de léxico-estatística textual. Essa abordagem focaliza na análise dos aspectos quantitativos do sistema da língua, a partir da quantificação e classificação das palavras do texto (oral ou escrito) com apoio de técnicas estatísticas (BIDERMAN, 1978).

4.2 Coleta de dados

O *corpus* analisado foi coletado em 2012 por pesquisadores do LAME em uma sala de aula com 20 alunos, matriculados no 2º ano do ensino fundamental de uma escola da rede privada, localizada na cidade de Maceió.

Esta instituição atende alunos da educação infantil ao 5º ano do ensino fundamental em dois turnos diferentes: matutino e vespertino. Sua proposta de ensino[16]de língua materna é apoiada em uma perspectiva sócio-histórica com foco na formação de leitores e escritores, sendo todas as disciplinas trabalhadas a partir de projetos didáticos fundamentados na teoria socioconstrutivista (VYGOTSKY, 1988) que compreende o

[16] Informações colhidas no site da escola, disponível no link: http://criarerecrear.com.br/portal/pt/ensino-fundamental.

conhecimento como uma construção social, desenvolvido através da interação com os outros sujeitos e com o meio.

A coleta foi conduzida por pesquisadores do LAME a partir do projeto didático de língua portuguesa "Contos do como e do porquê"[18]. Esse projeto foi aplicado na sala de aula dos alunos ao longo de 3 meses (abril a junho de 2012), sendo os dois primeiros meses destinados para a leitura de diferentes contos etiológicos (Anexo 1) e os dois últimos meses para a produção textual de "contos etiológicos inventados".

Durante a coleta, os alunos foram agrupados em díades pela professora para escreverem junto um conto etiológico, ou seja, uma produção de escrita colaborativa a dois. O critério seguido para a formação das díades era que os alunos tivessem uma boa relação interpessoal e, preferencialmente, que mantivessem relação de amizade fora da escola. Seguindo os procedimentos da coleta, se algum componente das díades faltasse, novas díades poderiam ser formadas, ou quando isso não fosse possível, o aluno poderia escrever sozinho ou até mesmo em trio.

Assim, uma vez por semana cada díade produzia um texto a partir de temas sugeridos pela professora ou temas livres. Para garantir que os componentes das díades tivessem as mesmas oportunidades de escrita, a cada proposta os alunos alternavam o papel de "escrevente" (aquele responsável pela inscrição na folha de papel) e "ditante" (responsável por ditar o texto a ser escrito). Essa variável ("escrevente x "ditante") não será discutida nesse estudo. Após essa coleta, os textos foram digitalizados e nomeados como Dossiê Criar_2012 e, posteriormente, armazenados no banco de dados do LAME.

4.3 Constituição do *Corpus*

Para uma análise mais consistente da densidade e diversidade lexicais nos textos, estabelecemos os seguintes critérios para a constituição do nosso *corpus:*

i) **Critério de exclusão**: díades cujos componentes faltaram em, pelo menos, uma das propostas de produção.
ii) **Critério de inclusão**: díades formadas pelos mesmos componentes que participaram das 6 propostas de produção.

[18] Contos etiológicos caracterizados como um tipo de narrativa ficcional que busca explicar o porquê e como das características, das origens, dos modos de comportamento, de objetos, animais, fenômenos da natureza, etc.

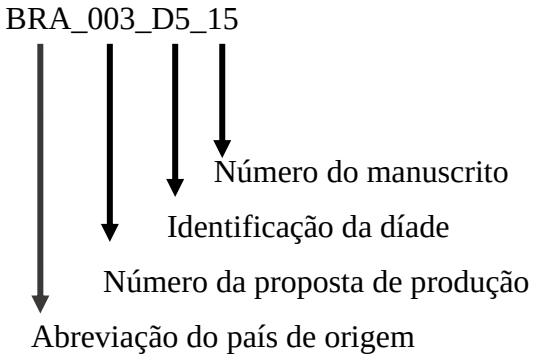
Das 10 díades que participaram da coleta, apenas 5 atenderam ao critério de inclusão. Assim, neste estudo analisamos apenas as produções dessas cinco díades. Logo, nosso *corpus* é composto por 30 textos, produzidos a partir das seguintes propostas temáticas:

Quadro 3 – Relação das propostas de produção dos contos etiológicos

Proposta	**Tema**
1	Livre
2	Sugerido (Por que o cachorro faz au, au? E o gato faz miau?)
3	Livre
4	Sugerido (Como surgiram as palavras?)
5	Livre
6	Livre

4. 4 Organização e tratamento do *corpus*

Com o *corpus* definido, catalogamos os 30 textos. Essa catalogação foi essencial para garantirmos um maior controle e organização dos textos ao longo de nossa análise. Desta maneira cada produção foi nomeada seguindo a estrutura da seguinte legenda:



Após essa catalogação, criamos um quadro de controle para registrar as informações dos textos de cada dupla. Nesse quadro também foram registrados os títulos escolhidos por elas nas produções com tema livre. Como pode ser observado no quadro abaixo.

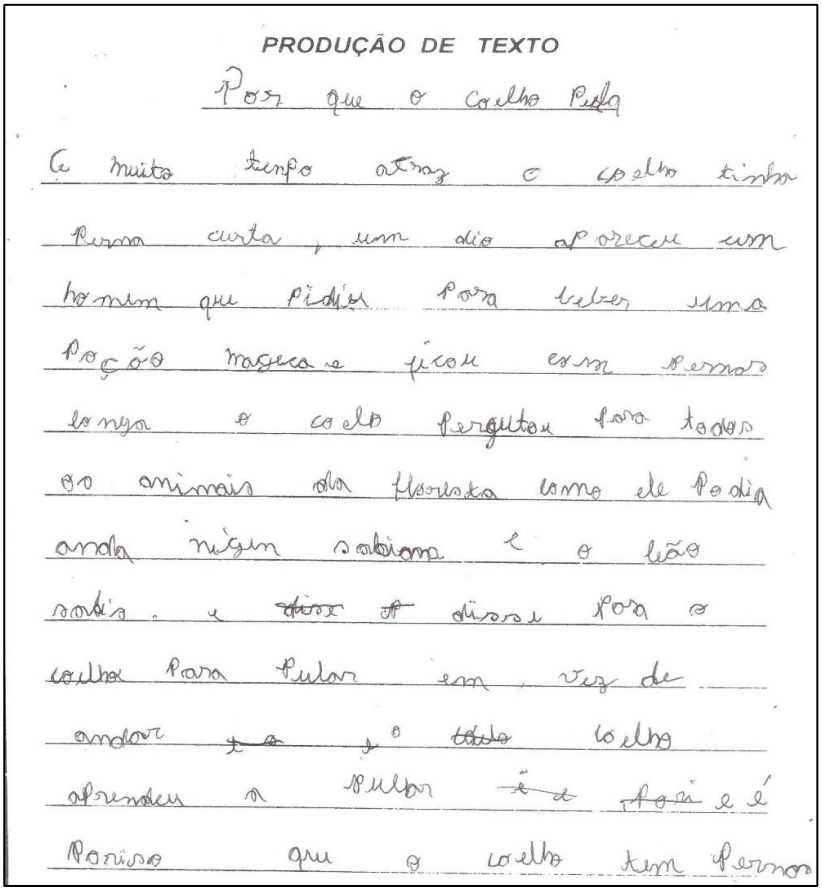
Quadro 4 – Relação dos textos e seus respectivos títulos

PRODUÇÃO	TEMA	TÍTULO
BRA_001_D1_01	Livre	O gamba que mentia
BRA_001_D2_02	Livre	Como surgiu o ambiente e os animais?
BRA_001_D3_03	Livre	O dragão que não voava
BRA_001_D4_04	Livre	Por que o sol brilha muito?
BRA_001_D5_05	Livre	Por que a girafa tem o pescoço longo
BRA_002_D1_06	Sugerido	Por que o cachorro faz au, au? E o gato faz miau?)
BRA_002_D2_07	Sugerido	
BRA_002_D3_08	Sugerido	
BRA_002_D4_09	Sugerido	
BRA_002_D5_10	Sugerido	
BRA_003_D1_11	Livre	Por que o gato tem unha grande?
BRA_003_D2_12	Livre	Por que existe o arco-íris?
BRA_003_D3_13	Livre	Por que o leão ruge?
BRA_003_D4_14	Livre	Por que a formiga é pequena?
BRA_003_D5_15	Livre	Por que o coelho pula?
BRA_004_D1_16	Livre	Por que o pássaro voa?
BRA_004_D2_17	Livre	A juba do leão
BRA_004_D3_18	Livre	Por que a barata é pequena?
BRA_004_D4_19	Livre	Por que o elefante tem tromba grande?
BRA_004_D5_20	Livre	Por que os esquilos gostam de nozes?
BRA_005_D1_21	Sugerido	Como surgiram as palavras?
BRA_005_D2_22	Sugerido	
BRA_005_D3_23	Sugerido	
BRA_005_D4_24	Sugerido	
BRA_005_D5_25	Sugerido	
BRA_006_D1_26	Livre	Por que o peixe nada?
BRA_006_D2_27	Livre	Por que os peixes moram no mar, no rio e nos lagos?
BRA_006_D3_28	Livre	Por que o leopardo é rápido?
BRA_006_D4_29	Livre	Por que o beija-flor beija a flor?
BRA_006_D5_30	Livre	Por que o sol brilha?

Dando prosseguimento a organização e tratamento do nosso corpus, fizemos a transcrição de cada texto através do Microsoft Word e normalização alguns aspectos deixados durante a produção textual, como erros de ortografia, segmentação de palavras e elementos rasurados. Procuramos preservar as mudanças das linhas feitas pelos alunos escreventes, para fazer corresponder graficamente o texto transcrito com seu texto original. Também enumeramos todas as linhas, desde a primeira palavra até a última palavra escrita. Apesar desses ajustes, não incluímos novas palavras (artigos, pronomes, conectivos etc.), caso a dupla não as tenha escrito.

Para uma melhor visualização das produções antes e depois desse processo de normalização e transcrição, trazemos na figura abaixo um texto original com as marcas gráficas deixadas pela díade (Figura 4).

Figura 4 - Produção BRA_003_D5_015



Fonte: Dossiê Criar - LAME, 2012.

Observando a figura (4), facilmente se identifica a presença de algumas separações de palavras, marcas de rasuras e erros de ortografia. Todos esses aspectos

foram ajustados durante o processo de transcrição. Como mostra a sua versão transcrita (Figura 5) a seguir:

Figura 5 -Transcrição do texto BRA_003_D5

1. Por que o coelho pula?
2. **Há** muito tempo **atrás** o coelho tinha
3. perna curta. Um dia apareceu um
4. homem que pediu para beber uma
5. **porção** mágica e ficou com pernas
6. **longas.** E o **coelho** perguntou para todos
7. os animais da floresta como eles podiam
8. **andar**. **Ninguém** sabia e o leão
9. sabia e disse para o
10. coelho para pular em vez de
11. andar e o coelho
12. aprendeu a pular e é
13. **por isso** que o coelho tem pernas
14. longas.

Como mencionamos, no momento da transcrição enumeramos todas as linhas do texto, desde o título até a última palavra escrita. Esse procedimento nos ajudou a localizar com maior facilidade a posição das palavras no corpo do texto. Como pode ser visualizado nesta transcrição, as palavras em destaque representam as correções que foram efetuadas. Como por exemplo, a inclusão da letra h na palavra “Há” na primeira linha, e a separação da palavra “por isso”, antes escrita por “poriso” na linha 13.

Vale ressaltar que essa forma de normalização é necessária para a contagem e classificação de palavras, para a extração da DiL e DeL, além de estabelecer um padrão comum para todos os textos analisados.

4.5 Categorias de análise

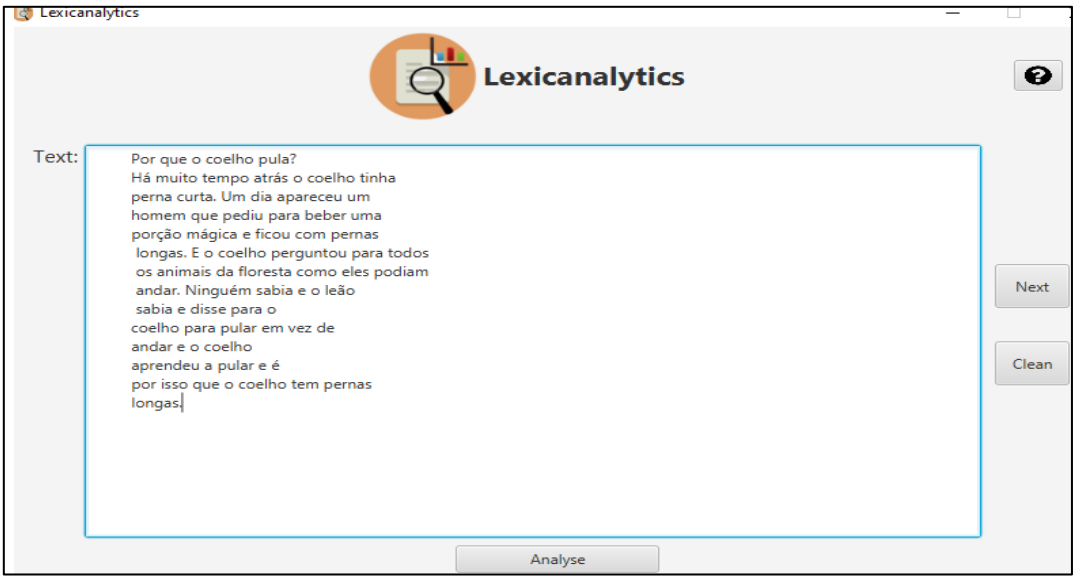
Em consonância com o referencial teórico e com os objetivos previstos neste estudo, analisamos as seguintes categorias de análise:

- **Palavras Escritas (PE):** representa o total de palavras escritas em um texto.
- **Diversidade Lexical (DiL):** representa a variedade de palavras *(types)* no texto.
- **Densidade Lexical (DeL):** representada a proporção de itens lexicais no texto.

4.6 Processo de mensuração das categorias de análise

Para a quantificação das categorias, contamos com o auxílio do software de análise lexical, *Lexicanalytics* (LEITE, 2016). Para medir o número de PE de cada texto, consideramos desde a primeira linha, em que geralmente era escrito o título da história, até a última palavra, comumente, a palavra "fim". Assim sendo, o primeiro passo foi copiar o texto transcrito no Microsoft Word e colar na tela inicial (Figura 6) do *lexicanalytics*, como podemos visualizar na figura abaixo:

Figura 6 - Tela 1 do *Lexicanalytics*

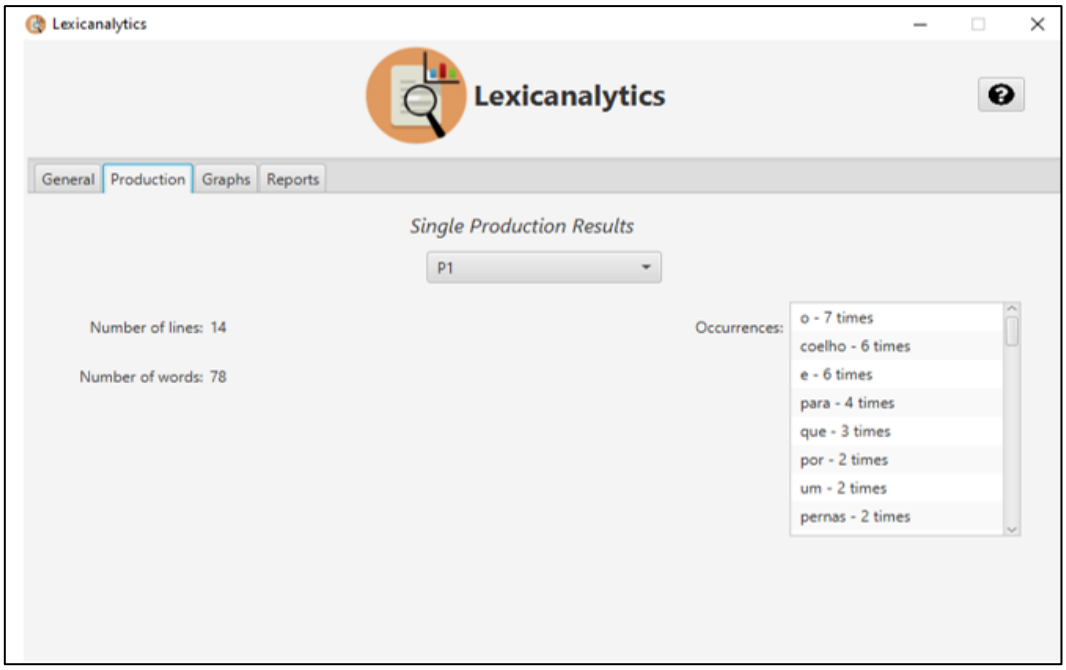


Fonte: Leite, 2016.

Uma funcionalidade importante desse s*oftware* é a análise simultânea de N textos o que, consequentemente, possibilita uma otimização de tempo para esse processo. Para analisar um texto por vez, após colá-lo na tela inicial, seleciona a tecla *analyse* (analisar). Caso o interesse seja a análise de um conjunto de textos, seleciona a tecla *next* (próximo) e insere os demais textos, sendo cada um por vez e finaliza selecionado a tecla *analyse*. Em seguida, uma nova tela (Figura 7) é aberta, contendo um resumo com as seguintes informações acerca do texto: número de linhas, número de palavras (PE) e do lado

esquerdo, a apresentação do registro de ocorrências (*occurrences*) de cada palavra de forma decrescente. No centro do texto, é possível selecionar a produção que se quer analisar. Todas essas informações podem ser conferidas na figura a seguir.

Figura 7 - Tela 2 do *Lexicanalytics*



Fonte: Leite, 2016.

Em suma, realizamos esse mesmo procedimento para cada produção e, ao final, construímos uma tabela para cada díade contendo as seguintes especificações:

Produção	Total de PE

Para mensuração da DiL também utilizamos o *Lexicanalytics* que, além de quantificar o número de palavras, tem como uma de suas funcionalidades o modelo VOCD. Assim, seguindo o mesmo procedimento de PE, os valores obtidos foram organizados em uma tabela com a seguinte estrutura:

Produção	DiL (%)

Diferente dessas categorias, a medição da DeL foi composta por três etapas. A primeira, através da análise do número de ocorrência de cada palavra fornecido pelo *lexicanalytics* (Tela 2). Na segunda etapa, retornamos à transcrição do manuscrito para fazer uma checagem e classificação manual de todos os itens lexicais, destacando-os de

acordo com sua classe gramatical (substantivo, adjetivo, verbos e advérbios modais terminados em mente).

Para uma melhor ilustração dessa etapa, apresentamos a seguir uma das histórias produzidas pela díade 2 (D2), cujo tema sugerido foi: “Como surgiram as palavras? Neste exemplo, estão destacados em cor laranja os itens pertencentes à classe dos substantivos e na cor verde, os itens que representam os verbos, vejamos:

BRA_004_D2_17

1 Como surgiram as palavras?
2 Era uma vez duas meninas que eram irmãs
3 Elas foram para uma casa e lá
4 haviam muitas letras e essas letras não
5 se encaixavam. A moça que morava lá
6 virou amiga das duas meninas e ela disse:
7 vocês querem tirar essas letras comigo?
8 E as meninas disseram: sim. E todas tiraram
9 as letras. Aquela moça pensava que aquelas
10 letras não formavam palavras. E elas
11 formaram muitas palavras.

Essa história foi escrita com um total de 72 palavras. Dessas palavras, 33 possuem valores lexicais (ou itens lexicais) e 39 valores gramaticais. Dos 33 itens lexicais, 17 são substantivos (cor laranja) e 16 são verbos (cor verde). No entanto, nenhum adjetivo e advérbio (terminado com o sufixo mente) foi escrito pela dupla. Após essa quantificação, utilizamos a equação proposta por Halliday (1985), aplicando os valores da quantificação dos itens lexicais da seguinte maneira:

$$D\text{eL} = \frac{\text{Itens Lexicais}}{\text{TP}} \text{X } 100 = \frac{33}{72} \text{ X } 100 = 45{,}83\%$$

Um fator relevante a ser citado é que não foi encontrado nenhum programa computacional que faça essa diferenciação de itens lexicais em textos. Por isso, seguimos

esse procedimento de medição. Por fim, ao aplicarmos esse procedimento para cada manuscrito, uma tabela também foi construída para cada díade contendo as seguintes informações:

Produção	DeL (%)

4.7 Procedimentos da análise das categorias

Após a mensuração de PE, DiL e DeL, os valores obtidos passaram por uma análise de estatística descritiva e inferencial. A seguir serão apresentados os instrumentos e as métricas utilizadas para o tratamento dos dados. Vale destacar que para essa análise adotamos como aporte teórico Lansor e Farber (2010), Magalhães e Lima (2010), Morettin e Bussab (2010), Moore (2003) e Triola (2013).

