

ESCOLA DE HUMANIDADES
PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS
MESTRADO EM LETRAS

JOANA PAIM DA LUZ

**ANÁLISE DE GRAFOS APLICADA A PRODUÇÕES TEXTUAIS DE ALUNOS DO
ENSINO FUNDAMENTAL E SEU POTENCIAL PREDITIVO DA DISLEXIA DO
DESENVOLVIMENTO**

Porto Alegre
2018

PÓS-GRADUAÇÃO - *STRICTO SENSU*



Pontifícia Universidade Católica
do Rio Grande do Sul

JOANA PAIM DA LUZ

**ANÁLISE DE GRAFOS APLICADA A PRODUÇÕES TEXTUAIS DE ALUNOS DO
ENSINO FUNDAMENTAL E SEU POTENCIAL PREDITIVO DA DISLEXIA DO
DESENVOLVIMENTO**

Dissertação apresentada como requisito para a
obtenção do grau de Mestre pelo Programa de
Pós-Graduação da Faculdade de Letras da
Pontifícia Universidade Católica do Rio Grande
do Sul.

Orientador: Dr. Augusto Buchweitz
Co-orientador: Dr. Felipe Meneguzzi

Porto Alegre
2018

Ficha Catalográfica

L979a Luz, Joana Paim da

Análise de grafos aplicada a produções textuais de alunos do Ensino Fundamental e seu potencial preditivo da dislexia do desenvolvimento / Joana Paim da Luz . – 2018.

166 f.

Dissertação (Mestrado) – Programa de Pós-Graduação em Letras, PUCRS.

Orientador: Prof. Dr. Augusto Buchweitz.

Co-orientador: Prof. Dr. Felipe Rech Meneguzzi.

1. Dislexia. 2. Teoria dos Grafos. 3. Habilidades de escrita. 4. Aprendizado de Máquina. I. Buchweitz, Augusto. II. Meneguzzi, Felipe Rech. III. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da PUCRS
com os dados fornecidos pelo(a) autor(a).

Bibliotecário responsável: Marcelo Votto Texeira CRB-10/1974

JOANA PAIM DA LUZ

**ANÁLISE DE GRAFOS APLICADA A PRODUÇÕES TEXTUAIS DE ALUNOS DO
ENSINO FUNDAMENTAL E SEU POTENCIAL PREDITIVO DA DISLEXIA DO
DESENVOLVIMENTO**

Aprovada em: ____ de _____ de _____

BANCA EXAMINADORA:

Dr. Natalia Bezerra Mota

Prof. Dr. Janaina Weissheimer

Prof. Dr. Felipe Meneguzzi

Prof. Dr. Augusto Buchweitz

Porto Alegre
2018

Aos meus pais, que me incentivaram a ler,
viabilizando assim muitos dos meus
aprendizados.

AGRADECIMENTOS

Aos meus pais, José Ricardo e Soraya, que, com amor incondicional, sempre me ofereceram o privilégio da certeza de que eu não estou só;

Aos meus irmãos, Arthur e Julia, que, desde que chegaram, têm me ajudado a ser uma pessoa melhor a cada conquista, derrota, experiência e olhar diferente acerca das coisas do mundo compartilhados;

Ao meu orientador, Prof. Augusto Buchweitz, pela receptividade, paciência e flexibilidade que me oportunizaram “dissertar” sobre interesses antigos e pouco ortodoxos no campo das Letras, enfrentar algumas inseguranças e, finalmente, (re)descobrir trajetórias possíveis rumo a minha Ilha Desconhecida, ao lado de queridos colegas de pesquisa;

Ao meu co-orientador, Prof. Felipe Meneguzzi, por me acolher, por me apresentar diálogos possíveis e pertinentes entre duas áreas que em muito me agradam e, não menos importante, por me ensinar, sem a pretensão de o fazer, que ciência séria também se faz com bom humor, armas de brinquedo e balas de borracha;

À Natália Mota, pelas trocas de email atenciosas e generosas, pelas considerações preciosas acerca do meu trabalho e, sobretudo, pelo estudo que me fascinou e me motivou a escrever esta dissertação;

À Prof. Janaína Weissheimer, que acompanhou a evolução do meu trabalho com uma delicada e acurada criticidade, contribuindo, assim, para o aprimoramento da discussão e da exploração que me propus a fazer;

À Mari, ao Bernardo e à Aline, colegas de pesquisa com quem esclareci incontáveis dúvidas e compartilhei algumas descobertas; em meio aos fracassos e êxitos diários da vida acadêmica, aprendi com eles a ser uma melhor aluna, orientanda, mestranda e aspirante a cientista;

À Ana Bassoa e à Rudi - colegas do ACERTA -, e ao Prof. Alexandre Franco, à Nathi, ao Luiz e ao Zé – parceiros do LABIMA -, que me socorreram em vários momentos com gentileza e generosidade, tornando menos nebulosas diferentes questões relacionadas à Neurociência e mais prazerosos meus dias de trabalho;

Ao Matheus e à Laura, orientandos do Meneguzzi, que, entre reuniões e conversas pelo Whatsapp, ajudaram a colocar em prática a interface central neste trabalho entre Psicolinguística e Inteligência Artificial – e, sem querer, mostraram que as pessoas da informática estão entre as mais legais da Academia;

A todos os colegas e professores que tornaram minha rotina acadêmica pouco enfadonha e muito divertida, em especial à Vergília (meu anjinho da guarda nas aulas de Fonologia – obrigada, Vergília!), à Ellen, ao Ricardo e à Bárbara, parceiros de cafés, companheiros de viagens e, agora, meus amigos;

Aos amigos e familiares que tanto amo e que fizeram mais felizes estes dois anos pelo simples fato de “estarem lá”, ainda que distantes da PUCRS;

Finalmente, às crianças e adolescentes participantes do Projeto ACERTA, que me permitiram melhor compreender a dislexia; graças a elas hoje compartilho meu aprendizado com outras pessoas interessadas em serem agentes de transformação da educação brasileira e da realidade de todos que convivem com os entraves cognitivos e sociais deste transtorno de aprendizagem chamado dislexia.

Change is the essential process of all existence.

Spock

*Just as the constant increase of entropy is the basic law
of the universe, so it is the basic law of life to struggle
against entropy.*

Vaclav Havel

RESUMO

A dislexia é um transtorno de aprendizagem de natureza neurobiológica, caracterizado por uma inesperada dificuldade em decodificar e compreender textos escritos, decorrente de um aprendizado insatisfatório do princípio alfabético e de uma associação grafo-fonêmica deficitária. A exploração científica da expressão escrita de disléxicos tem se dedicado, em grande medida, a investigar aspectos linguísticos associados à ortografia e às classes de palavras empregadas por sujeitos com dislexia. Estudos de natureza quantitativa focados em mensurar a estrutura dos textos deste grupo clínico são desconhecidos. O principal objetivo do presente trabalho foi identificar padrões de conectividade textual em bons leitores, maus leitores e disléxicos, a partir da análise de medidas de grafo extraídas de suas produções textuais e de técnicas de Aprendizado de Máquina. Essencialmente, buscou-se investigar (a) se o tipo de transcrição dos textos - corrigidos ou originais - e a normalização das medidas de grafo pelo número de palavras interferem significativamente na classificação das crianças entre grupos de fluência e escolaridade; (b) se há diferenças significativas entre os atributos do grafo dos bons leitores, maus leitores e disléxicos; (c) se as medidas de cada ano de coleta de dados convergem para valores semelhantes e se são significativamente diferentes entre os anos; (d) se os atributos de grafo obtidos com a utilização do *Speech Graphs*, quando associados às técnicas de Aprendizado de Máquina, podem prever níveis de fluência de leitura e, especificamente, a dislexia de desenvolvimento. As hipóteses, para as questões listadas, eram todas afirmativas. Para verificá-las, os textos produzidos por 181 crianças e adolescentes do Projeto ACERTA foram transcritos e divididos em dois grupos experimentais: Ambulatório (N = 52, todos disléxicos) e Escolas (N = 129, subdividido em leitores bons, médios e maus leitores). Esses textos transcritos serviram de entrada para o software *Speech Graphs*, que extraiu atributos de grafo representativos da estrutura de cada texto. As análises estatísticas descritivas e inferenciais revelaram (a) a prevalência de significância dos atributos de grafo extraídos de transcrições originais não normalizadas (63,07% de significância entre análises); (b) padrões de conectividade textual por grupo de fluência de leitura e (c) padrões de conectividade textual por cada ano de coleta de dados com base em diferenças significativas encontradas em cinco atributos de grafo: nós, arestas, maior componente conectado, densidade e caminho menor médio; (d) 2016 como o melhor ano para classificar as crianças em seus grupos de fluência de leitura fazendo uso de classificadores SVM, considerando que eles alcançaram a maior acurácia (85%), revocação (83%), precisão (85%) e pontuação F1 (83%) ao classificar textos de bons leitores e de crianças disléxicas, com base em medidas de grafo de 2016. Estas descobertas fornecem evidências que indicam a possibilidade de se explorar e aprimorar um novo aparato teórico-metodológico para avaliar a fluência de leitura, a partir de textos escritos e com base na Teoria dos Grafos.

Palavras-chave: Dislexia; Teoria dos Grafos; Habilidades de escrita; Aprendizado de Máquina.