4.7.1 Análise de estatística descritiva

Em nossa análise descritiva, organizamos e sumarizamos as principais informações das categorias estudadas (PE, DiL e DeL). Para isso, usamos medidas de posição, medidas de variabilidade e intervalo de confiança, detalhadamente apresentadas a seguir.

4.7.1.1 Medidas de posição

Para estimar onde a maioria dos dados está localizada, ou seja, sua tendência central, usamos a média aritmética, definida como a soma de todos os valores de uma amostra, dividida pelo número de valores.

4.7.1.2 Medidas de Variabilidade

Para analisar a variabilidade dos nossos dados, isto é, o quanto seus valores se aproximam ou se distanciam, adotamos duas medidas:

i) Amplitude de Variação (AV) para medir a diferença entre o maior e menor valor no conjunto de dados.

ii) Desvio-Padrão (DP) para medir o grau e variação dos dados. Desta maneira, quanto menor for o DP mais homogêneo os dados são, enquanto que um DP alto expressar uma heterogeneidade.

4.7.1.3 Intervalo de Confiança

Usamos o Intervalo de Confiança (IC) para estimar a posição da maior concentração de PE, DeL e DiL. Sendo essa estimação calculada através da seguinte operação:

$$\text{IC} = (\text{média} - \text{desvio padrão};\ \text{média} + \text{desvio padrão}) \quad (6)$$

4.7.2 Análise estatística inferencial

A estatística inferencial é definida como um estudo de técnicas que possibilitam a exploração de um grande conjunto de dados a partir das informações e conclusões obtidas de um subgrupo de valores, usualmente de dimensão muito menor (MAGALHÃES e LIMA, 2010, p.2).

Neste trabalho usamos as técnicas de estatística inferencial para estimar a existência de um comportamento padrão das nossas categorias de análise, além de verificar se há uma diferença significativa entre as díades analisadas. Para isso, foram aplicados quatro testes: coeficiente de correlação linear de Pearson, teste de normalidade *Shapiro-Wilk*, teste ANOVA e teste Kruskal-Wallis.

4.7.2.1 Correlação linear (r)

O coeficiente de correlação linear (r), também conhecido como correlação de *Pearson*, é definido como uma medida que quantifica a força da relação linear entre duas variáveis (TRIOLA, 2013).

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} \quad (7)$$

Em que:

- O símbolo x_i se refere ao i-ésimo elemento do primeiro conjunto de dados

- O símbolo y_i se refere ao i-ésimo elemento do segundo conjunto de dados
- O símbolo $\overline{x}$ representa a média do primeiro conjunto de dados
- O símbolo $\overline{y}$ representa a média do segundo conjunto de dados

O coeficiente de correlação (r) sempre assumirá valores entre -1 e 1 e pode ser interpretado da seguinte maneira:

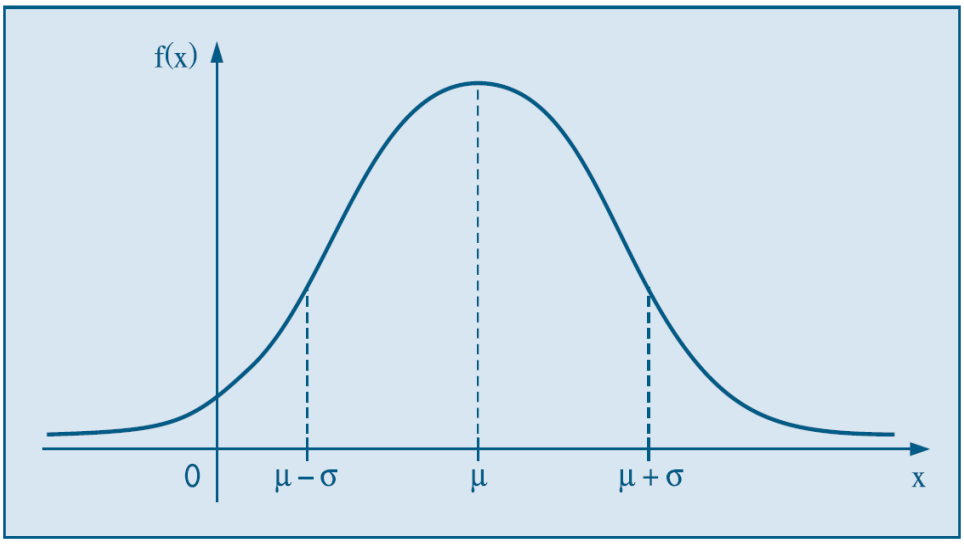
- r =1: correlação positiva (quando uma variável aumenta, a outra também aumenta)
- r =0: correlação inexistente (as variáveis não dependem linearmente uma da outra.
- r = -1 correlação negativa, (à medida que uma variável aumenta, a outra diminui).

Para avaliar a correlação de nossas categorias, adotamos o coeficiente de correlação de Pearson, em seguida, plotamos um gráfico de dispersão, pois, de acordo com Moore (2013, p.82), às relações entre duas variáveis quantitativas são melhores exibidas graficamente através de um gráfico de dispersão.

4.7.2.2 Teste de normalidade *Shapiro-Wilk*

O teste de normalidade *Shapiro-Wilk* é aplicado para identificar se um conjunto de dados segue ou não uma distribuição de probabilidade normal (ou gaussiana), sendo essa distribuição caracterizada pelo formato de sino, determinada em função da média e do desvio padrão. Sua representação matemática é $N(\overline{x}, DP^2)$ e seu gráfico tem a seguinte forma:

Gráfico 1 – Distribuição normal (Gaussiana)



Fonte: Morettin e Bussab (2010).

Para a aplicação do teste *Shapiro-Wilk* é necessário construir uma hipótese nula e uma alternativa. Geralmente, a hipótese nula assume que os dados se encaixam no modelo de distribuição normal e a hipótese alternativa assume que a distribuição de frequência observada não se encaixa no modelo de distribuição normal. Para avaliar o modelo de distribuição das nossas categorias estudadas (PE, DeL, DiL), construímos as seguintes hipóteses:

- Hipótese nula (H_0): os dados seguem uma distribuição normal
- Hipótese alternativa (H_a): os dados não seguem uma distribuição normal

Para verificar se as categorias estudadas seguem ou não um modelo de distribuição normal, avaliamos o resultado do p-valor do *Shapiro-Wilk* de cada uma. Adotando um nível de significância $\alpha = 0{,}05$, se $p > 0{,}05$ a hipótese nula é aceita, já se $p \leq 0{,}05$, a hipótese nula é rejeitada.

4.7.2.3 Testes de variância

Os testes de variância é uma técnica estatística usada para avaliar as informações da média dos dados, ou seja, se há alguma diferença entre elas. A depender da distribuição dos dados, a análise de variância pode ser feita através de testes paramétricos ou não-paramétricos.

Para analisar a diferença entre dados que seguem um modelo de distribuição normal, utilizam-se os testes paramétricos. Caso os dados não sigam uma distribuição normal, aplicam-se os testes não-paramétricos. Neste estudo, adotamos os seguintes testes para avaliar a diferença entre as díades: *ANalysis Of VAriance* (ANOVA) e *Kruskal-Wallis.*

ANOVA é um teste paramétrico usado para avaliar a diferença entre dadosque seguem um modelo de distribuição normal. Segundo Magalhães e Lima (2010), para a utilização dessa técnica algumas condições devem ser satisfeitas:

- Os dados[20] devem representar variáveis aleatórias independentes
- Os dados devem seguir um modelo normal
- Todas as populações devem apresentar variâncias iguais a DP^2

[20] No nosso caso, PE, DiL e DeL.

Para analisar a diferença entre dados que não seguem um modelo de distribuição normal, usamos o teste não-paramétrico *Kruskal-Wallis*. Para aplicação de ambos os testes, adotamos um valor de significância de 0,05 e avaliamos as seguintes hipóteses para cada categoria analisada em nosso estudo:

- Hipótese nula (H_0): As médias das díades são iguais
- Hipótese alternativa (H_a): Uma ou mais médias se distanciam significativamente das outras.

4. 8 Recursos tecnológicos para a análise os dados

Todas as medidas de estatística descritiva e testes de estatística inferencial foram executados com o auxílio dos *softwares* Microsoft *Excel*[22] e R[23]. Também usamos esses *softwares* para construção dos gráficos (de barra, linha e dispersão) que irão demonstrar os nossos resultados.

[22] Versão 2016.
[23]Um *software* livre para computação estatística e geração de gráficos. Disponível em https://www.r-project.org/. Versão usada: 3.6.1

5 RESULTADOS

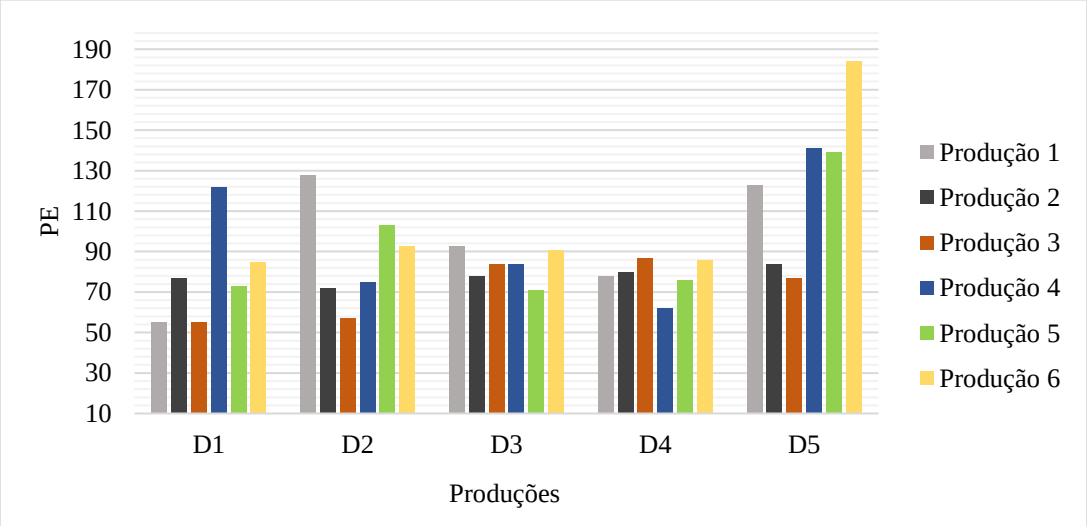
Esta seção apresenta os resultados referentes às nossas categorias investigadas: Palavras Escritas (PE), Diversidade Lexical (DiL) e Densidade Lexical (DeL), subdivididos em duas análises: descritiva e inferencial. Nos resultados da análise descritiva estão as médias de PE, DeL e DiL obtidas por cada dupla. Na análise inferencial estão os resultados dos testes de distribuição, variância e correlação. Posteriormente, trazemos a discussão dos resultados, considerando-se os objetivos gerais e específicos delineados para esta dissertação e respondendo as questões que conduziram nossa pesquisa.

5.1 Análise descritiva

5.1.1 Total de PE por díade

Após a quantificação do número de PE, identificamos que o texto mais extenso foi escrito pela D5, contendo 184 PE (TL)[24] e o texto de menor comprimento foi escrito pela D1 com 55 PE (TL), resultando em uma diferença de 129 PE. Em relação às demais díades, o gráfico abaixo traz uma visão geral sobre o número de PE escrita por cada uma ao longo das 6 propostas de produção.

Gráfico 2 - PE por díade



[24] Abreviação de Tema Livre

Seguindo as informações visuais expressas neste gráfico, é possível notar que nenhuma díade manteve o mesmo número de PE nas seis produções. Entretanto, comparando o conjunto de produções de cada uma, observamos que a D3 e D4 apresentaram uma certa aproximação, ou seja, ambas usaram um número parecido de PE para escreverem seus textos, indicando uma possível simetria. O mesmo também foi observado entre a D1 e D2.

Para uma melhor compreensão acerca do número de PE usado por cada díade, calculamos a média geral de cada uma com seus respectivos desvios padrões, como podemos visualizar na tabela abaixo.

Tabela 2 - Média de PE por díade

Díades	Nº de Produções	Média	DP
D1	6	78	24,77
D2	6	88	25,44
D3	6	83	8,16
D4	6	78	9,04
D5	6	124	39,80

Com essas médias de cada díade, podemos reafirmar que de fato as D1, D2, D3 e D4 usaram quantidades próximas de PE para produzirem seus textos. No entanto, a D5 foi a díade que mais se distanciou, escreveu uma média de 124 PE, ou seja, aproximadamente 46 palavras a mais de que outras díades.

Complementando esses resultados, temos os valores dos desvios-padrões que indicam com maior precisão o quanto cada díade variou em relação a sua média de PE. Esses valores dos DP ilustrados na tabela 2, mostram uma maior assimetria do número de PE ao longo das produções das D1, D2 e D5, quer dizer, ora escreviam seus textos com um número alto de palavras, ora com número baixo.

5.1.2 Taxas de DiL por díade

Antes de iniciar a apresentação dos resultados da DiL é importante ressaltar que neste subtópico estamos nos referindo apenas à proporção de palavras diferentes (*types*) de cada díade. Sabemos que durante a escrita de um texto em determinados momentos as palavras começam a se repetir, como já foi discutido anteriormente. Portanto, quanto

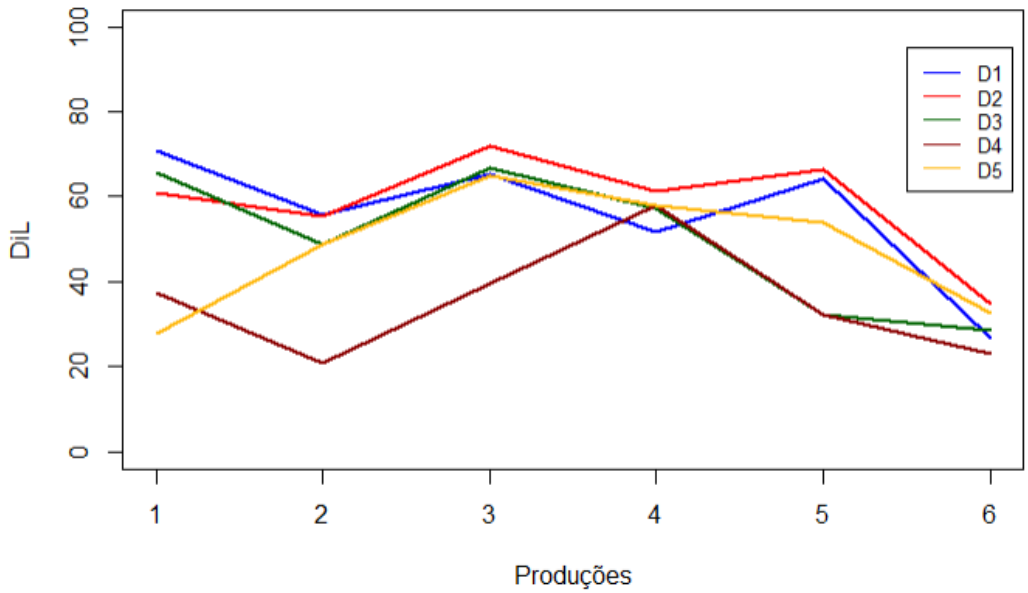
maior a taxa de DiL de um texto, menor repetição de palavras ele possui. Na tabela abaixo podemos observar as médias de DiL obtidas por cada díade.

Tabela 3 - Média de DiL por díade

Díades	Nº de Produções	Média (%)	DP
D1	6	55,80	15,89
D2	6	58,49	16,36
D3	6	49,86	13,49
D4	6	35,14	12,32
D5	6	47,72	14,64

Observando a tabela acima, nota-se que apenas duas díades produziram média de DiL superior a 50%, neste caso, a D1 com média de 55,81% e a D2 com média 58,49%. A D3 e a D4, apesar de ficarem com média inferior à 50%, não se distanciaram tantos das díades D1 e D2. Dentre as cinco díades, os textos produzidos com menor diversidade lexical foram os da D4 com uma média de apenas 35,14%, sugerindo uma alta concentração de repetições de palavras. Para complementar esses resultados, também analisamos a trajetória da DiL de cada dupla ao longo das seis produções. Vejamos o gráfico abaixo.

Gráfico 3 - Taxas de DiL por díade



Examinando esse gráfico, encontramos alguns pontos em comum. Primeiramente, todas as díades possuem crescimento na produção 3 e uma queda na produção 6. Vale

ressaltar que essas produções (3 e 6) foram produzidas a partir de “temas livres”. Esses resultados podem sugerir que as díades possuíam um conhecimento maior acerca dos temas escolhidos para os títulos de suas histórias na produção 3, do que nos temas escolhidos na produção 6. Assim, o domínio que os alunos têm sobre o que vão escrever é um fator que também pode contribuir para o uso de um vocabulário mais diversificado (SADEGHI e DILMAGHANI, 2013).

Em uma análise mais individual do texto que alcançou maior e menor DiL, é notável que a maior taxa foi adquirida na produção 1 da D1, com aproximadamente 71%, enquanto a história com a menor taxa DiL foi escrita pela D4 com aproximadamente 23%, o que ,consequentemente, refletiu no baixo valor de sua média final.

Portanto, considerando essas taxas, das cinco díades analisadas a D1, D2 e D3, D5 foram as que usaram um maior vocabulário na produção de seus textos, sendo aproximadamente 50% de palavras diferentes e 50% de palavras repetidas. Enquanto a D4, aproximadamente, 75% das palavras de suas produções são repetidas e apenas 25% das palavras são diferentes.

5.1.3 Taxas de DeL por díade

Em relação à DeL, a maior média foi produzida pela D4 (51,68%) e a menor média produzida pela D3 (46,14%), resultando uma diferença de aproximadamente 5% entre elas. Os valores dos desvios-padrões das díades também são próximos e com valores baixos, o que revela uma certa consistência nas taxas da DeL ao longo das produções de cada díade. Em síntese, as médias das díades apresentam-se mais uniformes, distribuídas no intervalo de 46,14% a 51,68%.

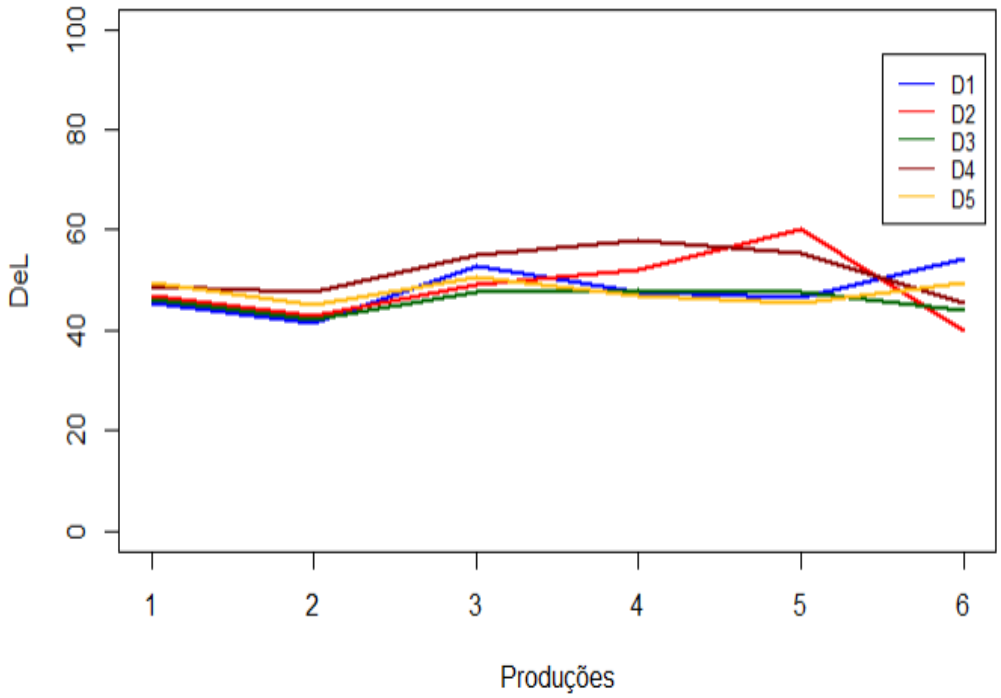
Tabela 4 - Média de DeL por díade

Díades	Nº de Produções	Média (%)	DP
D1	6	48,00	4,28
D2	6	48,51	6,55
D3	6	46,14	2,3
D4	6	51,68	4,69
D5	6	47,84	2,15

Essas médias nos revelam informações importantes, pois a díade (D4) que escreveu seus textos com maior DeL (51,68%) foi a mesma díade que também obteve a menor taxa de DiL (35,14%). Logo, podemos notar uma relação inversa entre a DeL e a DiL em suas produções.

Para termos uma compreensão mais abrangente sobre como as taxas de DeL de cada díade se comportaram ao longo dos seis processos de produção textual, construímos o gráfico (4) apresentado a seguir.

Gráfico 4 - Taxas de DeL por díade



O gráfico representa os valores da DeL para cada díade. Assim, observamos um comportamento interessante na 1ª produção, as díades iniciaram seus textos com uma DeL semelhante que se estende até a 3ª produção. A partir da 4ª produção, duas díades (D2 e D5) começam a se diferenciar das outras. Enquanto na última produção (6ª), todas as díades escreveram taxas diferentes.