ABSTRACT

Dyslexia is a learning disorder of neurobiological origin, which is characterized by an unexpected difficulty when decoding written texts, due to an unsatisfactory learning of the alphabetical principle and a deficient graphological-phonemic association. The ways dyslexia affects writing skills have been scientifically explored so as to investigate linguistic aspects associated with spelling and classes of words used by subjects who suffer from dyslexia. Quantitative studies focused on measuring the structure of their texts are unknown. The main objective of this study was to identify patterns of textual connectivity in good readers, bad readers and dyslexic children, based on the analysis of graph measures extracted from their texts and Machine Learning techniques. Essentially, it sought to investigate (a) whether the type of transcription of the texts - corrected or original - and the normalization of the graph attributes by the number of words of each text interfere significantly in sorting the children in their fluency and schooling groups; (b) whether there is significant differences among good readers', bad readers' and the dyslexic participants' graph attributes; (c) whether measures of each year of data collection converge to similar values and if they are significantly different among the years; (d) whether the graph attributes obtained with Speech Graphs, when associated with Machine Learning techniques, can predict reading fluency levels and, specifically, developmental dyslexia. The hypotheses for the questions listed were all affirmative. To verify them, texts produced by 181 children and adolescents from the ACERTA Project were transcribed and divided in two experimental groups: Ambulatory (N = 52, all dyslexic) and Schools (N = 129, subdivided into good, medium and bad readers). These transcribed texts served as input for Speech Graphs software, which extracted graph attributes representative of the structure of each text. Descriptive and inferential statistical analyzes revealed (a) the prevalence of significance among graphs attributes extracted from non-normalized original transcripts (63.07% significance between analyzes); (b) patterns of textual connectivity by each reading fluency group and (c) patterns of textual connectivity by year of data collection based on significant differences found in five graphs attributes - nodes, edges, largest connected component, density and average smallest path; (d) 2016 as the best year to sort the children in their reading fluency groups by making use of SVM classifiers, considering they reached the highest accuracy (85%), recall (83%), precision (85%) and F1 score (83%) when sorting good readers and dyslexic children texts, based on 2016 graph measures. These findings provide evidences that indicate the possibility to explore and improve a new methodological frame to assess reading fluency through written texts and based on Graph Theory.

Key-words: Dyslexia; Theory of Graphs; Writing skills; Machine Learning

LISTA DE ILUSTRAÇÕES

Figura 1: Distribuição das línguas europeias quanto à profundidade ortográfica	28
Figura 2: Afasias de Broca e Wernicke	30
Figura 3: Modelo conexionista da linguagem falada	31
Figura 4: A compreensão leitora no cérebro	34
Figura 5: Modelo de Representação Única e Modelo de Representação Dual	39
Figura 6: Fatores atuantes sobre o desempenho acadêmico infantil	41
Figura 7: Atividade cerebral de disléxicos e de leitores normais, durante a leitura	46
Figura 8: Relações interdisciplinares	53
Figura 9: Grafo	59
Figura 10: Grafo com ciclos	60
Figura 11: Grafo direcionado	60
Figura 12: Subgrafo do texto 143-2014-O-T-AC	61
Figura 13: Grafo desconectado do texto 58-2014-O-IE-LB	62
Figura 14: Atributos visualizados no grafo	66
Figura 15: LCC	67
Figura 16: LSC	67
Figura 17: Perfil dos pacientes do Ambulatório	71
Figura 18: Modelo da tirinha da tarefa Produção de Texto	77
Figura 19: Texto original escrito à mão, pelo participante 62-2014-IE-MD	78
Figura 20: Modelo de transcrição original do texto do participante 62-2014-IE-MD	79
Figura 21: Modelo de transcrição corrigida do texto do participante 62-2014-IE-MD	79
Figura 22: Grafo da versão original do texto do participante 62-2014-IE-MD	79
Figura 23: Grafo da versão corrigida do texto do participante 62-2014-IE-MD	80
Figura 24: Transcrição corrigida de 87-A-C-2014-LCM	88
Figura 25: Transcrição original de 87-A-C-2014-LCM	88
Figura 26: Grafo da transcrição corrigida do texto do participante 87-A-C-2014-LCM	88
Figura 27: Grafo da transcrição original do texto do participante 87-A-C-2014-LCM	89
Figura 28: Soma de L1 Estrato Socioeconômico	90
Figura 29: Soma de L2 por Estrato Socioeconômico	91
Figura 30: Atributos gerais de grafo (N e E) e contagem de palavras (WC) (Longitudinal)	93
Figura 31: Atributos de recorrência (PE, L1, L2 e L3) (Longitudinal)	95
Figura 32: Atributos de conectividade (ATD, LCC e LSC) (Longitudinal)	96
Figura 33: Atributos globais (Dens. Diam. e ASP) (Longitudinal)	98
Figura 34: Relação entre as médias de Dens. e LCC. (Longitudinal)	99
Figura 35: Atributos gerais (N e E) e número de palavras (WC) (Transversal Geral)	101
Figura 36: Atributos de recorrência (PE, L1, L2 e L3) (Transversal Geral)	102
Figura 37: Atributos de conectividade (ATD, LCC e LSC) (Transversal Geral)	104
Figura 38: Atributos globais (Dens., Diam. e ASP) (Transversal Geral)	105
Figura 39: Relação entre as médias de Dens. e N (Transversal Geral)	106
Figura 40: Relação entre as médias de N e Dens. (Transversal Geral)	107
Figura 41: Número de palavras (WC) e nós (E) (Transversal por ano e fluência de leitura)	109
Figura 42: Arestas (E) e arestas paralelas (PE) (Transversal por ano e fluência de leitura)	111
Figura 43: LCC e LSC (Transversal por ano de coleta e fluência de leitura)	112
Figura 44: ATD e ASP (Transversal por ano de coleta e fluência de leitura)	113
Figura 45: Dens. e Diam. (Transversal por ano de coleta e fluência de leitura)	115

Figura 46: Relação entre médias de WC e Dens.....	116
Figura 47: Relações entre médias de LCC e Dens.....	116
Figura 48: Relações entre as médias de E e Dens.....	117
Figura 49: Teia de médias de atributos de 2014.....	118
Figura 50: Teia de médias de atributos de 2015.....	120
Figura 51: Teia das médias de atributos de 2016.....	121
Figura 52: Gráfico de predição de 2014 - Originais.....	127
Figura 53: Gráfico de predição de 2014 - Corrigidas.....	128
Figura 54: Gráfico de predição de 2015 - Originais.....	130
Figura 55: Gráfico de predição de 2015 - Corrigidas.....	131
Figura 56: Gráfico de setores de 2016 - Originais.....	133
Figura 57: Gráfico de predição de 2016 - Corrigidas.....	134

LISTA DE TABELAS

Tabela 1: Exemplos de paradigmas de Aprendizado de Máquina	54
Tabela 2: Atributos de grafo (Adaptado de Mota et al. (2016)).....	63
Tabela 3: Características da população geral de estudo	71
Tabela 4: Características e critérios de seleção da amostra.....	72
Tabela 5: Distribuição dos participantes do ambulatório quanto à classe socioeconômica	73
Tabela 6: Tipo de escola de origem dos pacientes do Ambulatório.....	74
Tabela 7: Protocolo de diagnóstico do Ambulatório.....	75
Tabela 8: Instrumentos utilizados na avaliação fonoaudiológica do Ambulatório	76
Tabela 9: Análise longitudinal (Escolas)	83
Tabela 10: Análise transversal geral (Escolas e Ambulatório)	83
Tabela 11: Análise transversal geral por ano de coleta e fluência leitora (2014).....	83
Tabela 12: Transversal geral por ano de coleta e fluência leitora (2015)	83
Tabela 13: Transversal geral por ano de coleta e fluência leitora (2016)	84
Tabela 14: Públicas x Privadas (Ambulatório).....	85
Tabela 15: Estratos econômicos – B x C x D (Ambulatório).....	87
Tabela 16: Resultados dos classificadores SVM para originais.....	124
Tabela 17: Resultados dos classificadores SVM para corrigidas.....	125
Tabela 18: Matrizes de confusão de 2014.....	129
Tabela 19: Matrizes de confusão de 2015.....	132
Tabela 20: Matrizes de confusão de 2016.....	134

LISTA DE ABREVIACES

ABEP	Associao Brasileira de Estudos Populacionais
ABIPEME	Associao Brasileira dos Institutos de Pesquisa de Mercado.
ACERTA	Avaliao de Crianas Em Risco de Transtorno de Aprendizagem
AFVP	rea da Forma Visual das Palavras
AM	Aprendizado de Mquina
ASP	Menor caminho mdio (<i>Average Shortest Path</i>)
ATD	Grau Mdio Total (<i>Average Total Degree</i>)
CC	Cincia da Computao
Dens.	Densidade
Diam.	Dimetro
E	Arestas (<i>Edges</i>)
G	Grafo
IA	Inteligncia Artificial
InsCer	Instituto do Crebro do Rio Grande do Sul
L1	Ciclo de um n (<i>1 Node/ Loop</i>)
L2	Ciclo de dois ns (<i>2 Nodes/ Loop</i>)
L3	Ciclo de trs ns (<i>3 Nodes/ Loop</i>)
LC	Lingustica Computacional
LCC	Maior Componente Conectado (<i>Largest Connected Component</i>)
LING	Lingustica
LSC	Maior Componente Fortemente Conectado (<i>Largest Strongly Connected Component</i>)
N	Ns
PISA	Programa Internacional de Avaliao de Estudantes
PLN	Processamento da Linguagem Natural
PUCRS	Pontifcia Universidade Catlica do Rio Grande do Sul
QI	Quociente de Inteligncia
RMf	Ressonncia Magntica Funcional
SVM	Mquinas de Vetores de Suporte (<i>Support Vector Machines</i>)
TEA	Transtorno Especfico da Aprendizagem
TG	Teoria dos Grafos
UFRN	Universidade Federal do Rio Grande do Norte
UFSC	Universidade Federal de Santa Catarina
WC	Contagem de palavras (<i>Word Count</i>)