Esse gráfico de trajetória, nos revela que as díades não se diferenciam tanto na taxa da DeL, ou seja, apesar de serem cinco duplas diferentes, todas usaram uma quantidade de itens lexicais semelhante para escreverem seus textos.

5.1.4 Média de itens lexicais por díade

Como a DeL representa a proporção de itens lexicais do texto, calculamos a média de ocorrência das classes lexicais (substantivo, verbos, adjetivos e advérbios modais terminados em mente) usadas pelas díades, como mostra a tabela abaixo:

Tabela 5 - Média dos itens lexicais por díade

Díades	Substantivos	Verbos	Adjetivos	Advérbios
D1	16,5	18,5	2	0,3
D2	24,5	16,5	1,67	0,2
D3	19	15,5	3,67	0,3
D4	17,83	18,67	3,33	0,3
D5	28,3	27,5	3,3	0,5

Dentre as quatros classes gramaticais expressas nessa tabela, os substantivos e os verbos foram os mais recorrentes nos textos. Nas produções das díades D2, D3 e D5 os itens mais frequentes foram os substantivos e depois os verbos. No entanto, nos textos da D1 e D4 aconteceu o oposto, os itens mais frequentes foram os verbos, seguidos pelos substantivos. As classes com menor ocorrência nos textos de todas as díades foram os adjetivos e os advérbios modais terminados em -mente.

Sobre as ocorrências dos itens lexicais, Halliday (1989) ressalta que os substantivos é a classe de palavras mais frequente no vocabulário de toda língua. No entanto, em nosso trabalho os textos de duas díades contrariam essa tendência, tendo como classe mais frequente os verbos, como descrito no parágrafo acima.

Quanto aos adjetivos, Berman (2009) destaca que seu número de ocorrência possui uma correlação significativa com o fator idade e nível escolar, isto é, à medida que os alunos avançam na idade e no nível escolar, o número de adjetivo de suas produções escritas também tende a crescer significativamente.

Sobre a baixa ocorrência dos advérbios, Martins (2016), em seu estudo sobre progressão escolar analisou narrativas textuais de alunos portugueses do (5º, 7º e 10º) e também constatou que o percentual de advérbios terminados em -mente se manteve inalterado ao longo dos anos escolares. Isso sugere que comparando com as outras classes lexicais, os advérbios terminados em mente são os que apresentam a menor tendência de crescimento ao longo da progressão escolar.

Para verificar os substantivos e verbos mais frequentes nas produções das díades, fizemos um levantamento dos 30 textos e destacamos as cinco palavras que foram mais escritas ao longo de todo processo. O quadro abaixo apresenta cada palavra acompanhada de seus respectivos números de ocorrências.

Quadro 5- Substantivos e verbos mais frequentes

Substantivos mais frequentes	**Verbos mais frequentes**
Cachorro (26)	Tinha (16)
Girafa (23)	Era (14)
Gato (17)	Fazer (9)
Cobra (14)	Ficou (4)
Palavras (13)	Foi (4)

De acordo com as informações expressas no quadro 5, dos cinco substantivos mais escritos pelas díades, quatro fazem referência ao nome de algum animal. Essa alta taxa de ocorrência pode ser justificada através dos títulos escolhidos pelos alunos em suas histórias de tema livre (Quadro 4), uma vez que constatamos que das 20 histórias de temas livres escritas, 16 falam sobre animais. Em relação aos verbos, por se tratar de um conto etiológico que explica como algo aconteceu ou surgiu, os alunos usaram os verbos do pretérito imperfeito do indicativo (tinha e era), verbos do pretérito perfeito do indicativo (ficou e foi) e verbo no infinitivo (fazer).

5.1.5 Comparação entre as médias de PE, DiL e DeL

Para finalizar os resultados da nossa análise descritiva, apresentamos na tabela 6 as médias finais e o intervalo de confiança para cada categoria estudada.

Tabela 6 - Média geral de PE, DiL e DeL

Categorias	Média geral	DP	Intervalo de confiança
PE	90,00	28,87	61,13 ⟝ 118,87
DiL	49,40%	15,96	33,44 ⟝ 65,36
DeL	48,41%	4,75	43,66 ⟝ 53,16

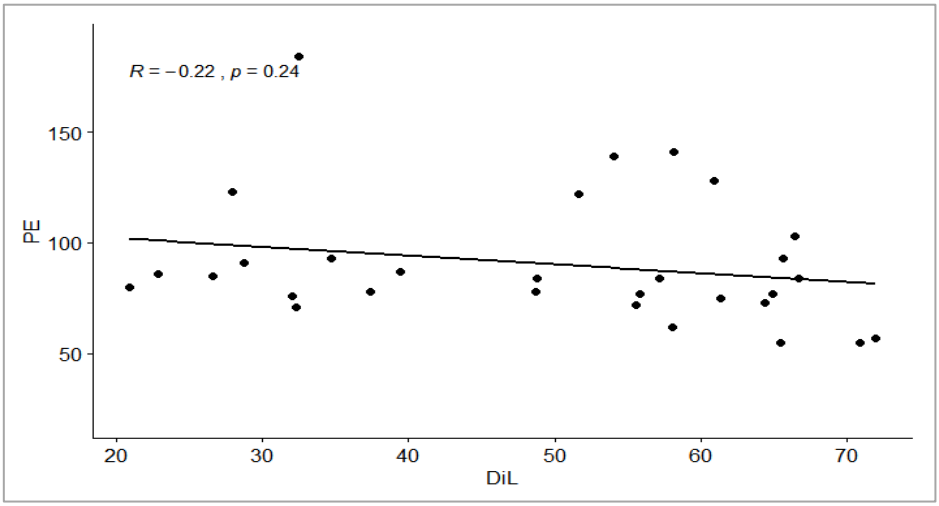
Analisando essas médias finais, notamos uma aproximação entre as médias da DiL e DeL, porém, comparando os seus respectivos intervalos de confiança, percebemos que a concentração das taxas da DiL ocupa um intervalo bem maior que o intervalo da DeL.

5.2 Resultados da análise inferencial

5.2.1 Correlação entre PE, DeL e DiL

Para determinar o grau de dependência entre as categorias estudadas (PE, DiL, DeL) e, consequentemente, verificar o quanto uma categoria pode interferir nos resultados da outra, aplicamos o teste de correlação de Pearson. Para essa análise, construímos através do *software* R um diagrama de dispersão verificando as seguintes relações: PE e DiL, PE e DeL, e DeL e DiL. Os resultados desse teste podem ser observados nos gráficos abaixo.

Gráfico 5 - Dispersão entre PE e DiL

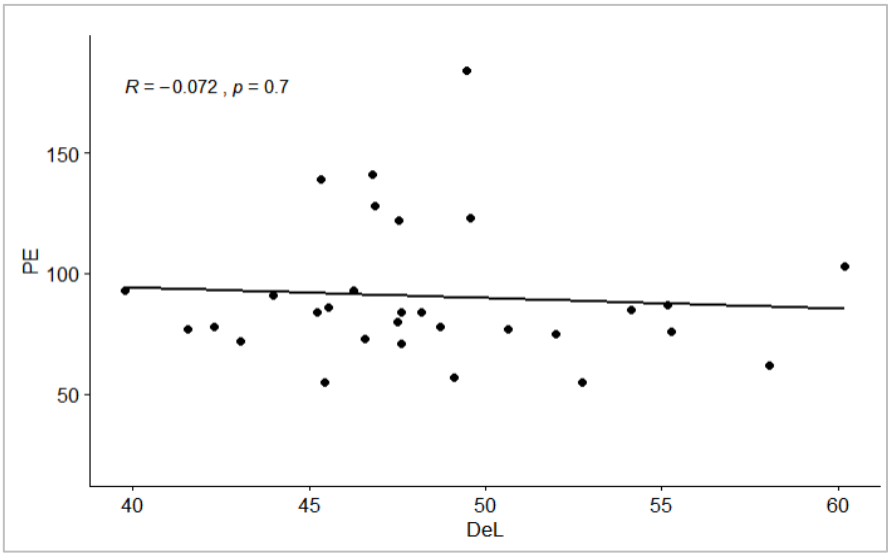


O p-valor = 0,24 e o coeficiente de Pearson R = - 0,22 apresentado no gráfico 5 indicaram que não há correlação entre PE e DiL, uma vez que os pontos do diagrama não seguem uma tendência linear. Esses resultados sugerem que o tamanho do texto não interfere no valor da DiL.

No teste entre PE e DeL, o valor resultante do coeficiente de Pearson R = - 0,072 e o p-valor = 0,7, apresentados no gráfico 6, também indicaram não haver correlação linear entre densidade lexical e número de palavras escritas. Em outras palavras, um texto

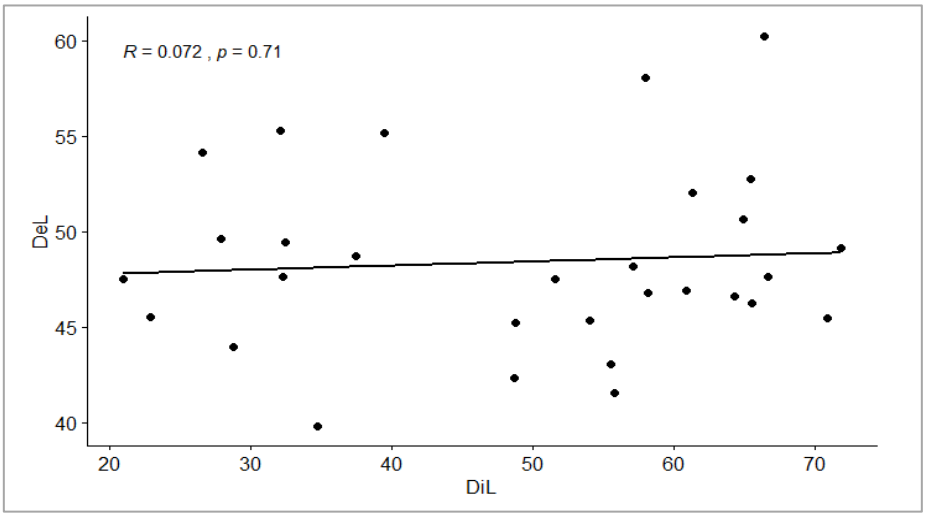
longo ou um texto curto pode possuir uma alta taxa de DeL. O contrário também pode ocorrer.

Gráfico 6 - Dispersão entre PE e DeL



E, por fim, no teste entre DeL e DiL, os valores do coeficiente Pearson R = 0,072 e o p-valor = 0,71, também nos confirmaram uma independência entre a densidade e diversidade lexical, isto é, uma não afeta a taxa da outra. Deste modo, podemos encontrar textos com alta taxa de diversidade lexical, mas com pequena taxa de densidade, assim, como textos com alta taxa de densidade e baixa taxa de diversidade. Essas informações podem ser conferidas no gráfico abaixo.

Gráfico 7 - Dispersão entre DeL e DiL



4.2.2 Análise da distribuição de probabilidade de PE, DeL e DiL

Para verificar a distribuição das nossas categorias de estudo, aplicamos com o auxílio do *software* R, o teste Shapiro-Wilk. Mas, para isso, construímos um teste de hipótese de distribuição para cada categoria, adotando um nível de significância de $\alpha = 0{,}05$. Os resultados da aplicação desse teste podem ser acompanhados a seguir.

i) Teste de hipótese de PE

- Hipótese nula (H_0): *PE* segue uma distribuição normal
- Hipótese alternativa (H_a): *PE* não segue uma distribuição normal

Para verificar essas hipóteses, observamos o p-valor resultante do teste Shapiro-Wilk a seguir:

Figura 9 - Resultado do teste *Shapiro-Wilk* para PE

```
# Testa se segue uma distribuicao normal
testeShapiro <- shapiro.test(
     x = data$PE
)

# Extrai o p-value
testeShapiro$p.value
```

```
[1] 0.000644318
```

Como podemos constatar na figura acima, o resultado indicou um p_{valor} menor que 0,05, logo, a hipótese nula é rejeitada. Assim, concluímos que PE não segue uma distribuição normal (gaussiana).

ii) Teste de hipótese da DiL

- Hipótese nula (H_0): *DiL* segue uma distribuição normal.
- Hipótese alternativa (H_a): *DiL* não segue uma distribuição normal.

Verificando esse teste de hipótese a partir do teste Shapiro-Wilk, obtemos o seguinte resultado.

Figura 10 – Resultado do Teste de Shapiro-Wilk para DiL

```
# Testa se segue uma distribuicao normal
testeShapiro <- shapiro.test(
     x = data$DiL
)

# Extrai o p-value
testeShapiro$p.value
```

```
[1] 0.0170622
```

O resultado do teste para DiL também indicou um p_{valor} menor que 0,05, portanto, podemos considerar a hipótese nula rejeitada. Assim, concluímos que a DiL não segue um modelo de distribuição normal (gaussiana).

iii) Teste de hipótese da DeL

- Hipótese nula (H_0): *DeL* segue uma distribuição normal
- Hipótese alternativa (H_a): *DeL* não segue uma distribuição normal

Para verificar essas hipóteses observamos o p-valor resultante do teste Shapiro-Wilk a seguir:

Figura 11- Resultado do Teste de Shapiro-Wilk para DeL

```
# Testa se segue uma distribuicao normal
testeShapiro <- shapiro.test(
     x = data$DeL
)

# Extrai o p-value
testeShapiro$p.value
```

```
[1] 0.3176753
```

Diferentemente das outras categorias, o resultado do teste da DeL indicou um p_{valor}valor maior que 0,05. Portanto, a hipótese nula é aceita. Assim, podemos concluir que a DeL segue uma distribuição normal (gaussiana).

4.2.3 Testes de variância de PE, DeL e DiL

A partir dos resultados do teste de distribuição, foi possível identificar que as categorias PE e DiL, não seguem um modelo de distribuição normal, apenas a DeL. Diante disso, aplicamos o teste de variância não-paramétrico Kruskal-Wallis para PE e DiL e o teste paramétrico ANOVA para DeL. Em ambos os testes, adotamos um nível de significância de α = 0,05. Vale ressaltar que esses testes foram usados para verificar se há diferença significativa entre as díades estudadas, em termos de PE, DiL e DeL. Para isso, construímos duas hipóteses para cada categoria.

i) Análise de variância de PE

- Hipótese nula (H_0): $\overline{PE}_1 = \overline{PE}_2 = \overline{PE}_3 = \overline{PE}_4 = \overline{PE}_5$ (dadas eventuais flutuações)
- Hipótese alternativa (H_a): Uma ou mais médias não respeita(m) a igualdade

Para avaliar essas hipóteses, aplicamos o teste *Kruskal-Wallis* com o auxílio do programa R, usando a função *kruskal.test*, como mostra a figura abaixo:

Figura 12 - Saída da função *kruskal.test* para os dados de PE.

```
kruskal.test(Y ~ X, data=dados)


    Kruskal-Wallis rank sum test

data:  Y by X
Kruskal-Wallis chi-squared = 6.9996, df = 4, p-value = 0.1359
```

Como podemos observar, o teste gerou um $p_{valor} = 0{,}1539$ maior que 0,05. Assim sendo, a hipótese nula pode ser aceita, indicando que não há diferença significativa entre as médias de PE das díades.

ii) Análise de Variância da DiL

- Hipótese nula (H_0): $\overline{DiL_1} = \overline{DiL_2} = \overline{DiL_3} = \overline{DiL_4} = \overline{DiL_5}$ (dadas eventuais flutuações)
- Hipótese alternativa (H_a): Uma ou mais médias não respeita(m) a igualdade

Aplicando o teste *Kruskal-Wallis* com o auxílio do programa R, usando a função *kruskal.test* chegamos na seguinte saída:

Figura 13 - Saída da função *kruskal.test* para os dados de DiL

```
kruskal.test(Y ~ X, data=dados)

    Kruskal-Wallis rank sum test

data:  Y by X
Kruskal-Wallis chi-squared = 7.2559, df = 4, p-value = 0.123
```

O resultado desse teste indicou um $p_{valor} = 0,123$. Como esse valor é maior que 0,05, a hipótese nula é aceita. Logo, também podemos afirmar que não existe diferença significativa entre as médias de DiL das díades observadas.

ii) Teste de Variância da DeL

- Hipótese nula (H_0): $\overline{DeL_1} = \overline{DeL_2} = \overline{DeL_3} = \overline{DeL_4} = \overline{DeL_5}$ (dadas eventuais flutuações)
- Hipótese alternativa (H_a): Uma ou mais médias não respeita(m) a igualdade

Para verificar essas hipóteses, aplicamos o teste ANOVA com o auxílio do programa R, usando a função *lm.* Os resultados podem ser observados na figura a seguir*:*

Figura 14 - Saída da função *lm* para os dados de DeL

```
modelo = lm(Y ~ X, data=dados)
summary(modelo)
```

```
Call:
lm(formula = Y ~ X, data = dados)

Residuals:
    Min      1Q  Median      3Q     Max
-8.7250 -2.5904 -0.1025  2.6517 11.6850

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  47.9967     1.9199  24.999   <2e-16 ***
XD2           0.5083     2.7152   0.187    0.853
XD3          -2.0083     2.7152  -0.740    0.466
XD4           3.7100     2.7152   1.366    0.184
XD5          -0.1517     2.7152  -0.056    0.956
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.703 on 25 degrees of freedom
Multiple R-squared:  0.1575,    Adjusted R-squared:  0.02274
F-statistic: 1.169 on 4 and 25 DF,  p-value: 0.3484
```

O teste gerou um $p_{valor} = 0,3484$ maior que 0,05. Assim sendo, a hipótese nula pode ser aceita, indicando que não há diferença significativa entre as médias de DeL das díades.

5.3 Discussão dos resultados

A partir da quantificação das categorias de estudo (PE, DiL, DeL) e da análise estatística descritiva e inferencial, foi possível alcançar as respostas para as nossas perguntas de pesquisa. Os resultados referentes à nossa primeira questão, revelaram que as díades usaram em média 90 PE para produzirem seus textos, com um desvio-padrão de 28,87, indicando a ocorrência de uma variação entre os números de PE nas 30

produções. Quanto à DiL, as díades produziram uma média de 49,40%, com desvio-padrão de 13,71, ou seja, das 90 palavras escritas, 44 são diferentes e 46 são repetidas.

Na DeL, os valores obtidos demonstraram que elas escreveram seus textos com uma média de 48% e desvio-padrão de 3,99, equivalente a aproximadamente 43 itens lexicais de um texto de 90 palavras. Respondendo então, nossa primeira questão.
Como podemos observar, os valores das médias gerais de DiL e DeL foram próximos, diferenciando-se apenas 1,4%. Entretanto, apesar dessa aproximação, constatamos através dos valores de seus respectivos desvios-padrões que as taxas da densidade foram mais estáveis, em outras palavras, as díades mantiveram uma proporção de itens lexicais semelhantes ao decorrer de suas 6 de produções.

Na diversidade lexical aconteceu o oposto, as médias das díades apresentaram maior variação ao decorrer das produções, com um desvio-padrão de 13,71. Porém, essa variação pode estar relacionada ao domínio do tema escolhido em suas produções de tema livre, pois notamos que algumas duplas escreveram seus textos (tema livre), ora com taxas de DiL altas, ora com taxas de DiL baixa. Assim, compreendemos que o domínio que os alunos têm sobre o que vão escrever é um fator que também influencia no uso de um vocabulário mais diversificado e, consequentemente, com menor proporção de palavras repetidas (SADEGHI e DILMAGHANI, 2013).

No que se refere à correlação entre o número de palavras escritas, densidade e diversidade, os resultados indicaram que não houve uma correlação linear entre essas três categorias, sugerindo que a extensão do texto não influencia nas taxas de DeL e DiL. Logo, esse resultado contraria, aquela expectativa de que um texto extenso é mais rico, em termos lexicais, que um texto mais curto. Essas considerações respondem a segunda questão da nossa pesquisa.

Entretanto, cabe ressaltar que essas observações no teste de correlação nos fornecem informações importantes acerca dessas métricas como variáveis independentes, ou seja, uma não exerce influência relevante sobre a outra. Desse modo, podemos dizer que PE, DiL e DeL são categorias de estudos que merecem análises individuais.

Por fim, respondendo à terceira questão, constatamos que apesar das variações nas taxas de PE, DeL e DiL ao longo das seis produções, os testes de variância indicaram não haver diferença significativa entre as díades, confirmando a hipótese principal do nosso estudo.

Comparando nossos achados a outros trabalhos que também avaliaram a densidade e diversidade em textos de alunos com idades próximas das díades brasileiras, percebemos que há uma certa aproximação apenas em relação à DeL.

No estudo de Rodrigues (2010), que analisou produções textuais de alunos portugueses do 2º ano, notamos uma aproximação entre a média de DeL dos textos produzidos pelas díades (48%) com a média de DeL dos textos produzidos pelos alunos portugueses (52%). No entanto, em termos de DiL, nossa média (49,4%) se distancia desses alunos (68%).