SUMÁRIO

INTRODUÇÃO	17
1 FUNDAMENTAÇÃO TEÓRICA	25
1.1 LINGUAGEM: DA ORALIDADE AO ARTEFATO CULTURAL DA ESCRITA	25
1.1.1 Ortografias: transparentes e opacas, rasas e profundas	27
1.2 O CÉREBRO HUMANO E A LINGUAGEM ORAL E ESCRITA.....	29
1.2.1 O cérebro que compreende e produz a linguagem oral	29
1.2.2 O cérebro que compreende e produz a linguagem escrita	32
1.2.3 Modelos de Compreensão Leitora	36
1.2.4 Convergências e divergências entre leitura e escrita	38
1.3 DISLEXIA: UM TRANSTORNO DE APRENDIZAGEM.....	40
1.3.1 O cérebro com dislexia	44
1.3.2 A alfabetização e a produção escrita de disléxicos	46
1.4 CIÊNCIA DA COMPUTAÇÃO E LINGUÍSTICA: UMA INTERFACE PRODUTIVA E PERTINENTE.....	51
1.4.1 Aprendizado de Máquina aplicado a dados linguísticos: um encontro em evidência	53
1.4.2 A Teoria dos Grafos como ferramenta de análise textual	56
1.4.3 Definições e atributos de grafo	59
2 MÉTODOS	68
2.1 PARTICIPANTES.....	70
2.2 RECRUTAMENTO.....	72
2.2.1 Procedimentos anteriores ao recrutamento	74
2.3 MATERIAIS E PROCEDIMENTOS.....	76
2.3.1 Tarefa de produção textual - (SALLES, 2005)	76
2.3.2 Avaliação de compreensão leitora de textos expositivos (SARAIVA, R., MOOJEN, S., MUNARSKI, 2007)	80

3 ANÁLISE E DISCUSSÃO DE DADOS	82
3.1 ANÁLISES ESTATÍSTICAS DOS DADOS	82
3.1.1 Análises prévias - Ambulatório	85
3.1.2 Análises estatísticas intra e intergrupos – Escolas e Ambulatório.....	92
3.1.3 Análise Longitudinal - Escolas	92
3.1.4 Análise transversal geral	99
3.1.5 Análise transversal geral entre categorias de fluência de leitura, por ano de coleta	107
3.1.6 Teias de médias de atributos	117
3.2 RESULTADOS DO CLASSIFICADOR SVM.....	123
CONSIDERAÇÕES FINAIS	136
REFERÊNCIAS	142
ANEXOS	153
APÊNDICES	164

INTRODUÇÃO

O presente trabalho reporta o estudo realizado acerca dos padrões de conectividade textual¹ observados em produções textuais escritas por pacientes com diagnóstico de dislexia do desenvolvimento do Ambulatório de Aprendizagem do projeto ACERTA (Avaliação de Crianças Em Risco de Transtorno de Aprendizagem), crianças com dificuldade de leitura (maus leitores) e leitores típicos (bons leitores), todas participantes do projeto. A compreensão da pesquisa empreendida é favorecida por uma breve contextualização, apresentada a seguir.

Atualmente, rendimentos acadêmicos insatisfatórios são recorrentes já no início do ensino fundamental, em diferentes países. No Brasil, especificamente, estima-se que 59% das crianças cursando o quarto ano do ensino fundamental não apresentem um grau de alfabetização adequado (CASELLA et al., 2011). Avaliações oficiais, como PISA² (OECD, 2012) indicam que a educação brasileira, no que diz respeito ao ensino da leitura e da escrita, merece especial atenção, uma vez que não parece estar garantindo o domínio satisfatório e pleno destas habilidades da língua aos brasileiros (GUARESI; ALMEIDA, 2015). As causas deste baixo rendimento na alfabetização podem estar associadas a dificuldades de aprendizagem oriundas de quadros clínicos – caso de síndromes genéticas, deficiência intelectual, baixa acuidade auditiva ou visual, transtornos mentais, entre outros -, a fatores socioambientais – tais como pobreza, fome, ausência de moradia, inadequações do sistema de ensino e baixo estímulo à educação no ambiente familiar – ou a algum tipo de Transtorno Específico da Aprendizagem (TEA), como é o caso da dislexia, caracterizada pelo prejuízo da leitura e da escrita (HOEFT et al., 2011; MEYLER et al., 2007; SHAYWITZ et al., 1998).

¹ O termo “conectividade textual”, sempre que empregado neste trabalho, fará referência à “coesão textual” enquanto unidade textual. Defende-se que a adjacência entre os itens lexicais de um texto escrito, aqui representada e mensurada a partir da Teoria dos Grafos, consiste em uma alternativa de se obter uma medida quantitativa da unidade de um texto, ou seja, da associação entre os elementos que compõem a sua totalidade e determinam a sua legibilidade. Logo, a perspectiva de “coesão textual” subjacente a este estudo não está relacionada aos pressupostos da Linguística Textual, tampouco a qualquer perspectiva da coesão como propriedade qualitativa do texto escrito; trata-se de uma concepção de unidade textual referente à conexão ou à ligação entre as palavras ou itens lexicais de um texto, passível de ser medida quantitativamente.