Também observamos uma aproximação dos nossos resultados com o trabalho de Martins (2016) quando analisou textos narrativos de alunos portugueses de anos escolares mais avançados (5º e 7º ano). Seus resultados indicaram uma média de densidade de 44,47% nos textos dos alunos do 5° ano, e 43,49% para os textos dos alunos do 7° ano. Em relação à diversidade, os alunos portugueses obtiveram médias superiores a 74%.

De modo geral, foi possível constatar que a média da DeL, usada pelas díades brasileiras, se mostrou superior às médias obtidas pelos alunos portugueses dos estudos de Rodrigues (2010) e Martins (2016), mesmo estando em anos escolares mais avançados. Em contrapartida, na DiL ocorreu o inverso, isto é, as médias obtidas pelos alunos portugueses foram superiores às médias das díades.

Para finalizar, destacamos os resultados do trabalho de Santos e Calil (2018) que adotou uma metodologia de escrita colaborativa semelhante ao nosso trabalho e concluiu que as díades de alunos portugueses do 2º ano escreveram seus textos com média de DeL (54,40%) e DiL (53,90%).

Comparando com os resultados deste trabalho, percebemos dois pontos semelhantes. Primeiramente, uma aproximação das médias de DeL e DiL entre as díades brasileiras e portuguesas. O segundo ponto a se destacar é que contrastando as médias da DiL (53,90%) e a média da DeL (54,40%) das díades portuguesas, encontramos uma diferença de 0,5%. Fazendo essa mesma comparação entre as médias da DiL (49,40%) e da DeL (48%) das duplas analisadas em nosso estudo, encontramos uma diferença de 1,4%. Logo, é possível notar que nos textos escritos em duplas há um equilíbrio entre a proporção de densidade e diversidade lexical.

Em contrapartida, nos textos individuais essa diferença é maior, pois, como podemos notar no estudo de Rodrigues (2010), a diferença entre as médias de DeL (52%) e DiL (68%) em textos de alunos portugueses do 2° ano foi de 16%. Essa diferença também foi observada entre as médias de DeL e DiL, obtidas no estudo de Martins (2016),

tanto nos textos dos alunos portugueses do 5° ano, quanto nos alunos do 7º ano, porém, com um percentual maior, aproximadamente 31%.

Portanto, nossos resultados sugerem que a escrita colaborativa pode influenciar na produção da densidade e diversidade, isto é, escrever a dois pode auxiliar nas escolhas lexicais, reunindo em um único texto um vocabulário diversificado e com densidade informacional. Logo, a produção textual escrita de forma colaborativa pode contribuir para uma maior riqueza lexical dos alunos.

6 CONSIDERAÇÕES FINAIS

O presente trabalho analisou a riqueza lexical em manuscritos escolares produzidos por díades de alunos brasileiros recém-alfabetizados, tomando como categorias de análise o número de Palavras Escritas (PE), Densidade Lexical (DeL) e Diversidade Lexical (DiL). Para isso, selecionamos 30 contos etiológicos escritos por 5 díades de alunos brasileiros recém-alfabetizados, coletados durante o desenvolvimento de um projeto didático intitulado "Contos do como e do porquê" no ano de 2012, em uma escola privada de Maceió.

Adotando como base teórica os estudos de Henriksen (1999), Nation (2001) e Daller (2007), fizemos uma explanação sobre as concepções de conhecimento lexical, compreendendo seus diferentes níveis/dimensões. Assim, identificamos que um índice de grande confiabilidade para a avaliação e acompanhamento desse conhecimento é a riqueza lexical, composto por alguns indicadores linguísticos, dentre eles, a densidade e a diversidade lexical.

Nos estudos de Ure (1971) e Halliday (1985, 1989), exploramos os conceitos de densidade lexical, assim como as principais medidas para sua mensuração. Sobre a diversidade, em Thompson (1915) e Carroll (1938) encontramos as primeiras definições e nos situamos sobre as métricas propostas para sua mensuração ao longo dos anos, como a TTR, RTTR, CTTR, medidas D e VOCD. No entanto, constatamos uma problemática antiga em torno da TTR, RTTR e CTTR, devido às suas limitações para avaliação de textos de diferentes dimensões. Para superar essa deficiência, duas novas técnicas foram desenvolvidas, a medida D e VOCD, consideradas como as mais confiáveis para examinar a diversidade lexical de um texto.

Em nossa análise bibliográfica, também verificamos os resultados de alguns estudos que avaliaram o conhecimento lexical de alunos de diferentes nacionalidades, idades e níveis escolares a partir da avaliação da densidade e diversidade lexical de seus textos. Assim, identificamos que alguns fatores como a idade, nível escolar, idioma, gênero textual e proposta metodológica de escrita podem contribuir para densidade e diversidade do texto.

Através de uma abordagem metodológica de léxico-estatística textual, quantificando o número de PE, DeL e DiL das produções textuais das 5 díades brasileiras. Submetendo os valores obtidos a uma análise estatística descritiva e inferencial, buscamos responder às nossas questões de pesquisa: i) Qual a taxa de palavras escritas,

de densidade e diversidade lexical em manuscritos escolares produzidos por díades de alunos recém-alfabetizados quando escrevem seus primeiros contos etiológicos? ii) Existe correlação entre taxas de palavras escritas, de densidade e diversidade lexical desses contos? iii) Há diferença significativa entre as díades, e termos de PE, DeL e DiL?

Através dos nossos resultados, constatamos que as díades escreveram seus textos com uma média de 90 palavras, usando uma densidade lexical de 48% e diversidade lexical de 49,40%. Além disso, também não foi encontrada correlação entre PE, DeL e DiL, sugerindo que são categorias independentes e merecem uma avaliação individual no texto.

Outro resultado importante é que na comparação entre as produções das díades, não foi encontrada diferença significativa, ou seja, os alunos usaram taxas semelhantes de PE, DeL e DiL para produzirem seus textos, confirmando assim, a nossa hipótese principal. Ademais, comparando o valor da média geral de DeL e DiL encontramos uma aproximação entre elas.

Com esses resultados, foi possível alcançar as respostas para as nossas questões de pesquisa e verificar as nossas hipóteses acerca dos textos das duplas analisadas, todavia, para uma maior confirmação da generalização desses resultados é necessário aplicar essa mesma técnica em um *corpus* maior, seguindo a mesma proposta metodológica de escrita colaborativa.

Contudo, também vale ressaltar que para um maior acesso acerca das escolhas lexicais das duplas é importante considerar, além da avaliação dos textos físicos, a avaliação do processo de escrita em curso, isto é, as imagens e os áudios que mostram as diversas situações de negociações, tensões e decisões.

Por fim, esperamos que os resultados do nosso trabalho possam contribuir para a produção de conhecimento acerca desses indicadores linguísticos, como também proporcionar aos professores novas possibilidades metodológicas para auxiliar no ensino da produção textual em sala de aula e oferecer novos parâmetros para avaliação do conhecimento lexical dos alunos.

REFERÊNCIAS

ARNAUD, P. J. **The Lexical Richness of L2 Written Productions and the Validity of Vocabulary Tests.** In: Practice and Problems in Language Testing, vol. 7, p. 14-28,1984.

BERMAN, R., VERHOEVEN, L. **Cross-Linguistic Perspectives on the Development of Text-Production Abilities in Speech and Writing**. Written Languages, vol. 5, p. 1-43, 2002.

BEZERRA, M. A. **Dificuldade no uso adequado de vocabulário em textos escolares escritos**. In: LEFFA, V. J. As palavras e suas companhias: o léxico na aprendizagem da língua. Pelotas: EDUCAT, p, 217-228, 2000.

BIDERMAN, M. T. **Teoria Linguística: Linguística Quantitativa e Computacional**. Rio de Janeiro: Livros Técnicos e Científicos, 1978.

CALIL, E. **Autoria: a criança e a escrita de histórias inventadas**. Londrina: Ed. da UEL, 2009.

CALIL, E. **O sentido das palavras e como eles se relacionam com o texto em curso: estudo sobre comentários semânticos feitos por uma díade de alunas de 7 anos de idade**. ALFA: Revista de Linguística (UNESP. Online), vol. 60, p. 531-555, 2016.

CALIL, E.; PEREIRA, L. A. **Reconhecimento antecipado de problemas ortográficos em escreventes novatos: quando e como acontecem**. Alfa: revista de linguística. Universidade Estadual de São Paulo. vol. 62, p. 91-123, 2018.

CARROLL, J. B. **Diversity of vocabulary and the harmonic series law of word-frequency distribution**. The Psychological Record, vol. 2, p. 379-386, 1938.

CARROLL, J. B. **Language and Thought.Englewood** Cliffs. Prentice-Hall, Inc. Francais Moderne, vol. 46, p. 25-32, 1964.

CORDEIRO, M.; CALIL, E. **Estudo comparativo das rasuras orais comentadas em processos de escritura de duas díades de alunos recém alfabetizadas**. Revista Educação e Linguagem vol. 7, p. 107-128, 2018.

COSTA U.R.C. **Aspectos da complexidade gramatical uma contribuição para o ensino**. Dissertação em Linguística e Ensino da Língua Portuguesa. Universidade Federal do Ceará, 1997.

COVINGTON, M. A., MCFALL, J. D. **Cutting the Gordian Knot: The Moving-Average Type-Token Ratio (MATTR)**. Journal of Quantitative Linguistics ,vol. 17, n. 2, p. 94-100, 2010.

DALLER, H., MILTON, J., TREFFERS-DALLER, J. **Modelling and assessing vocabular knowledge**. Cambridge University Press, Cambridge, 2007.

DÓCZI, B., KORMOS, J. **Longitudinal developments in vocabulary knowledge and lexical organization.** Oxford University Press, New York, 2016.

FELIPETO, C. **Escrita colaborativa e individual em sala de aula: uma análise de textos escritos por alunos do ensino fundamental**. ALFA: Revista de Linguística (UNESP ONLINE), v. 63, p. 133-152, 2019.

FERRAZ, A P. **O léxico do português em estudo na sala de aula II.** Araraquara: Letraria, 2017.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. - São Paulo: Atlas, 2008

GUIRAUD, P. **Les caractères statistiques du vocabulaire**. PUF, Paris,1954.

_____. **Les problèmes et méthodes de la statistique linguistique**. D. Reidel Publishing Company / Dordrecht – Holland, 1959.

HALLIDAY, M. A. K **Spoken and written modes of meaning**. In Comprehending oral and writ- ten language, (Eds, Horowitz, R. & Samuels, S.J.) Academic Press, Orlando, 1985.

_____. **Spokcn and writtcn language**. Oxford University Press, 1989.

JABORANDY, B., LOPES, A., CALIL, E. **Análise de manuscritos escolares por meio de tópicos de intertextualidade e criatividade**. Entretextos (UEL), vol. 17, p. 77-104, 2017.

HARMON, J., WOOD, K. D., MEDINA, A. **Vocabulary learning in the content areas: Research-based practices for middle and secondary school classrooms**. In K. D. Wood & W. E. Blanton (Eds.), Literacy instruction for adolescents: Research based practice. New York, p. 344–367, 2009.

HENRIKSEN, B. **Three Dimensions of Vocabulary Development**. In: Studies in Second Language Acquistion (SSLA), vol. 21, p. 303-317, 1999.

JARVIS, S. **Short texts, best fitting curves, and new measures of lexical diversity**. Language Testing *19*, 57-84, 2002.

_____. **Capturing the Diversity in Lexical Diversity**. Language Learning Research Club, University of Michigan, p. 87- 106, 2013.

JOHANSSON, V. **Lexical density and lexical density in speech and writing: a developmental perspective**. Working Papers, Lund University, p. 61-79, 2009.

JOHNSON, W. **Studies in Language Behavior**. A. program of research, Psychol, p. 1-15, 1944.

KOCH, I. G.; ELIAS, V. M**. Ler e escrever: estratégias de produção textua**l. 2. Ed. São Paulo: Contexto, 2010.

LANSOR, R., FARBER, B. **Estatística Aplicada**. Tradução: Luciane Ferreira e Paulo Viana. Ed. 4, São Paulo: Pearson Pretince Hall, 2010.

LEITE, G., SANTOS, E., CALIL, E.**, Uma ferramenta para auxiliar na análise lexical em produções textuais**. Alagoas, 2016.

LEKI, I., CARSON, J. **Students‟ perceptions of EAP writing instruction and writing needs across the disciplines**. *TESOL Quarterly, 28*, 81-101. 1994.

LINNARUD, M. **Lexical density and lexical variation – An analysis of the lexical texture of Swedish students written work**. University of Lund, 1973.

MAGALHÃES, M, LIMA, A. **Noções de Probabilidade e Estatística**. 7 ed. São Paulo: Editora da Universidade de São Paulo, 2010.

MALVERN, D., RICHARDS, B. **Anew measure of lexical diversity**. In Ryan, A. and Wray, A., editors, Evolving models of language. Clevedon: Multilingual Matters, p. 58 - 71, 1997.

MALVERN, D., RICHARDS, B., CHIPERE, N., DURÁN P. **Lexical Diversity and Language Development: Quantification and Assessment**. Basingstoke, Hampshire: Palgrave Macmillan, 2004.

MARTINS, M. **Complexidade textual e progressão escolar em dois registos: um estudo de correlação baseado em um *corpus quasi*-longitudinal**. Tese (Doutorado em Linguística) - Faculdade de Letras, Universidade de Lisboa, Lisboa, 2016.

_____. **Densidade Lexical na escrita de textos escolares**. IGNUM: Estud. Ling., Londrina, n. 20/1, p. 218-240, 2017.

MARZANO, R., PICKERING, D. **Building academic vocabulary: Teacher's manual**. Alexandria, VA: Association for Supervision and Curriculum Development, 2005.

MCCARTHY, M., JARVIS, S. **VOCD: A theoretical and empirical evaluation**. Language Testing. University of Memphis, USA, and Ohio University, USA. vol. 24 p. 459–488, 2007.

_____. **MTLD, VOCD-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment.** Behavior Research Methods. University of Memphis, vol. 42, p. 381-392, 2010.

MCKEE. G., MALVERN, D., RICHARDS, B. **Measuring vocabulary diversity using dedicated software**. Literary and Linguistic Computing, vol. 15, p. 323-337, 2000.

MORETTIN, A., BUSSAB, W. Estatística Básica. Pedro - 6. ed. – São Paulo: Saraiva, 2010.

ME'NARD, N. **Measure de la richesse lexicale**. Geneva: Slatkine, 1983.

MILLER, J. F. **Assessing language production: experimental procedures**. London, 1981.

MOORE, D. **The Basic Practice of Statistics**. Third Edition, 2003.

MORETTIN, P. A., BUSSAB, O. W. **Estatística Básica**. 6. ed. – São Paulo: Saraiva, 2010.

NATION, I. **Learning Vocabulary in Another Language**. Cambridge University Press, 2001.

OLINGHOUSE, N. G., WILSON, J. **The relationship between vocabulary and writing quality in three genres.** Reading and Writing, 26(1), p.45 - 65, 2012.

O'LOUGHLIN, K. **The equivalence of direct and semi-direct speaking tests*:* Studies in language testing 13**. Cambridge, U.K, *2001*.

RAIMES, A. **What unskilled ESL students do as they write: A classroom study of composing**. TESOL Quarterly, 19(2), 229-258, 1985.

READ, J. **Assessing vocabulary**. Cambridge University Press, p.188-210, 2000.

REHDER, B., CHREINER, M., WOLFE, M., LAHAM, D., LANDAUER, T., KINTSCH, W. **Using latent semantic analysis to assess knowledge: Some technical considerations**. Discourse Processes, 1998.

RIFFO, K., OSUNA, S., LAGOS, P. S. **Lexical Diversity and Lexical Density Description of News Written by Journalism Student**. Rev. Bras. Linguíst. Apl., vol. 19, n. 3, p. 499-528, 2019.

RODRIGUES, S. **Escrita espontânea: Desenvolvimento das capacidades de composição escrita de em crianças do 1º ao 4º ano de escolaridade**. (Dissertação de Mestrado). Universidade Fernando Pessoa-Porto, p.107, 2008.

SADEGHI, K.; DILMAGHANI, S. **The Relationship between Lexical Diversity and Genre in Iranian EFL Learners Writings**. Journal of Language Teaching and Research, vol. 4, n. 2, p. 328-334, 2013.

SANTOS, E., CALIL, E., PEREIRA, L., COIMBRA, R. **Diversidade e densidade lexical em textos escritos por alunos recém-alfabetizados: um estudo descritivo de produções individuais e em díades.** Calidoscópio-Unisinos vol. 16, n. 1, p. 25-32, 2018.

SCHERER S, CASARIM F, ZART P, RAMOS A**. Perfil evolutivo da relação type/token de crianças de 3, 4 e 5 anos de idade**. Porto Alegre: Trabalho de especialização, CEFAC 2002.

SCHERER, S., SOUZA, A. S. ***Types* e t*okens* na aquisição típica de linguagem por sujeitos de 18 a 32 meses falantes do português brasileiro.** Revista CEFAC, vol. 12, n. 5, p. 838-845, 2011.

SOO-KYUNG, P. **Lexical analysis of korean university student's narrative and argumentative Essays**. Korea University: *English Teaching*, vol. 68, n. 3, 2013.

STROMQVIST, V., JOHANSSON, V., KRIZ, H., AISENMAN, R., RAVID, D. **Toward a cross-linguistic comparison of lexical quanta in speech and writing. I**n: Written Language and Literacy, vol. 5, p. 45-67, 2002.

TEMPLIN M. C. **Certain language skills in children: their development and interrelations**. Westport, CT: Greenwood, 1957.

THOMSON, G., THOMPSON, J., **Outlines of a method of the quantitative analysis of writing vocabularies**. British Journal of Psychology, vol. 8, p. 52–69, 1915.

TREFFERS-DALLER, J. **Measuring lexical diversity among L2 learners of French: an exploration of the validity of D, MTLD and HDD as measures of language ability**. In: Jarvis, S. and Daller, M. (eds.) Vocabulary knowledge: human ratings and automated measures. Benjamins, Amsterdam, p.79 -104, 2013.

TRIOLA, F. M. **Elementary Statistics**. Twelfth Edition, 2013.

URE, J. **lexical density and register differentiation**. In: applications of linguistics. selected papers of the second international congress of applied linguistics. Cambridge, p, 443-452, 1971.

UZAWA, K., CUMMINGS, A. **Writing strategies in Japanese as a foreign language: Lowering or keeping up the standards**. The Canadian Modern Language Review, vol. 46, p.179–191, 1989.

VYGOTSKY, L., LEONTIEV, L. **Linguagem, desenvolvimento e aprendizagem**. São Paulo: Ícone, 1988.

WANG, X. **The relationship between lexical diversity and EFL writing proficiency**. University of Sydney, Australia, Papers in TESOL, p. 65-88, 2006.

WOERFEL, T.; YILMAZ, S. **Lexical development of German-Turkish bilinguals: A comparative study in written discourse** In Chris Cummins et al. (eds.) Proceedings of the 6th Cambridge Postgraduate Conference in Language Research. Cambridge: Cambridge Institue of Language Research. p. 240-251, 2011.

ZIPF, G. The **Psychobiology of Language: An Introduction to Dynamic Philology**. Cambridge, Mass.: M.I.T. Press, 1935.

ANEXOS

Anexo 1 – Relação dos contos sugeridos para a leitura

Ordem	Contos lidos	Autores
1	A Festa no céu: Um conto do nosso folclore	(LAGO, 2005)
2	Como surgiu o mundo e (quase) tudo o que tem nele.	(ZATZ, 2010)
3	O bico do Tucano	
4	O fedor do Gamba	
5	Como surgiram os morcegos?	
6	Peixando daqui, peixeando dali, os peixes foram morar no mar	
7	Por que os caranguejos não tem cabeça	
8	Como a lebre foi parar na Lua	
9	Como surgiram as línguas	
10	A casa do caracol	
11	O pescoço da girafa	
12	Por que o coelho tem orelhas compridas	
13	Amigos, mas não para sempre	(BARBOSA, 2004)
14	O Jabuti de Asas	
15	A pele nova da mulher velha	
16	Por que o sol anda tão devagar?	
17	Por que o morcego só voa de noite	
18	Por que o porco vive no chiqueiro	
19	Por que o camaleão muda de cor	
20	Por que o cachorro foi morar com o homem	
21	Por que a zebra é toda listrada	
22	Por que a girafa não tem voz	
23	Por que o macaco se esconde nas árvores	
24	Como os primeiros homens nasceram	(LODY, 2009,
25	Ogum: Aquele que veio para ensinar	
26	Odé, o caçador	
27	Catendê, o dono das folhas	
28	Quianda e Quicimbe	
29	Ibejis, os gêmeos	
30	Por que o cachorro e inimigo de gato... E gato de rato	(CASCUDO, 2006)
31	A goela e o rabo da baleia	
32	Por que a galinha-d'angola tem pintas brancas?	(BARBOSA, 2010)
33	Por que o porco tem o focinho curto?	
34	Por que o sol e a Lua foram morar no céu	(BRAZ, 2005)
35	A origem da morte	

ANEXO 2 - Produção Tema Livre 001_D1_01

PRODUÇÃO DE TEXTO

O ganba que mentia

~~Era~~ Era uma vez um ganbá muito mentiroso
ele gostava muito de engana ~~os~~ os animais.
Mais um dia a ~~ra~~ ratinha que se chamava
Lua viu ele mentindo pros animais e ela ~~decidio~~
~~demiti ele~~ e desidiram coloca nele um fedo que
quando ele mentia o fedo saia dele* ~~mais~~
~~mesmo ele fez isso sinão sinão a raina~~
~~sinão a rainha ia demiti ele~~ *
e ele foi demitido porque ele mintio novamente

ANEXO 3 - Produção Tema Sugerido 002_D1_06

PRODUÇÃO DE TEXTO

Por que o cachorro faz au,au?
E o gato faz miau?

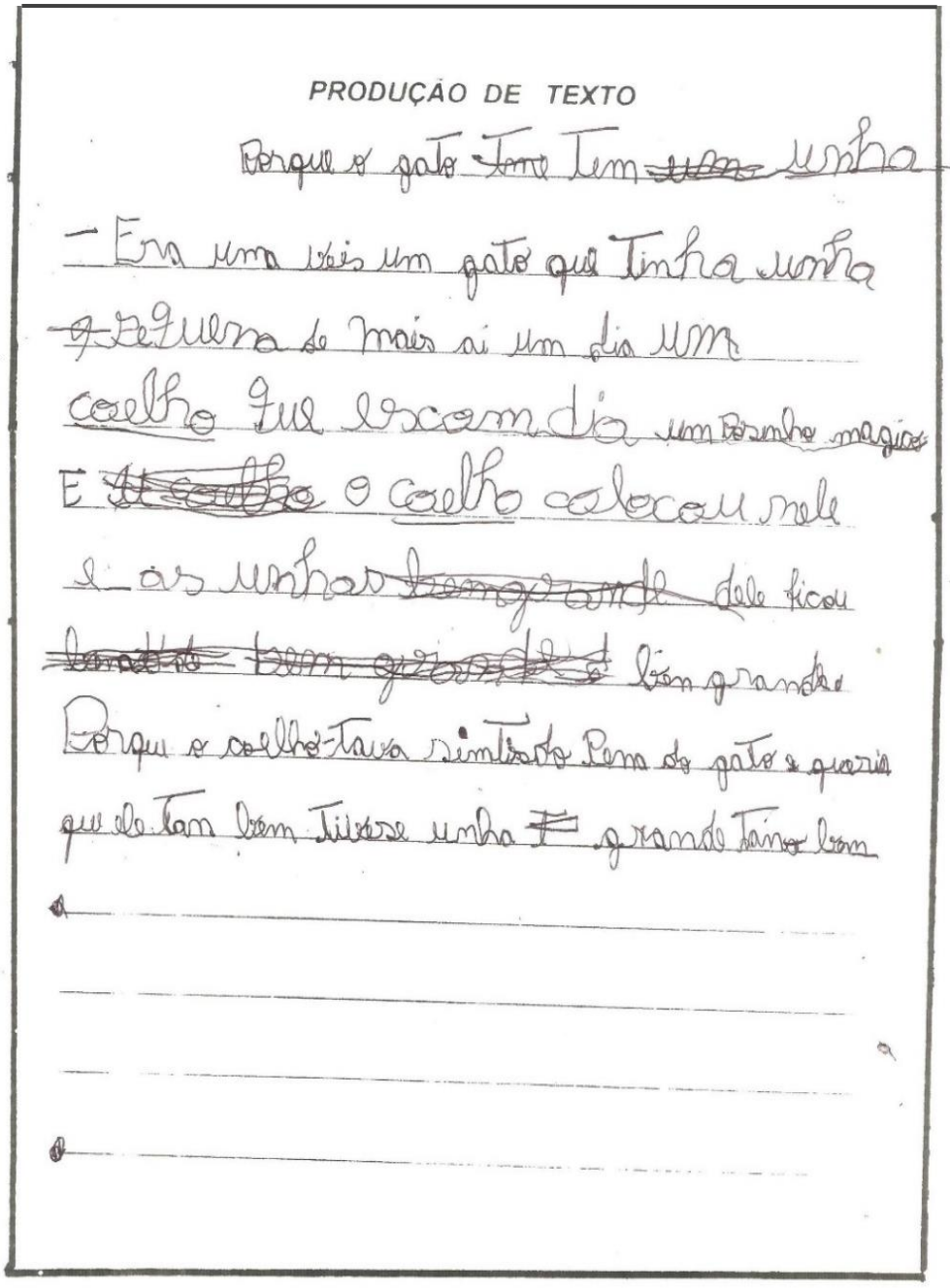
Era uma vez um gato e ~~o cachorro~~ uma cachorra o ~~gato~~ cachorro gostava muito de core atrais do gato. E o gato não ~~aguentava~~ core muito é sim aguentava core muito é sim, e o cachorro parava e ia para casilha dele. Ele foi ~~dormi~~ dormi da casilha dele E ele dormiu até o outro dia e esqueceu do gato e o gato ficou muito Bem.

~~o cachorro late Por que o gato late o gato mia Por que o gato late~~

~~Por que o cachorro faz au, au?~~

~~E o gato faz miau?~~ Por que eles brigavam. E Miau para os cachorros

ANEXO 4 - Produção Tema Livre 003_D1_11



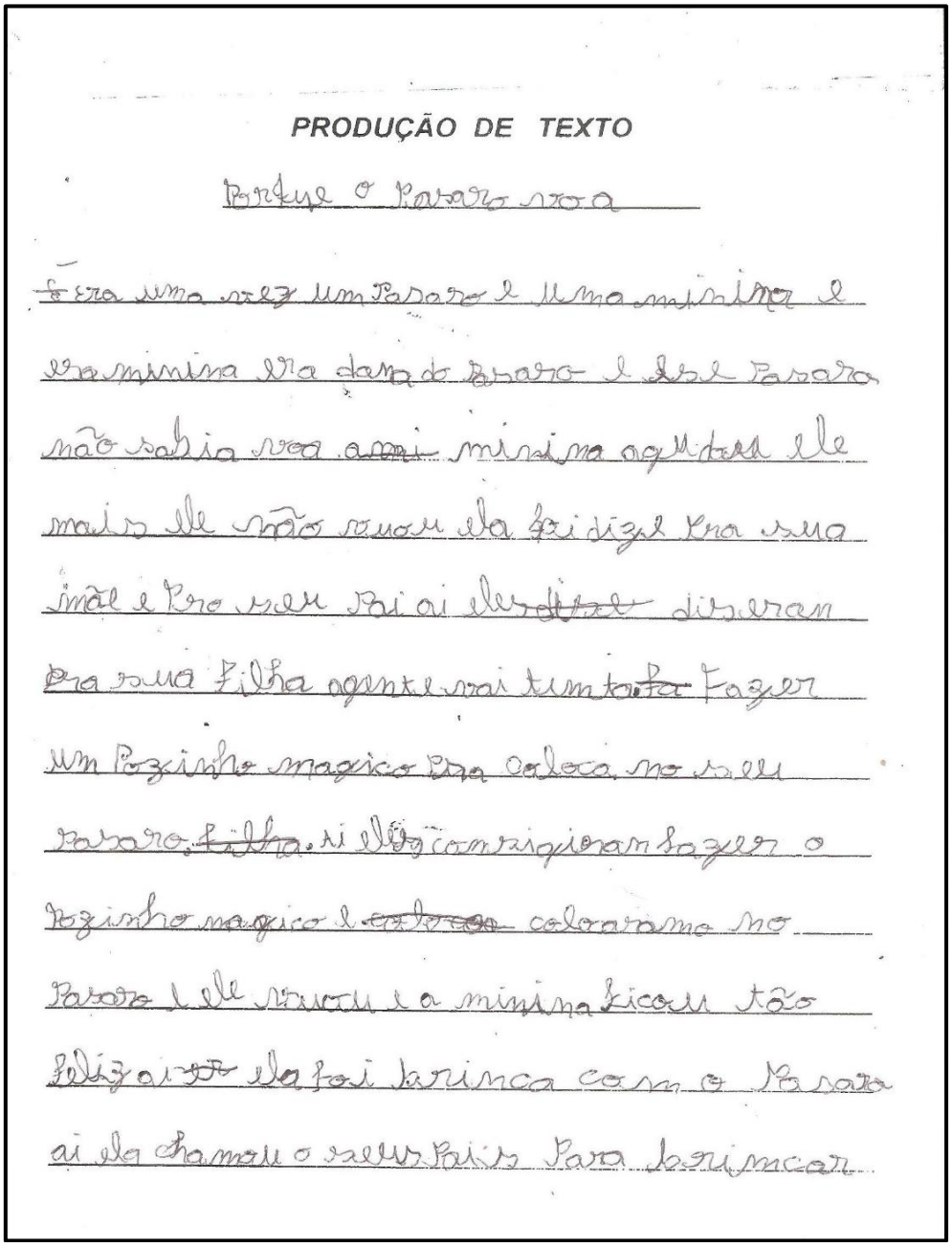
PRODUÇÃO DE TEXTO

Porque o gato ~~Jane~~ Tem ~~uma~~ unha grande

– Era uma vez um gato que Tinha unha ~~grande~~ grande de mais aí um dia um coelho que se chamava um pouco magico E ~~o coelho~~ o coelho colocou nele as unhas ~~bem grande~~ dele ficou ~~bem grande~~ bem grande.

Porque o coelho Tava sintindo Pena do gato e queria que ele Tem bem Tivesse unha ~~F~~ grande Tinha bem

ANEXO 5 - Produção Tema Livre 004_D1_16



PRODUÇÃO DE TEXTO

Porque o Papagaio voa

Era uma vez um Papagaio e uma menina e essa menina era dona do Papagaio e esse Papagaio não sabia voar a menina ajudava ele mais ele não voava ela foi dizer pra sua mãe e pra ver pai aí eles disseram pra sua filha agente vai tentar fazer um Pozinho magico para colocar no seu papagaio. filha. aí eles começaram fazer o pozinho magico e colocaram no Papagaio e ele voou e a menina ficou tão feliz aí ela foi brincar com o Papagaio aí ela chamou o seu Pai e Para brincar

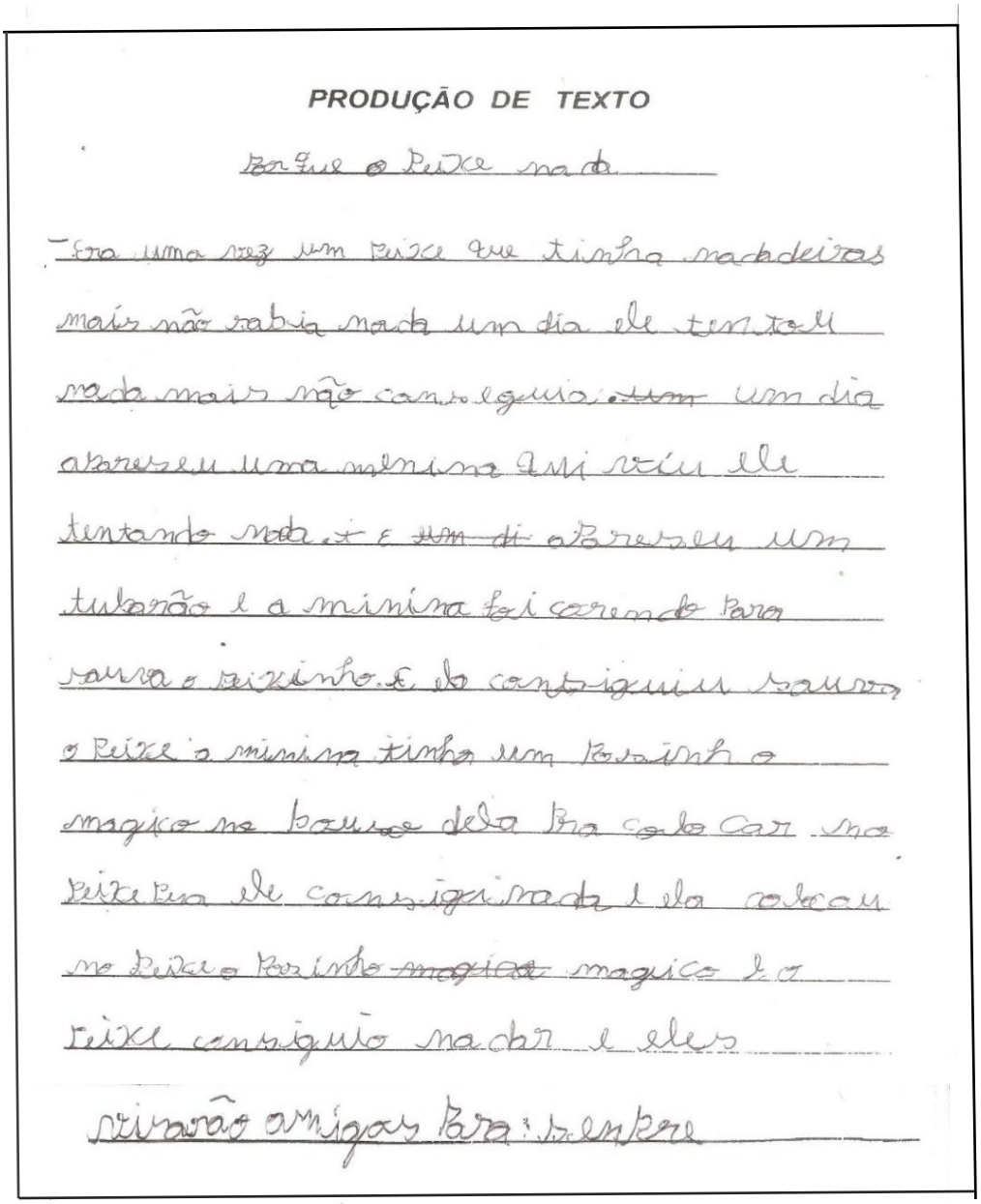
ANEXO 9 - Produção Tema Sugerido 005_D1_21

PRODUÇÃO DE TEXTO

como sugestão as ~~letras~~ Palavras

– Era uma vez duas meninas qui erão ~~irmas~~
irmãs. E elas forão para uma casa ~~[illegible]~~ e la
avia muitas letras e essas letras não
se encachavão aí a moça que morava la
virou amiga das duas meninas e ela dise
vocês querem tiro essas letras com migo
la e as meninas ~~[illegible]~~ diserão sim. E todas tirarão
as letras aquela moça pensava qui ~~as~~ aquelas
letras não formavão Palavras. E elas
formarão muitas Palavras.

ANEXO 7 - Produção Tema Sugerido 006_D1_26

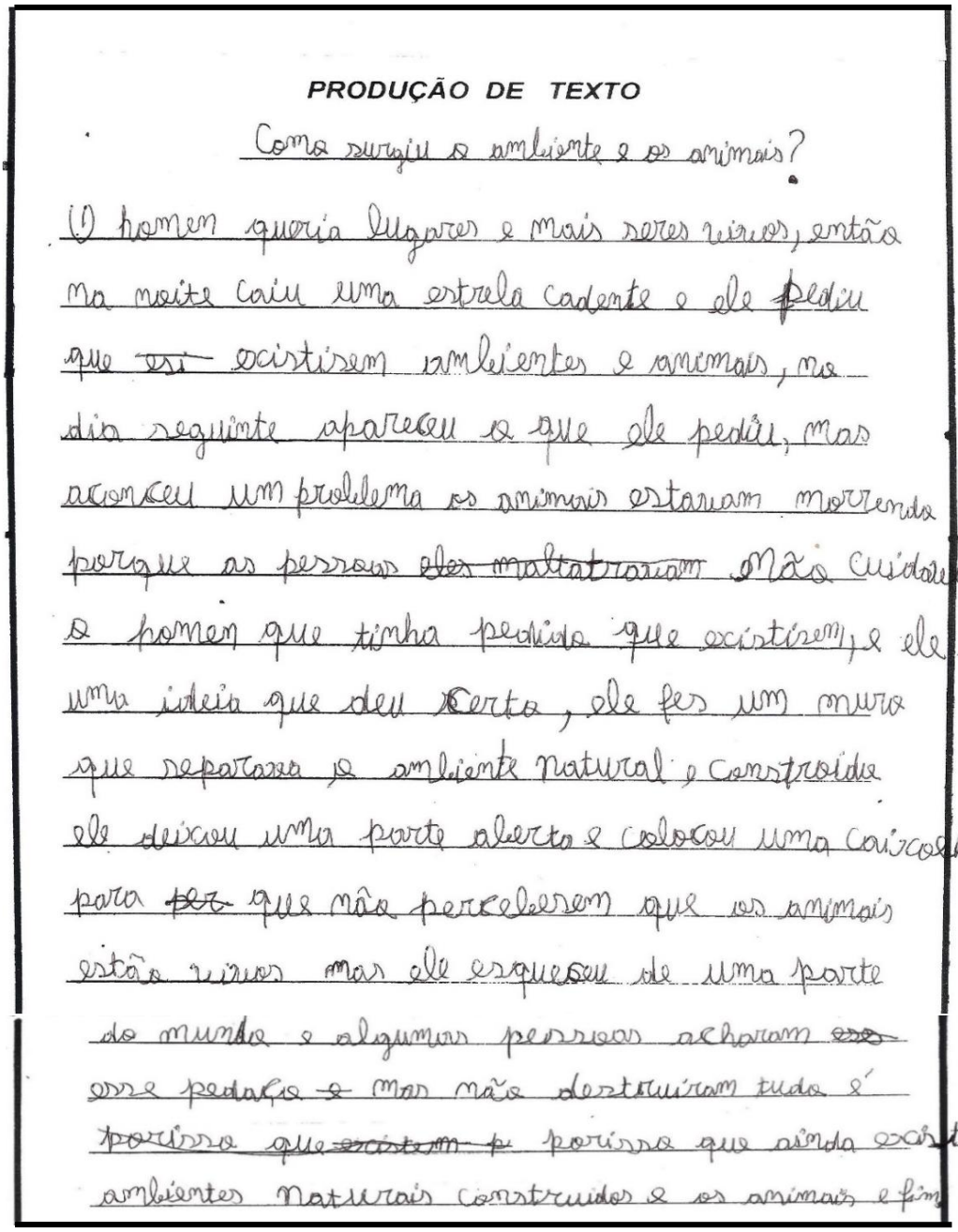


PRODUÇÃO DE TEXTO

O Peixe nada

Era uma vez um Peixe que tinha nadadeiras mais não sabia nada um dia ele tentou nada mais não conseguia. um dia apareceu uma menina ai viu ele tentando nada. e apareceu um tubarão e a minina foi correndo para salva o peixinho. e ela conseguiu salvar o Peixe a minina tinha um Pozinho magico na bolsa dela Pra colocar no Peixe Pra ele conseguir nadar e ela colocou no Peixe o Pozinho magica e o Peixe conseguio nadar e eles viraram amigos Para sempre

ANEXO 8 - Produção Tema Livre 001_D2_02



PRODUÇÃO DE TEXTO

Como surgiu o ambiente e os animais?