² O PISA (Programa Internacional de Avaliação de Estudantes) é coordenado pela OCDE (Organização para a Cooperação e Desenvolvimento Económico) e consiste em uma iniciativa internacional de avaliação comparada, aplicada a estudantes próximos aos 15 anos, por ser a faixa etária em que se pressupõe o término da escolaridade básica obrigatória na maioria dos países (GUARESI; ALMEIDA, 2015).

De acordo com os critérios propostos pelo DSM-IV (APA, 2013), a dislexia compromete a performance de leitura do indivíduo, tornando-a substancialmente inferior à esperada para a sua idade cronológica, inteligência e escolaridade. Buchweitz (2016) define o transtorno como uma dificuldade de aprender a ler imprevista, uma vez que ocorre em crianças que apresentam condições normais de aprendizado, sem patologias ou déficits cognitivos que possam comprometer seu aproveitamento escolar. A incidência da dislexia, como sustenta Dehaene (2012), é controversa visto que se trata de um transtorno de aprendizagem associado a diversas variáveis, tais como a profundidade ortográfica e a própria concepção do que pode ser considerado um comportamento atípico em termos de compreensão leitora (GUARESI; ALMEIDA, 2015). No entanto, segundo a literatura, a dislexia do desenvolvimento acomete 1 em cada 5 crianças só nos Estados Unidos, o que traduz em cerca de 10 milhões de sujeitos disléxicos (SHAYWITZ, 2008, p. 16).

Sabe-se que a Área da Forma Visual das Palavras (AFVP) consiste em uma região do cérebro especializada na decodificação das palavras escritas, com papel crucial no circuito que executa esse processamento e viabiliza a leitura. Em crianças disléxicas, este circuito não se forma da maneira como ocorre com as demais crianças, que se mostram confortáveis com o aprendizado e não apresentam dificuldades de leitura expressivas (BUCHWEITZ, 2016). A dislexia, portanto, é a manifestação de uma atividade funcional cerebral anômala; regiões recrutadas normalmente para o processamento da escrita são subativadas em disléxicos quando confrontados com textos escritos, como a região temporal posterior esquerda (SHAYWITZ, 1998); igualmente comum é a superativação do córtex frontal inferior esquerdo - região da Área de Broca – durante a leitura (DEHAENE, 2012; GEORGIEWA, 2002), como um indício de compensação do cérebro, necessário diante da fraca atividade de regiões posteriores de decodificação.

Alunos disléxicos decodificam o texto escrito com vagar, de modo pouco fluente, se comparados aos demais colegas de classe. Há fatores preditivos da dislexia, relacionados a desempenhos observáveis nos anos pré-escolares, como o conhecimento do som das letras, a consciência fonológica, a nomeação automática rápida e as competências executivas. O diagnóstico do transtorno e sua adequada remediação implica em uma distinção entre dificuldades de aprendizagem e TEAs. Essa diferença têm motivado pesquisas interdisciplinares, comuns aos campos da Linguística, Psicologia, Neurociências e Ciências da Computação (MORKEN et al., 2014; OLSON, 2006; RICHARDSON; LYYTINEN, 2014).

O ACERTA corresponde a uma dessas iniciativas de estudo. Trata-se de um projeto de caráter multicêntrico – executado por pesquisadores da Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), da Universidade Federal de Santa Catarina (UFSC) e da Universidade Federal do Rio Grande do Norte (UFRN) -, longitudinal – realizado ao longo de quatro anos, com início em 2013 e conclusão em 2016 – e transversal – com coletas de dados realizadas em circunstâncias escolares e ambulatoriais (no que diz respeito ao trabalho realizado na PUCRS).

Com o propósito de ampliar e aprofundar o conhecimento científico sobre a aprendizagem em crianças falantes do português brasileiro, em fase de alfabetização, o ACERTA foi desenhado a partir de quatro grandes objetivos: identificar fatores preditores de dificuldades e transtornos de aprendizagem, compreender os mecanismos neurais relacionados à alfabetização e os obstáculos que a eles se impõem quando há um TEA, desenvolver um instrumento de psicometria válido para o diagnóstico clínico da dislexia e contribuir para o avanço de políticas de educação e do próprio diagnóstico, a partir da investigação de métodos de diagnóstico e de biomarcadores indicativos do transtorno, através da utilização de neuroimagens funcionais e estruturais. Em vista disso, o ACERTA foi organizado em três etapas: (I) revisão sistemática de literatura, (II) produção técnica e (III) testagem de campo. Todas as fases foram concluídas, de modo que, atualmente, estão sendo analisados os dados coletados ao longo dos quatro anos de duração do projeto.