O homem queria lugares e mais seres vivos, então
na noite caiu uma estrela cadente e ele pediu
que ~~os~~ existisem ambientes e animais, no
dia seguinte apareceu o que ele pediu, mas
aconteu um problema os animais estavam morrendo
porque as pessoas ~~eles maltratavam~~ não cuidavam
o homem que tinha pedido que existisem, e ele teve
uma ideia que deu certa, ele fes um muro
que separasse o ambiente natural e construido
ele deixou uma parte aberta e colocou uma cachoeira
para ~~ver~~ que não percebesem que os animais
estão vivos mas ele esqueceu de uma parte
do mundo e algumas pessoas acharam
esse pedaço e mas não destruiram tudo é
porisso que ~~existem~~ porisso que ainda existem
ambientes naturais construidos e os animais e fim

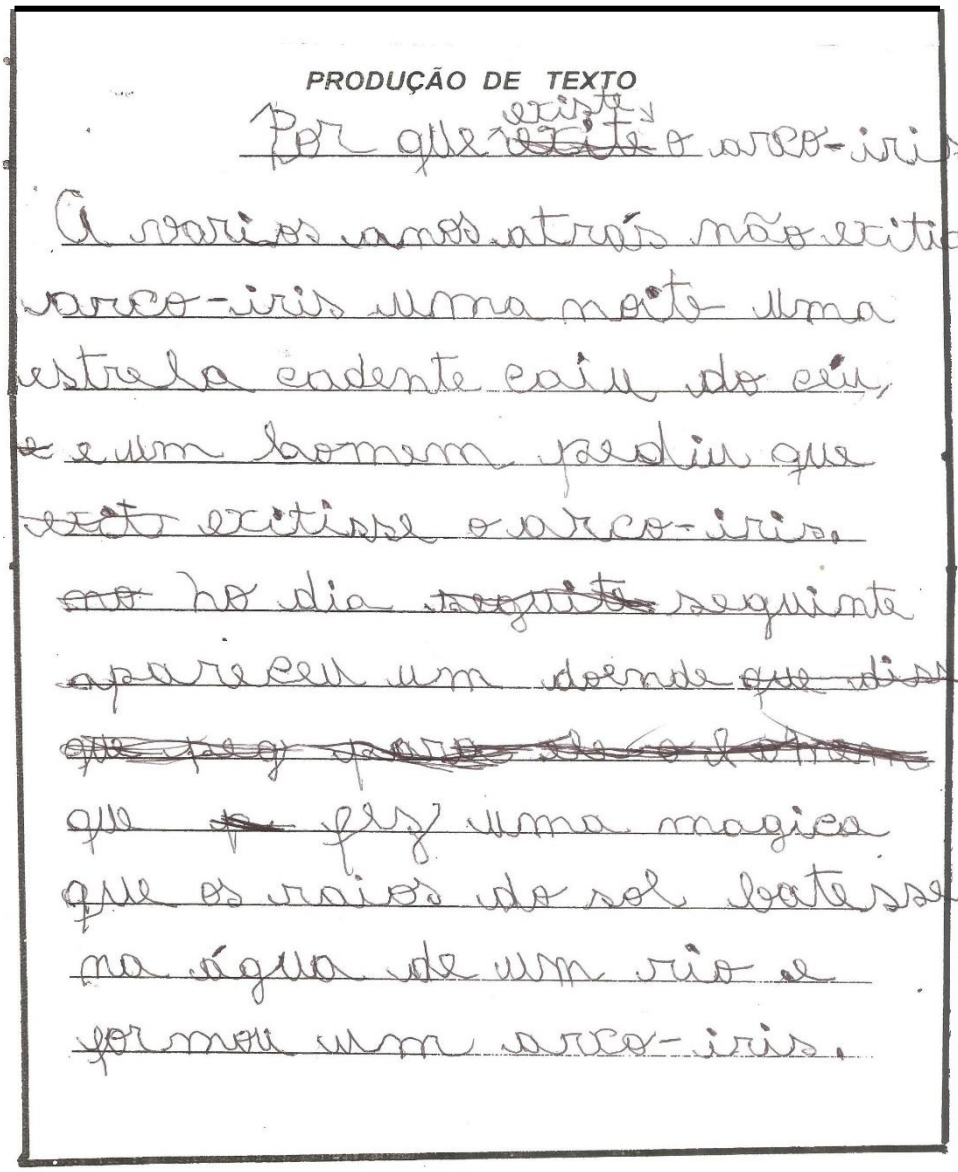
ANEXO 9 - Produção Tema Sugerido 002_D2_07

PRODUÇÃO DE TEXTO

Porque o carhorro faz au au?
E o gato faz miau?

Numa casa vivia ~~os~~ um gato
e um cachorro ~~cacho~~. Um homem queria
saber o sons que ~~o~~ eles faziam.
O gato e o ~~cachoro~~ Cachorro ~~fal~~ fa-
laram que ~~si~~ se ~~om~~ ~~ome~~ o homem
desse um susto neles ~~ele~~ Ele fez ~~i~~
~~s~~ isso ~~o~~ para ~~os~~ eles fazerem
o son deles, ~~o~~ o homem conseguiu
depois de fazer doze vezes,
~~ele conse~~ e foi assim que o
cachorro e o gato ~~fi~~ fazem
~~o~~ os sons deles.

ANEXO 10 - Produção Tema Livre 003_D2_012



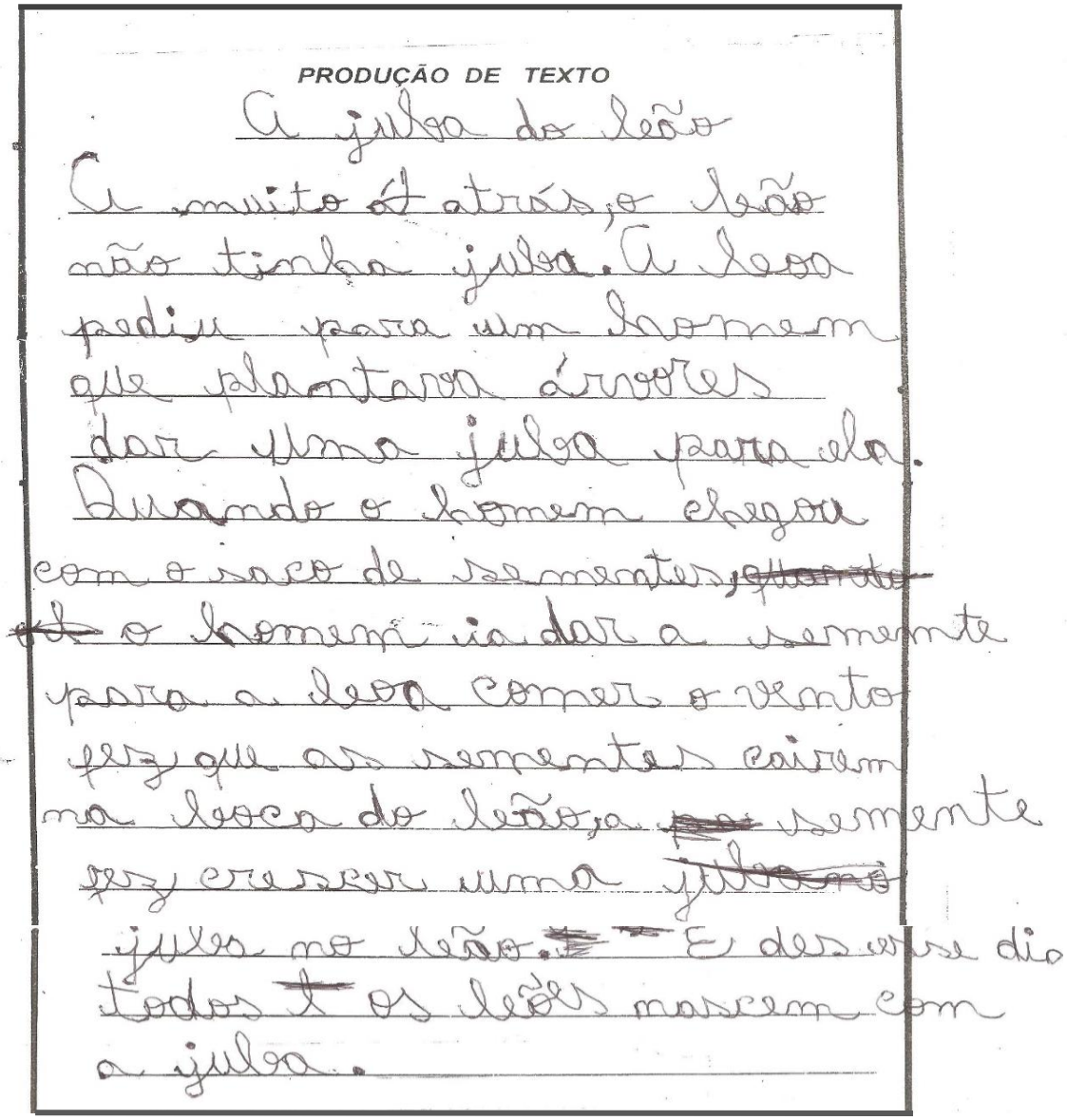
PRODUÇÃO DE TEXTO

Por que existe o arco-iris

A muito muito atrás não existia arco-iris uma noite uma estrela cadente caiu do céu, e um homem pediu que existisse o arco-iris.

no dia seguinte apareceu um doente que disse que foi uma magica que os raios do sol batesse na água de um rio e formou um arco-iris.

ANEXO 11 - Produção Tema Livre 004_D2_017



PRODUÇÃO DE TEXTO

A juba do leão

Há muito ~~é~~ atrás, o leão não tinha juba. A leoa pediu para um homem que plantava árvores dar uma juba para ele. Quando o homem chegou com o saco de sementes, ~~ele~~ o homem ia dar a semente para a leoa comer o vento fez que as sementes caíssem na boca do leão, a ~~se~~ semente fez crescer uma ~~juba~~ juba no leão. E desde dia todos ~~t~~ os leões nascem com a juba.

ANEXO 12 - Produção Tema Sugerido 005_D2_022

PRODUÇÃO DE TEXTO

Como surgiram as palavras.

Á muito tempo atrás, não existiam as palavras E as pessoas se comunicavam através de ~~telefones.~~ telefones. Um homem ~~este~~ queria escrever as palavras, mas elas não existiam, e veio dizer que tinha um mago na montanha mais alta que poderia ajudala. Para ele chegar na montanha teria que passar por florestas e rios. Na floresta ele ~~em~~ encontrou tigres, cobras, aranhas, morcegos e escorpiões, mas conseguia passar. No rio ele encontrou, cobras d'agua, crocodilos, piranhas e outros animais aquaticos, mas passou por tudo e ~~chegou~~ chegou na topo da montanha e pedio ao mago que ele criasse as palavras, e fim.

ANEXO 13 - Produção Tema Livre 006_D2_27

PRODUÇÃO DE TEXTO

Por que os peixes moram no mar ~~e~~, no rio e nos lagos.

Ai muito tempo atrás os ~~peixes~~ peixes não morarom na água, e sim na terra.

Um dia os peixes encontraram um grupo de gaviões que siguiram os peixes até o rio para ~~col~~ comelos. O gato-do-mato ~~sigu~~ seguiu ~~et~~ os ~~os~~ ~~ot~~ outros peixes do ~~gu~~ grupo até o lago para comer os ~~peix~~ peixes ~~o~~. E as gaivotas ~~si~~ seguiram eles até o mar. E todos os peixes pularam no seu ambiente aquático. E fim.

É porisso que os peixes moram no ~~m~~ mar, rios, e lagos. E fim.

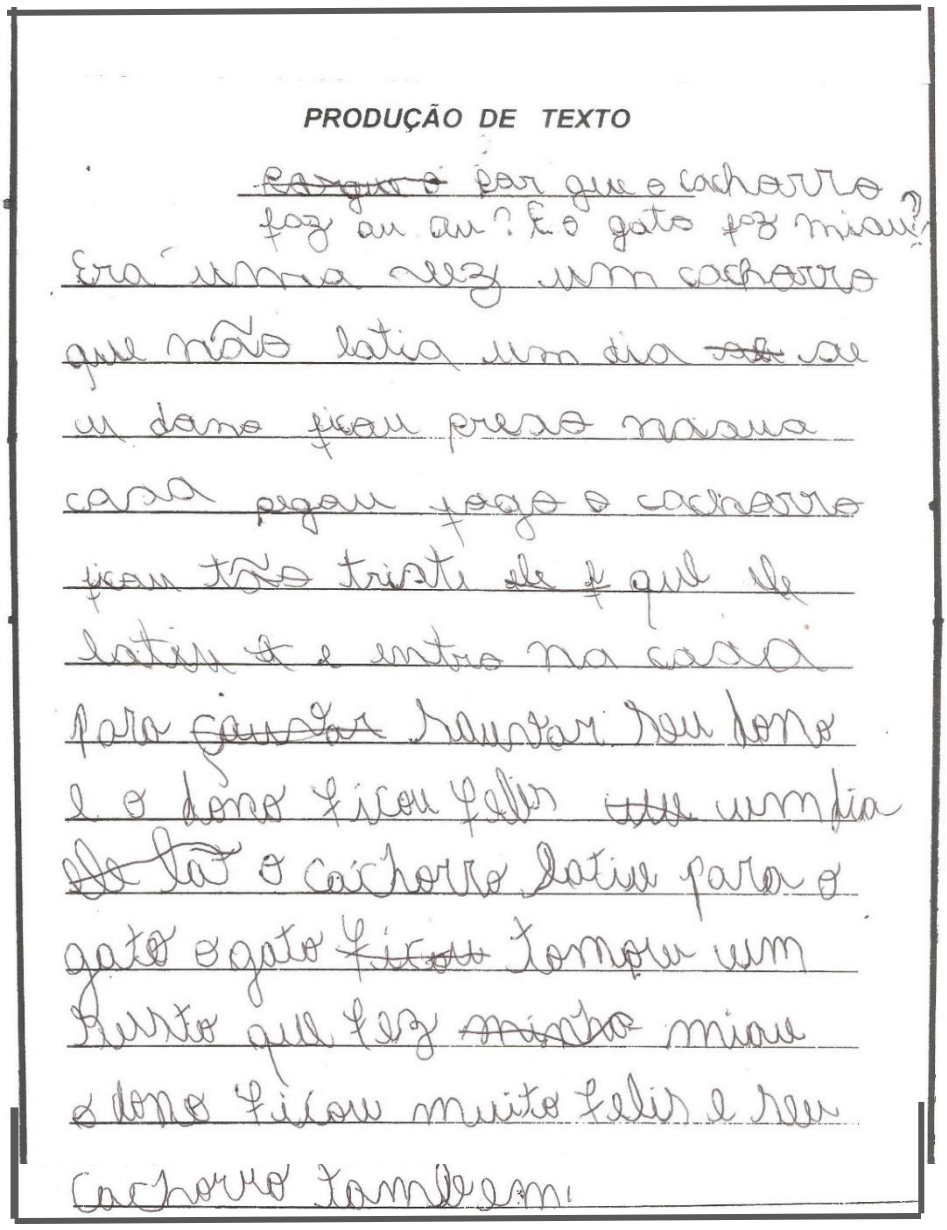
ANEXO 14 - Produção Tema Livre 001_D3_03

PRODUÇÃO DE TEXTO

O dragão que não voava

era uma ves um ~~d~~ dragão
ele era muito bom mas tinha
um proble ele não voava ele
tentou de tudu mas não consseguia
mas umdia ele Pulou da torre
para ~~assiviar~~ assaurear um leão
ele tambem mas derrepente
as brasas dele gom com
esaram a ficar muits l
angas o leão muito felis ficou
amigo do dragão quando eles
chegaram na terra todo m
undo siperguntou como o dragão
consiguil voar o dragão respo
ndeu eu sofis polar da torre
e saurear o leão i eles ficaram
felizes para sempre

ANEXO 15 - Produção Tema Sugerido 002_D3_08



PRODUÇÃO DE TEXTO

~~Cargo é~~ por que o cachorro faz au au? E o gato faz miau?

Era uma vez um cachorro que não latia um dia ~~ele~~ o dono ficou preso na sua casa pegou fogo o cachorro ficou tão triste ~~de~~ que ele latiu e entro na casa para ~~ajudar~~ salvar seu dono e o dono ficou feliz ~~ele~~ um dia ~~ele~~ o cachorro latiu para o gato e gato ~~ficou~~ tomou um susto que fez ~~miau~~ miau o dono ficou muito feliz e seu cachorro também.

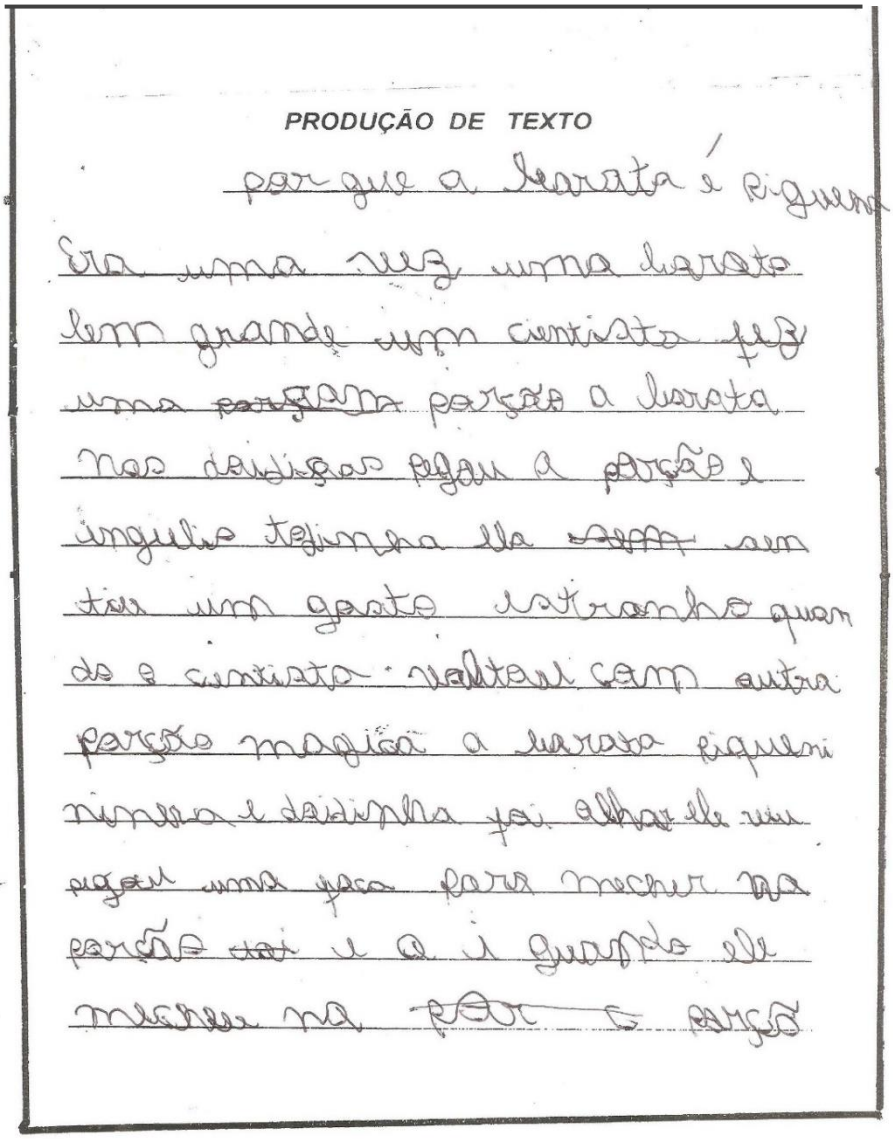
ANEXO 16 - Produção Tema Livre 003_D3_013

PRODUÇÃO DE TEXTO

Por que o leão ruge

era uma vez um leão que não
rugia ele irritava todos os animais
atarde estava jogando o coelho falou
um plano com os animais ate que enfim chegou
a noite o leão sonhou que todos os animais
riam dele o leão muito triste correu
o mais rapido que pode ate agua cara
ele acordou dosonho e foi para a
floresta quando ele abriu a boca de
deu um rugido bem auto e ele ficou
amigo dos animais.

ANEXO 17 - Produção Tema Livre 004_D3_018

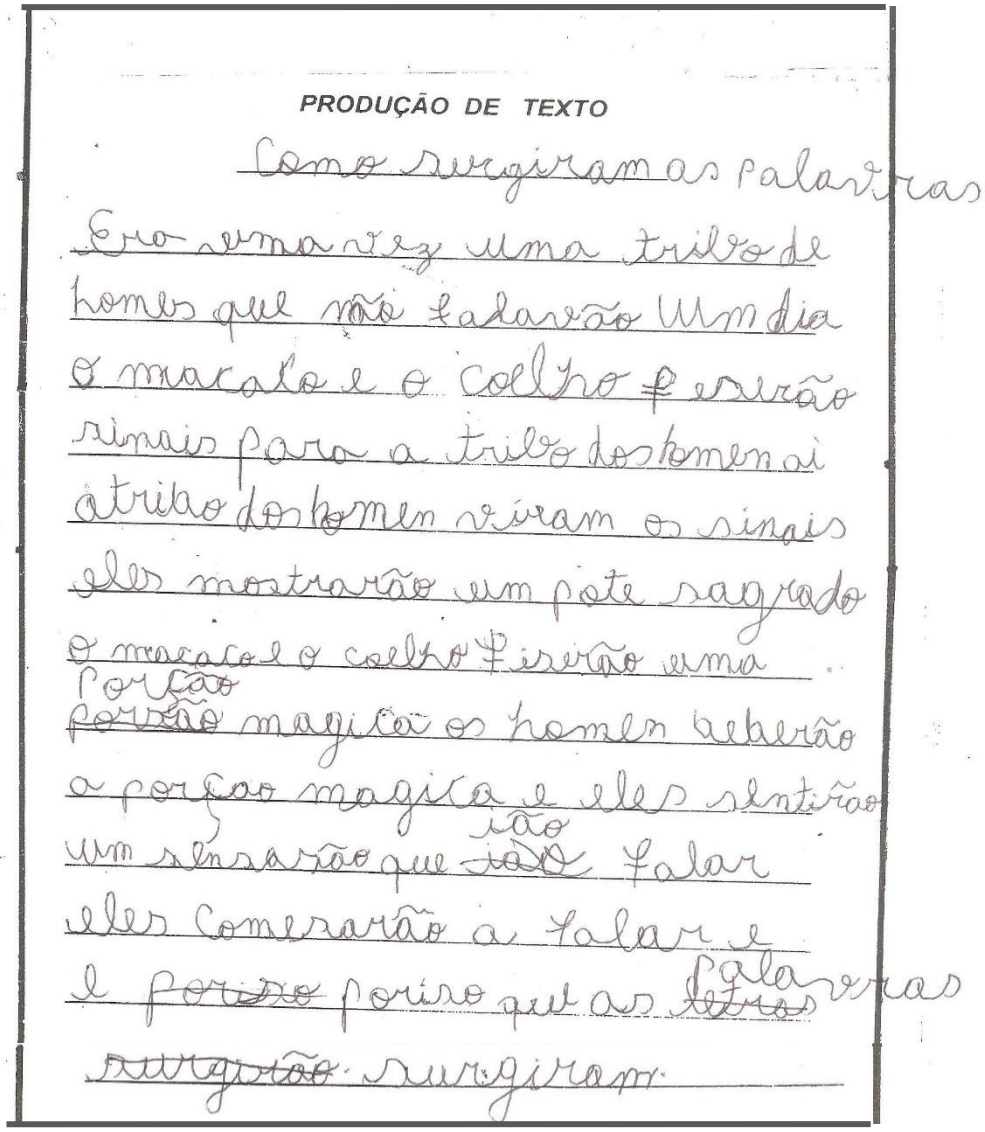


PRODUÇÃO DE TEXTO

por que a lagarta é preguiçosa

Era uma vez uma lagarta bem grande um cientista fez uma ~~poção~~ poção a lagarta nos desenhos pegou a poção e ingeriu ficou ela ~~com~~ com tipo uma gosto estranho quando o cientista voltou com outra poção mágica a lagarta esqueceu nunca e desenha foi olhar ele viu pegou uma poca para mexer na poção foi e a i quando ele mexeu na poção

ANEXO 18 - Produção Tema Sugerido 005_D3_023



PRODUÇÃO DE TEXTO

Como surgiram as palavras

Era uma vez uma tribo de homens que não falavão Um dia o macaco e o coelho ~~f~~ serão sinais para a tribo dos homen ai a tribo dos homen viram os sinais eles mostravão um pote sagrado o macaco e o coelho ~~f~~ serão uma ~~porção~~ porção magica os homen beberão a porção magica e eles sentião um sensação que ~~ião~~ não falar eles comesarão a falar e ~~l porizo~~ poriso que as ~~letras~~ palavras ~~surgirão~~ surgiram.