Duas frentes de trabalho e coleta de dados foram adotadas na execução da terceira etapa do projeto: a das visitas às seis escolas³ da rede pública de Porto Alegre participantes do projeto – destinada à obtenção de dados comportamentais acerca do desenvolvimento da habilidade de leitura de crianças dos anos iniciais do ensino fundamental - e a do Ambulatório de Aprendizagem do Instituto do Cérebro do Rio Grande do Sul (InsCer) - destinada não só à obtenção de dados comportamentais associados ao desenvolvimento da habilidade de leitura, mas também à investigação de índices neurobiológicos de desenvolvimento da leitura, por meio da coleta e do estudo de neuroimagens funcionais e estruturais.

A coleta de dados realizada nas escolas ocorreu em três momentos: nos anos de 2014, 2015 e 2016. Os participantes realizaram testes comportamentais encontrados na literatura individual e coletivamente. A sequência de testes coletivos, aplicados em sala de aula, em todos

³ Escola Estadual de Ensino Médio Anne Frank, Escola Estadual de Ensino Fundamental Gonçalves Dias, Escola Estadual de Ensino Fundamental Professora Dinah Néri Pereira, Escola Estadual de Ensino Fundamental Prudente de Moraes, Instituto de Educação Rio Branco e Colégio Estadual Florinda Tubino Sampaio.

os alunos da turma, compreendeu um ditado ortográfico (Moojen, 2009), uma tarefa de cópia textual (Saraiva et al., 2007) e uma produção textual inspirada em uma ilustração (Salles, 2005). A sequência de testes individuais, aplicados individualmente em um espaço reservado, oferecido por cada uma das escolas, incluiu uma tarefa de compreensão textual (Saraiva et al. 2007) e outra de leitura de palavras e pseudopalavras⁴ em letra bastão (Salles, 2005).

O Ambulatório de Aprendizagem do Projeto ACERTA, ao longo dos quatro anos de realização do projeto, utilizou-se da estrutura do Instituto do Cérebro do Rio Grande do Sul (InsCer) e da PUCRS na sua realização enquanto centro de pesquisa e avaliação de crianças com dificuldades e TEA, especificamente no domínio da leitura e da escrita. Escolares com baixo rendimento acadêmico, oriundos ou não das escolas participantes do projeto, com idade entre 8 e 16 anos, foram avaliados após solicitarem atendimento no InsCer.

A adoção das duas frentes de trabalho e coleta de dados reflete a busca por evidências que sejam complementares nas respostas para as questões que fomentam o projeto ACERTA, de modo a contribuir para o diagnóstico clínico da dislexia e estabelecer um programa de intervenção em leitura, baseado em evidências e resultante de um trabalho de pesquisa experimental interdisciplinar.

Por esse motivo, o estudo e o aprimoramento de ferramentas validadas cientificamente, capazes de beneficiar o diagnóstico do transtorno, mostram-se interessantes às diferentes áreas de pesquisa envolvidas no projeto, pertinente para o ACERTA (dados os seus objetivos) e relevante para a comunidade de modo geral.

O presente trabalho reporta uma análise das produções escritas de disléxicos, de crianças com dificuldade de leitura (más leitoras) e de leitores fluentes (bons leitores), participantes do projeto ACERTA, com base na Teoria dos Grafos (TG) - uma vertente de estudos oriunda da Matemática que trata das relações entre elementos de um mesmo conjunto – e no Aprendizado de Máquina, subcampo da Ciência da Computação especializado em programar máquinas (ou computadores) para o reconhecimento de padrões, a fim de executarem tarefas.

O estudo compreendeu, em um primeiro momento, a realização de uma revisão de literatura que sustenta a interface proposta neste trabalho entre a Linguística, a Ciência da Computação, a Matemática e a Psicologia Cognitiva. Posteriormente, elegeu, para análise,

⁴ Pseudopalavras são palavras inventadas, comuns em estudos da Psicologia Cognitiva e áreas afins que investigam o processamento da linguagem (ELLIS, 1995).

dados oriundos de uma das tarefas comportamentais aplicadas nas escolas, de produção escrita (Salles, 2005), coletados em 2014, 2015 e 2016. Em seguida, os textos produzidos pelas crianças foram transcritos para, na etapa final, serem analisados.

Dado o pioneirismo do presente estudo em abordar a caracterização da dislexia e da fluência leitora a partir da análise da representação gráfica de textos produzidos por disléxicos, leitores típicos e crianças com dificuldades de leitura, assumiu-se, como propósito geral, a seguinte missão:

Identificar padrões de conectividade textual recorrentes entre crianças e adolescentes com diferentes níveis de escolaridade e de fluência de leitura – em especial disléxicos -, a partir da análise de grafos representativos de suas produções textuais e de técnicas de aprendizado de máquina.