ANEXO 19 - Produção Tema Livre 006_D3_028

PRODUÇÃO DE TEXTO

~~porqu~~ Por que o leopardo é rápido

~~Um~~ Era uma vez um leopardo ele não conseguia
correr muito rápido um dia o coelho chamou a girafa
para pegar uma fruta para ele misturar na sua porção
ele ia dar para o leopardo mas o leopardo derrubou
a porção o rei leão deu um castigo para o leopardo o
castigo foi ele correr muito rápido ~~[illegible]~~ se ele não
conseguia se controlar toda vez que ele corria ele
ia para o lado que ele não queria mas depois ele
se acostumou e é por isso que o leopardo é rápido

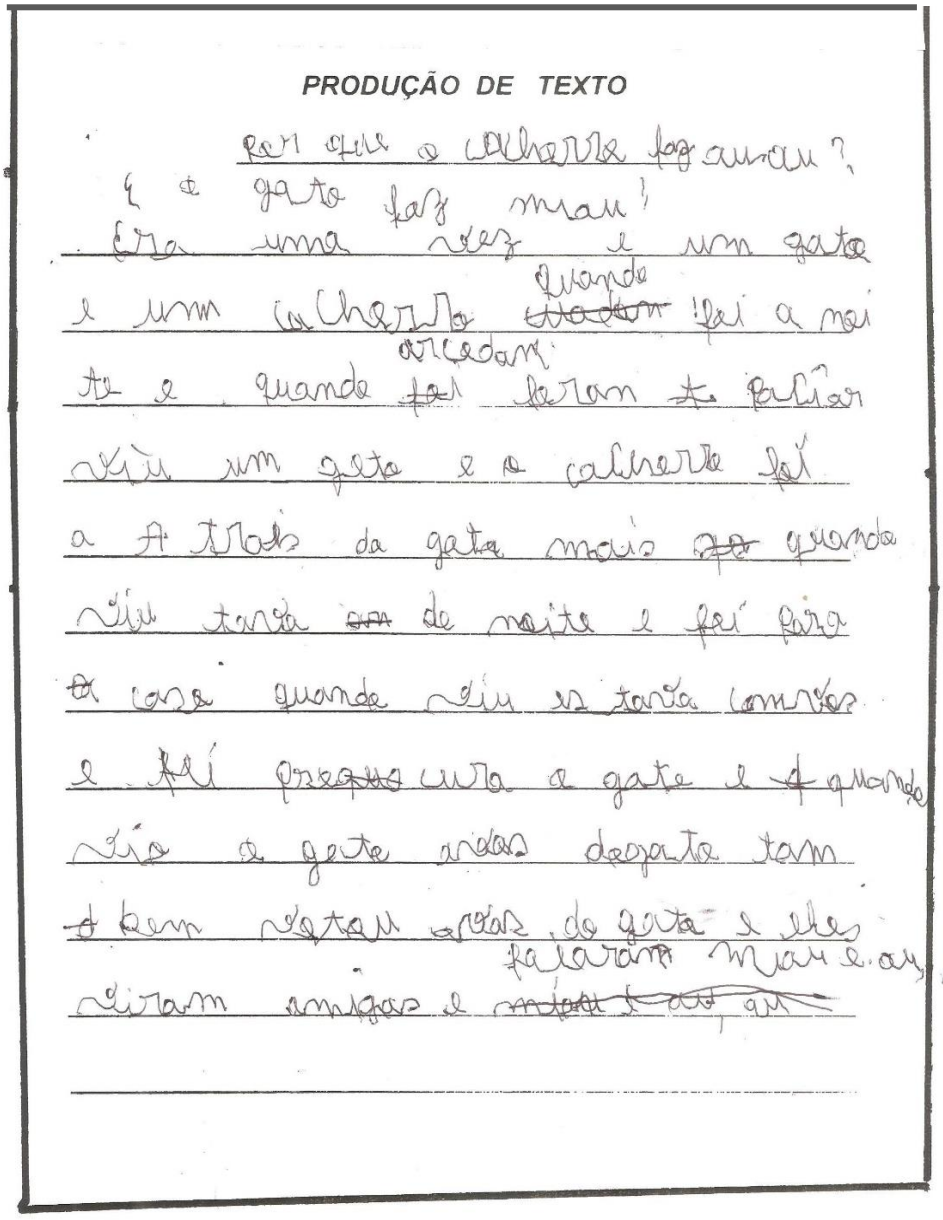
ANEXO 20 - Produção Tema Livre 001_D4_04

PRODUÇÃO DE TEXTO

Porqlle o sol brilha. a muito

tenpo atrais um indio queria qlle tivese um ~~lua~~ sol
e o idio pediu aos outros indios e comceguiram
fazer o sol e botaram no seu e comesaram a
somemoraram e depois foram dormi e no outro
dia acordaram e cuando acordaram viram que
ainda estava de noite e fizeram uma outra sol
mas ian botar fogo e encontraram um vulcam
vulcam e queimaram sol e a parecem
uma Ral cimitias do 2 paras maso

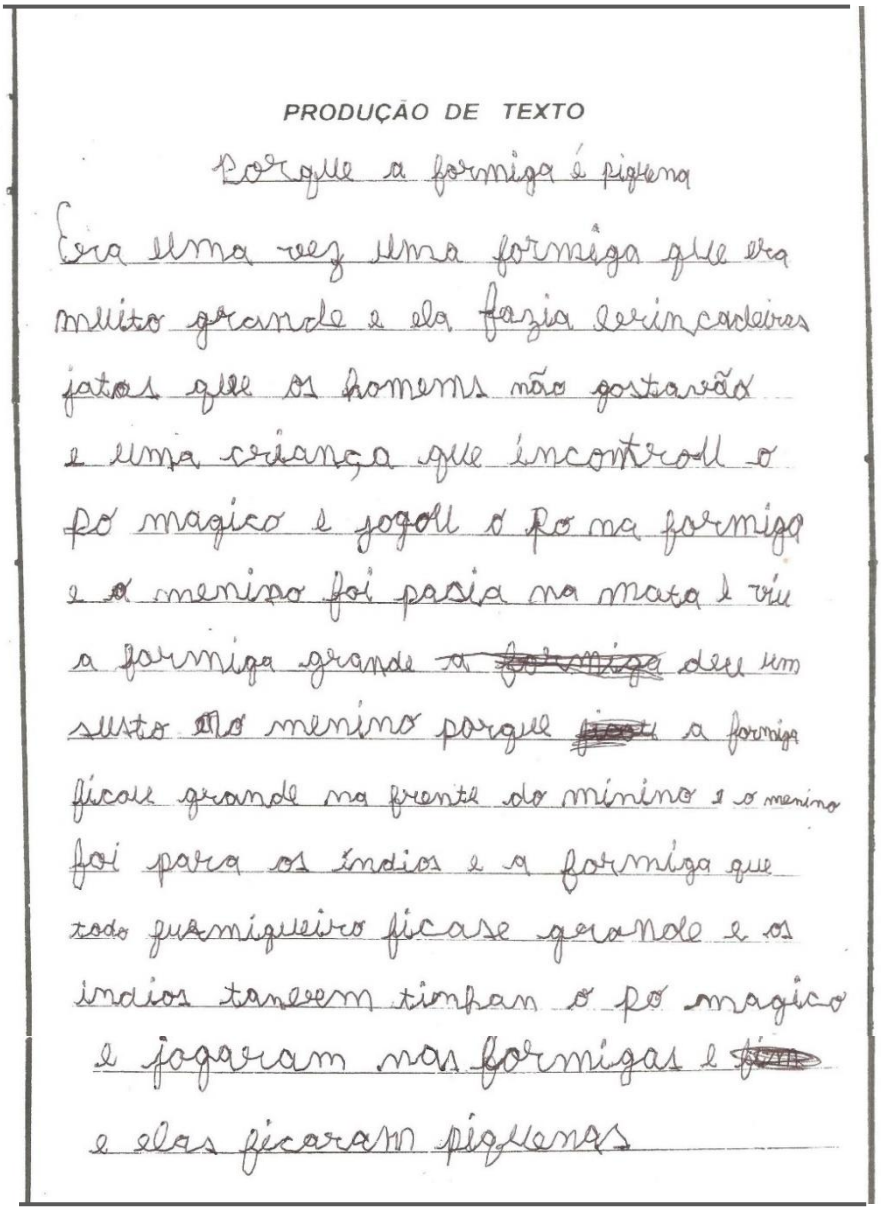
ANEXO 21 - Produção Tema Sugerido 002_D4_09



PRODUÇÃO DE TEXTO

Por que a colherde faz auau?
E o gato faz miau!
Era uma vez e um gato
e uma colherde quando
Foi a noi
te e quando foram te pegar
viu um gato e a colherde foi
a atrás da gata mais quando
viu tarde de noite e foi para
a casa quando viu a tarde comer
e foi procura a gata e quando
viu a gata viram da gata tam
bem viram atrás da gata e lhes
falaram miau e au au
viram amigas e

ANEXO 22 - Produção Tema Livre 003_D4_014

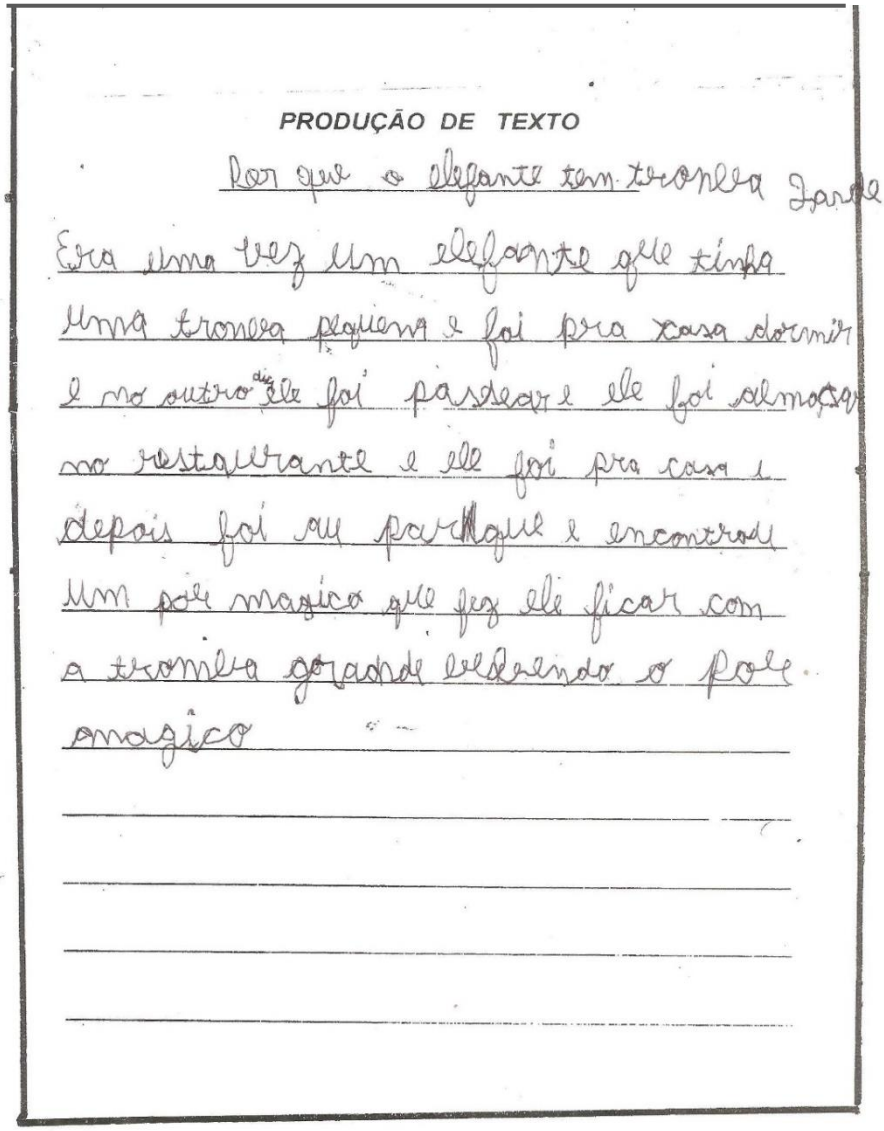


PRODUÇÃO DE TEXTO

Porque a formiga é pigmea

Era uma vez uma formiga que era
muito grande e ela fazia serin cadeiras
fatas que os homens não gostavão
e uma criança que encontrou o
pó magico e jogou o pó na formiga
e o menino foi pasia na mata e viu
a formiga grande ~~a formiga~~ deu um
susto do menino porque ~~ela~~ a formiga
ficou grande na frente do minino e o menino
foi para os índios e a formiga que
todo formigueiro ficou grande e os
indios tambem tinham o pó magico
e jogaram nas formigas e ~~ficou~~
e elas ficaram pigmeas

ANEXO 23 - Produção Tema Livre 004_D4_019

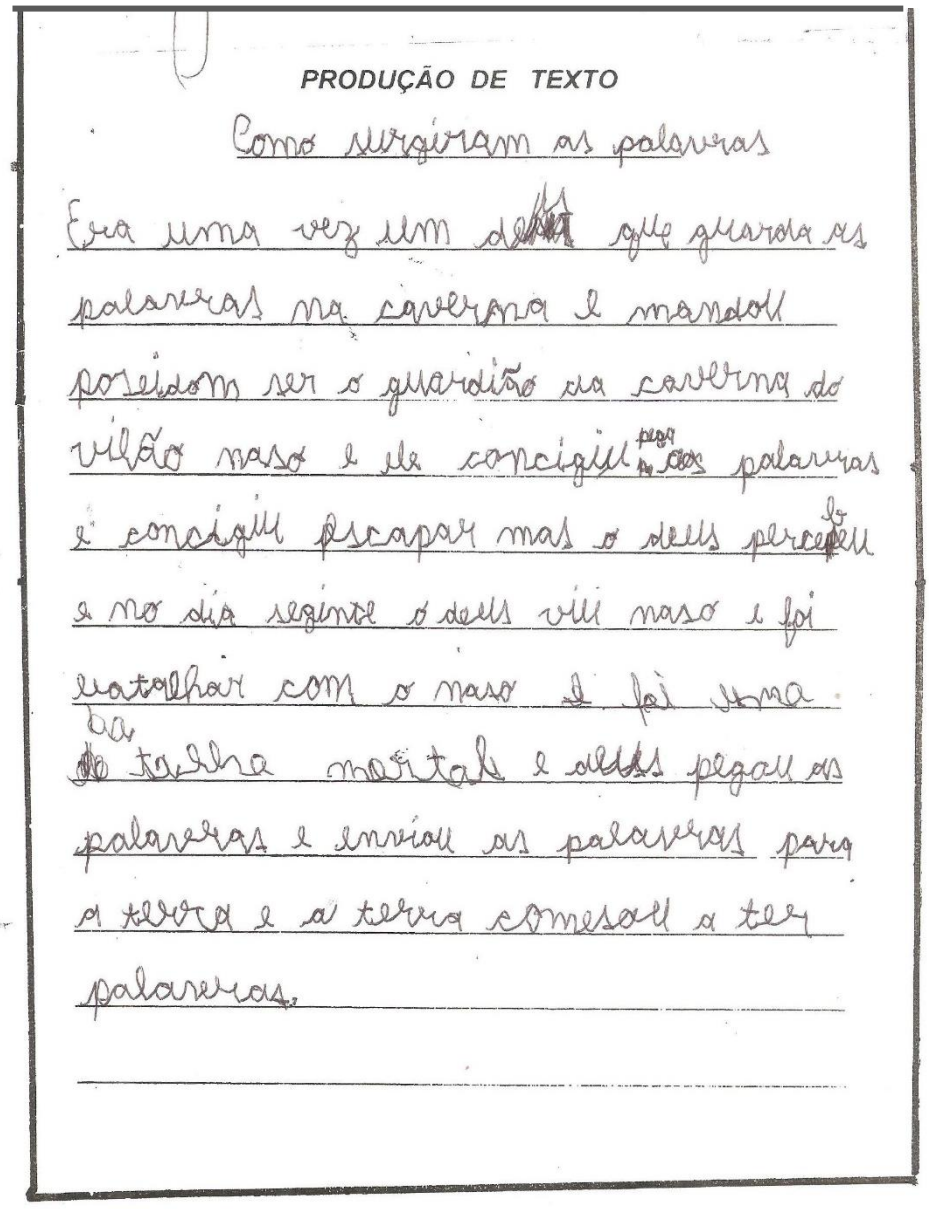


PRODUÇÃO DE TEXTO

Por que o elefante tem tromba Jarde

Era uma vez um elefante que tinha uma tromba pequena e foi pra casa dormir e no outro dia ele foi passear e ele foi almoçar no restaurante e ele foi pra casa e depois foi ao parque e encontrou um pó mágico que fez ele ficar com a tromba grande obedecendo o pó mágico

ANEXO 24 - Produção Tema Sugerido 005_D4_024

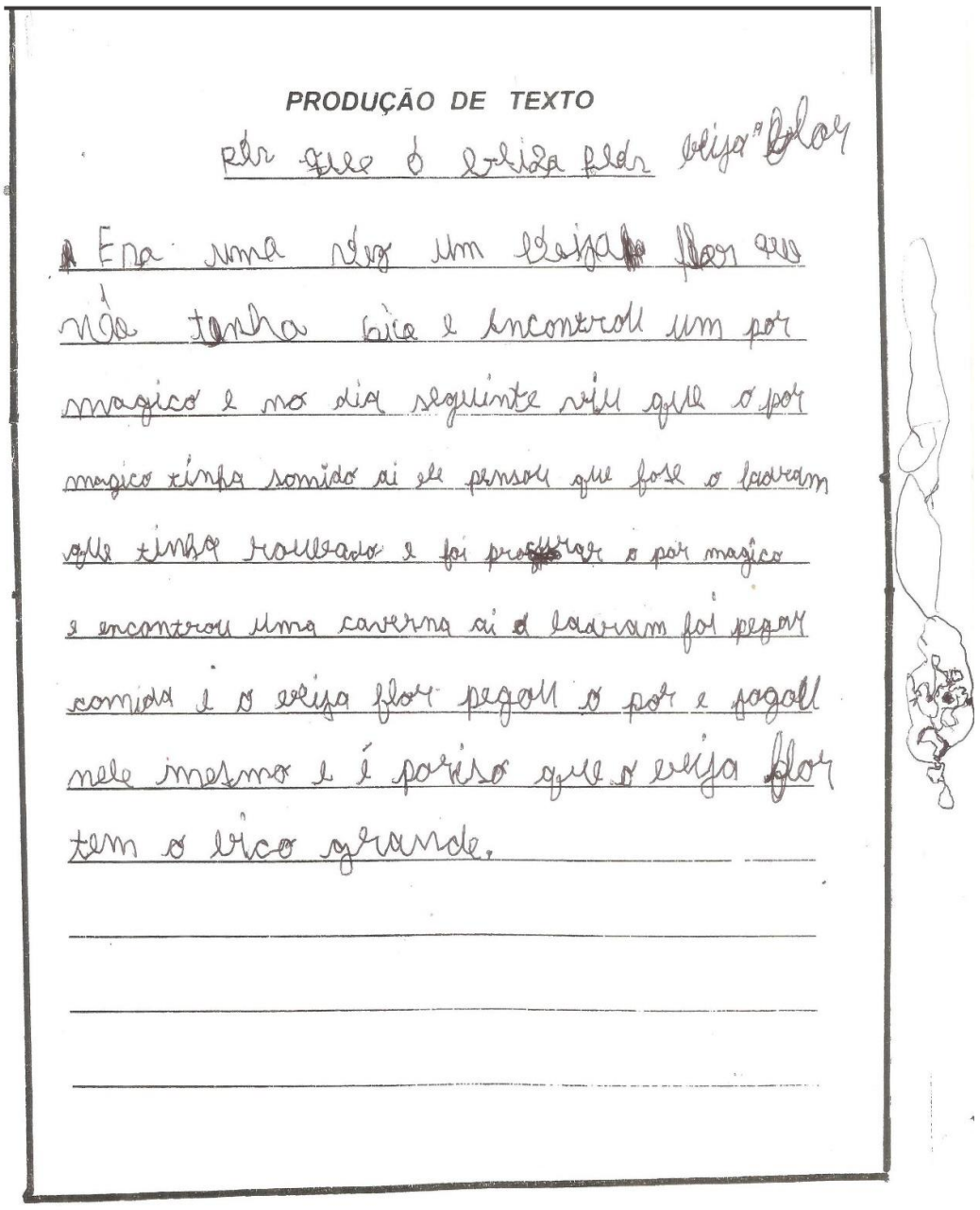


PRODUÇÃO DE TEXTO

Como surgiram as palavras

Era uma vez um deus que guardava as palavras na caverna e mandou Poseidon ser o guardião da caverna do vilão naso e ele conseguiu pegar as palavras e conseguiu escapar mas o deus percebeu e no dia seguinte o deus viu naso e foi batalhar com o naso e foi uma batalha mortal e deus pegou as palavras e enviou as palavras para a deusa e a terra comeu a tal palavras.

ANEXO 25 - Produção Tema Livre 006_D4_029



PRODUÇÃO DE TEXTO

Era uma vez um Beija flor que não tinha bico e Encontrou um por mágico e no dia seguinte viu que o por mágico tinha sumido ai ele pensou que foi o ladrão que tinha roubado e foi procurar o por mágico e encontrou uma caverna ai o ladrão foi pegar comida e o beija flor pegou o por e pegou nele mesmo e é por isso que o beija flor tem o bico grande.

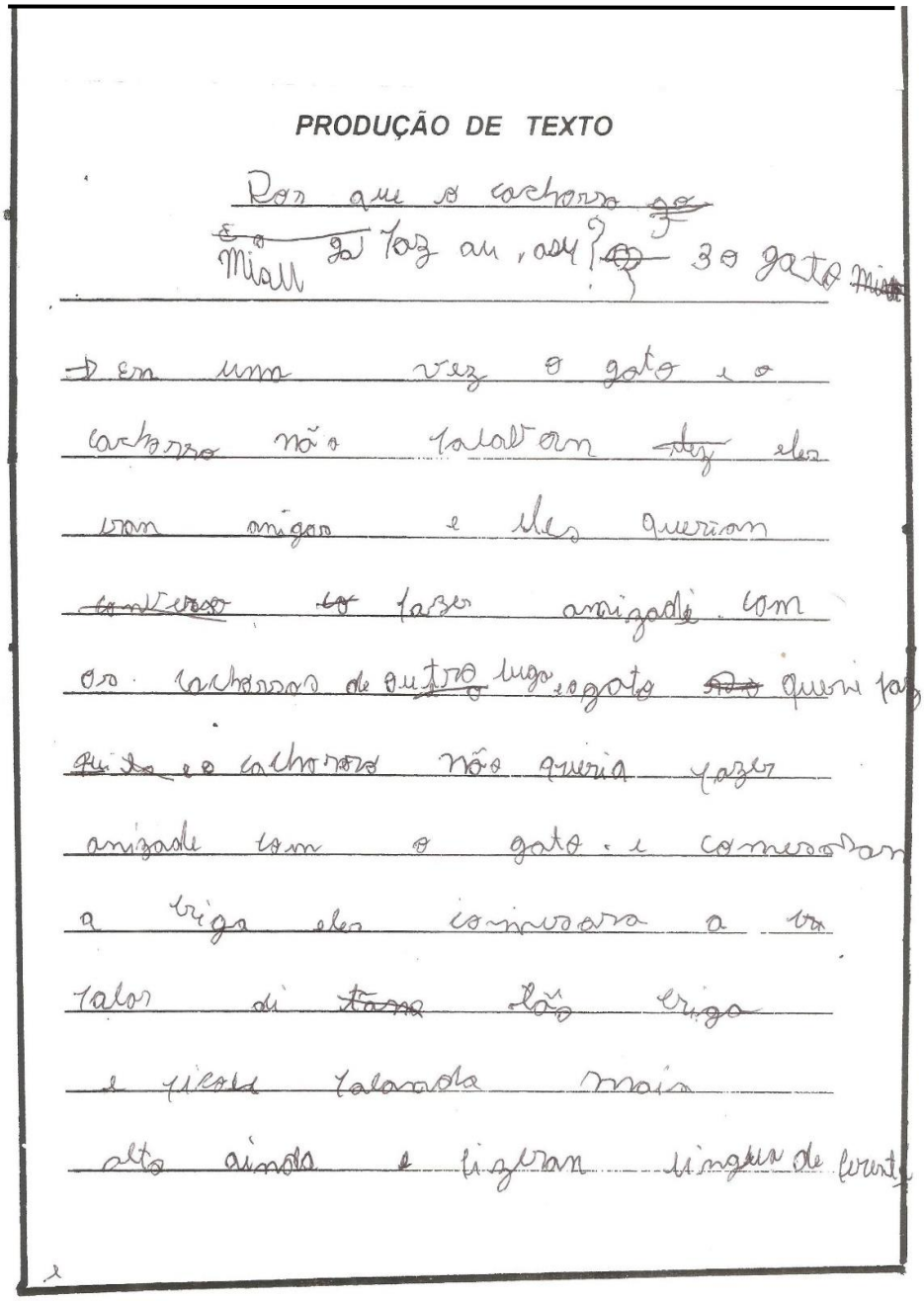
ANEXO 26 - Produção Tema Livre 001_D5_05

PRODUÇÃO DE TEXTO

Porque a girafa tem o pescoço longo

A Muito tempo atras uma girafa que
~~que~~ não tinha ~~pesçoço~~ pescoço ~~que derepe-~~
~~nte~~ derrepente apareceu uma ~~co~~ Cobra grande
que estava passando por perto da girafa a gira-
fa não gostou, a Cobra Cançou de ~~paso~~
~~s~~ passar toda vez la e foi falar
Com a girafa e viraram amigos
derrepente a Cobra tinha uma irmã
que não gostava dela e passou por perto
~~e disse que~~ a Cobra disse que ia ~~Com~~
Competir, a girafa não queria ~~o~~ Competir
mas a Cobra empediu e a girafa não
gostou disso e ficou Com raiva da Cobra
~~e~~ a girafa pisou na Cobra e ingoliu
~~e~~ a Cobra ficou em pé e o pescoço da
girafa foi Crecendo e porisso que a girafa tem
o pescoço longo.

ANEXO 27 - Produção Tema Sugerido 002_D5_10

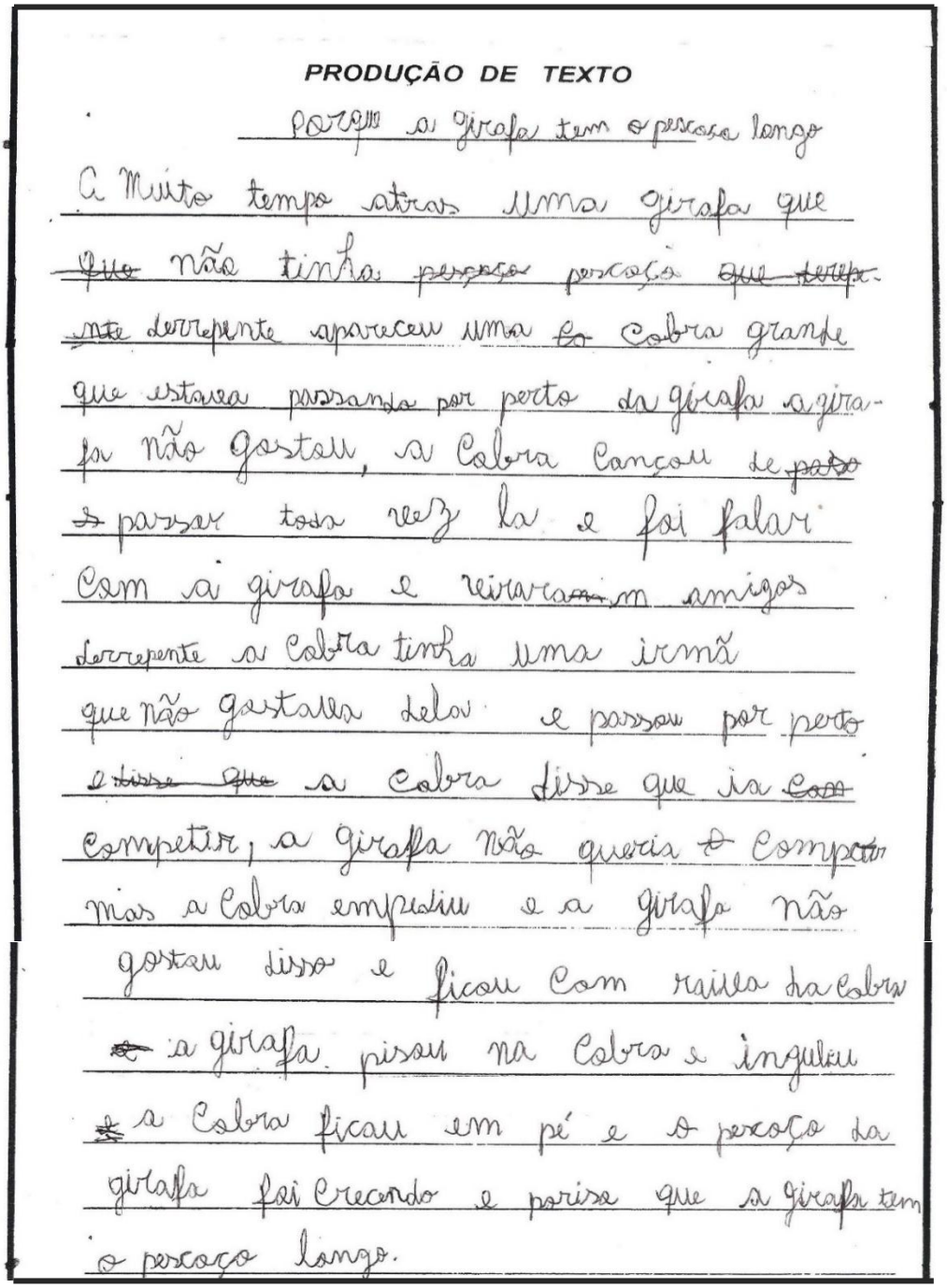


PRODUÇÃO DE TEXTO

Por que o cachorro ~~go~~ Miall foi faz au, au? ~~O~~ O gato Miall

Em uma vez o gato e o cachorro não falavam ~~dez~~ eles eram amigos e eles queriam ~~fazer~~ ~~fazer~~ fazer amizade com os cachorros de outro lugar, o gato ~~não~~ queria fazer amizade ~~amizade~~ ~~Miall~~ e o cachorro não queria fazer amizade com o gato. e começaram a briga eles começaram a falar ai ~~tava~~ lá briga e ficou falando mais alto ainda e ficaram ~~ficaram~~ brigando de verdade

ANEXO 28 - Produção Tema Livre 003_D5_015



PRODUÇÃO DE TEXTO

Porque a girafa tem o pescoço longo

A muito tempo atras uma girafa que não tinha ~~pescoço~~ pescoço ~~que tinha~~ ~~nte~~ derrepente aparecen uma ~~Co~~ Cobra grande que estavea passando por perto da girafa a girafa não gostou, a Cobra Cançou de ~~pass~~ passar toda vez la e foi falar Com a girafa e virarem amigos derrepente a Cobra tinha uma irmã que não gostava delas e passou por perto e ~~disse que~~ a Cobra disse que ia ~~Com~~ Competir, a girafa não queria ~~o~~ Competir mas a Cobra empurrou e a girafa não gostou disso e ficou Com raiva da Cobra ~~a~~ a girafa pisou na Cobra e ingulio ~~a~~ a Cobra ficou em pé e o pescoço da girafa foi Crecendo e porisso que a girafa tem o pescoço longo.

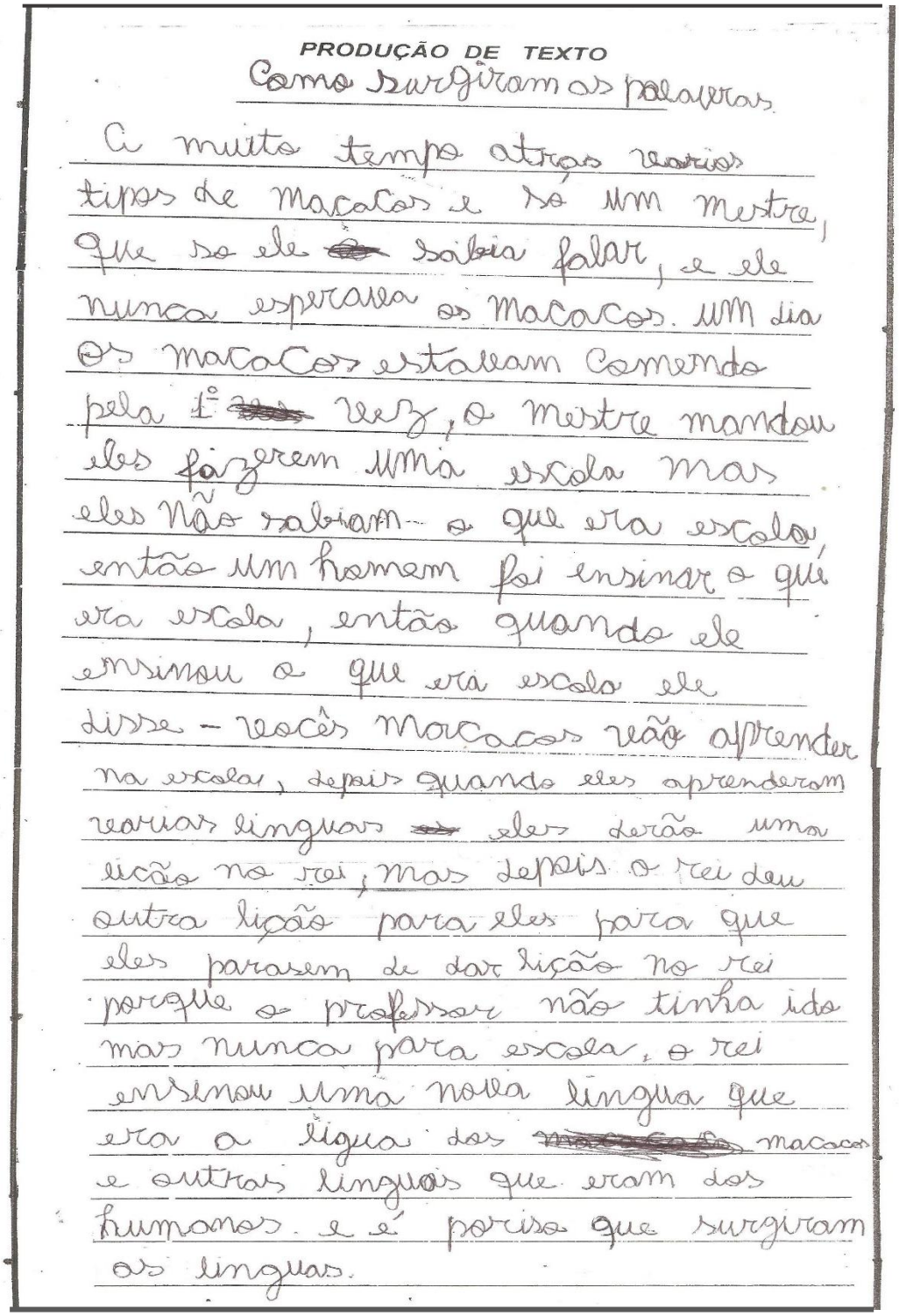
ANEXO 29 - Produção Tema Livre 004_D5_020

PRODUÇÃO DE TEXTO

Por que o esquilo gosta de nozes.

a muito ~~tempo~~ ~~at~~ tempo atras o esquilo não
tinha casa nem comida, um dia o esquilo
tinha perguntado para o leão onde ~~[illegible]~~ tinha
uma casa com comida, ele disse eu
não sei se você pode mora com
migo o esquelo perguntou a lebre ela
disse que era muito perigoso mora em
baixo, entono ele viu um àrvori bem bonita
~~e~~ é subiu para ver o que tinha ~~[illegible]~~
dentro, tinha varios nozes, água e cama.
ele disse para todos os animais e
voltou um animau esse animau
tentou matar ~~[illegible]~~ o esquilo ~~[illegible]~~ mais
o esquilo entrou em sua casa e disse
para os lebre que tinha ido como visita.
~~lebre você tinha [illegible] razão [illegible] a~~
. quando a lebre saiu ~~[illegible]~~ o esquilo si ~~[illegible]~~ sentiu sozinho
~~[illegible]~~ ficou com ele para ~~[illegible]~~
na nova casa ~~do~~ esquelo. o esquilo nunca
mais sais da sua casa e é porisso que
o esquilo come nozes. fim.

ANEXO 30 - Produção Tema Sugerido 005_D5_025



PRODUÇÃO DE TEXTO

Como surgiram as palavras

Ci muito tempo atrás varios tipos de macacos e so um mestre, que so ele ~~e~~ sabia falar, e ele nunca esperava os macacos. um dia os macacos estavam comendo pela 1ª ~~vez~~ vez, o mestre mandou eles fazerem uma escola mas eles não sabiam o que era escola, então um homem foi ensinar o qué era escola, então quando ele ensinou o que era escola ele disse - vocês macacos vão aprender na escola, depois quando eles aprenderam varias linguas ~~eles~~ eles terão uma lição no rei, mas depois o rei deu outra lição para eles para que eles parasem de dar lição no rei porque o professor não tinha ido mas nunca para escola, o rei ensinou uma nova lingua que era a ligua dos ~~macacos~~ macacos e outras linguas que eram dos humanos. e é porisso que surgiram as linguas.

ANEXO 31 - Produção Tema Livre 006_D5_030

PRODUÇÃO DE TEXTO

Porque o sol brilha.

A Muito tempo atras o sol não brilhava, um dia o dragão pediu ao vucão por emprestar um poco do seu fogo, o dragão disse por o vucão, o vucão disse que ele ia ficar sem fogo, o dragão Maugrindo foi e pegou o fogo do vucão Mais quando estava saindo pisou no rabinho do vucão disse bem alto - esse e o seu castigo, o dragão gritou e saio fogo direto - muito dele ate o sol o sol ficou feliz por esta brilhando de novo, e o dragão perguntou ao sol se devolveria o fogo, ele poderia devolver o sol disse não por que ele não podia sai de lá e o dragão ficou sem voo de tão auto que tava voando, quando o dragão teve sua voz de volta ele voltou com o seu fogo, o dragão pedio a o sol obrigado por te min dado um pouco do fogo, Mais o sol esqueseu de colocar fogo no vucão e foi jogar fogo no vucão e o vucão ficou feliz. e é por isso que o sol brilha.

Fim.

ANEXO 32 - Resumo das produções de cada díade.

D1

Produções	PE	DiL	DeL
1	55	70,91	45,45
2	77	55,84	41,56
3	55	65,45	52,73
4	122	51,64	47,54
5	73	64,38	46,58
6	85	26,62	54,12

D2

Produções	PE	DiL	DeL
1	128	60,94	46,88
2	72	55,56	43,06
3	57	71,93	49,12
4	75	61,33	52,00
5	103	66,43	60,19
6	93	34,74	39,78

D3

Produções	PE	DiL	DeL
1	93	65,59	46,24
2	78	48,72	42,31
3	84	66,67	47,62
4	84	57,14	48,18
5	71	32,3	47,62
6	91	28,74	43,96

D4

Produções	PE	DiL	DeL
1	78	37,43	48,72
2	80	20,94	47,5
3	87	39,47	55,17
4	62	58,06	58,06
5	76	32,08	55,26
6	86	22,87	45,53

D5

Produções	PE	DiL	DeL
1	123	27,91	49,59
2	84	48,81	45,24
3	77	64,94	50,65
4	141	58,16	46,81
5	139	54,06	45,32
6	184	32,47	49,46