Especificamente, este trabalho foi orientado pela expectativa de compreender:

- a) Se o tipo de transcrição textual impactaria as medidas de grafo extraídas dos textos, uma vez que o software utilizado para a extração dos atributos é sensível à quebra de linha e a correções na ortografia, componente textual expressivo de limitações impostas pela dislexia (BASSO et al., 2017);
- b) Se a normalização das medidas de grafo pelo número de palavras de cada texto seria igualmente distintiva de categorias de fluência de leitura e anos de escolaridade quanto as transcrições não normalizadas, uma vez que esta medida eliminaria o número de palavras de um texto, que consiste em um aspecto textual relevante em análises discursivas (MOTA et al., 2014; BERTOLA et al., 2014);
- c) Se leitores fluentes, leitores pouco fluentes e crianças diagnosticadas com dislexia apresentariam padrões específicos de conectividade textual, haja vista as correspondências, previstas na literatura, entre a compreensão e a produção escrita (CAPOVILLA; CAPOVILLA, 2007; NAVAS; PINTO; DELLISA, 2009; AFFONSO et al., 2011; BASSO et al., 2017);
- d) Se uma análise longitudinal evidenciaria diferenças significativas entre as produções textuais dos participantes das escolas que produziram textos ao longo dos três anos do ACERTA, de modo a registrar mudanças no desempenho acadêmico decorrentes do avanço escolar (CAPOVILLA; CAPOVILLA, 2004; CAPOVILLA; CAPOVILLA, 2007; AFFONSO et al., 2011; GÂNDARA; BEFI-LOPES, 2010);

- e) Se as médias dos atributos de grafo dos textos produzidos em 2014, pelo grupo Escolas, apresentariam proximidade com as dos textos escritos por pacientes do Ambulatório, uma vez que a produção escrita de disléxicos tende a ser substancialmente inferior à esperada para suas idades e escolaridades, e o grupo de disléxicos é bem heterogêneo quanto à idade e escolaridade (DEHAENE, 2012; (SHAYWITZ, 2008);
- f) Se o potencial preditivo da ferramenta seria, assim, corroborado com o uso de técnicas de aprendizado de máquina, dentro deste contexto de pesquisa.

As hipóteses que orientaram este estudo foram as seguintes:

- O tipo de transcrição impactaria as medidas de grafo extraídas dos textos, uma vez que o software utilizado para a extração dos atributos é sensível à quebra de linha e a correções ortográficas; textos corrigidos seriam processados e categorizados com mais acurácia pelo classificador uma vez que “normalizam” representações semânticas do discurso ao eliminarem erros ortográficos, que são “ruído” em análises discursivas;
- A normalização das medidas de grafo pelo número de palavras de cada texto, se produzisse atributos com diferença significativa entre os grupos experimentais, faria em menor escala que as transcrições não-normalizadas, uma vez que eliminaria o número de palavras que consiste em um aspecto relevante dos textos em análises discursivas (BERTOLA et al., 2014);
- Leitores fluentes, leitores pouco fluentes e crianças diagnosticadas com dislexia apresentariam padrões específicos de conectividade textual, haja vista as correspondências, previstas na literatura, entre a compreensão e a produção escrita (CAPOVILLA; CAPOVILLA, 2007; NAVAS; PINTO; DELLISA, 2009; AFFONSO et al., 2011; BASSO et al., 2017);
- Uma análise longitudinal evidenciaria que as produções textuais dos leitores fluentes das escolas que produziram textos ao longo dos três anos do ACERTA apresentariam diferenças entre suas medidas de grafo correspondentes aos anos de 2014 e 2016, por conta do avanço da escolarização (CAPOVILLA; CAPOVILLA, 2004; CAPOVILLA; CAPOVILLA, 2007; AFFONSO et al., 2011; GÂNDARA; BEFI-LOPES, 2010);

- As médias dos atributos de grafo dos textos produzidos em 2014, pelos alunos das escolas envolvidas no ACERTA, apresentariam proximidade ou similaridade com as dos textos escritos por pacientes do ambulatório de aprendizagem, uma vez que a produção escrita de disléxicos tende a ser substancialmente inferior à esperada para suas idades e escolaridades, e o grupo de disléxicos é bem heterogêneo quanto à idade e escolaridade (DEHAENE, 2012; SHAYWITZ, 2008);
- O potencial preditivo da ferramenta seria corroborado com o uso de técnicas de aprendizado de máquina, que identificariam, com razoável acurácia, correspondências entre medidas de grafo dos textos e classificações de fluência de leitura prévias, obtidas por meio do diagnóstico dos pacientes do Ambulatório de Aprendizagem do ACERTA de escores de fluência de leitura adquiridos a partir de tarefas comportamentais.

A espinha dorsal deste trabalho se estrutura em três capítulos. O primeiro deles, intitulado FUNDAMENTAÇÃO TEÓRICA, apresenta um panorama do desenvolvimento da linguagem. Inicialmente, propõe-se um percurso de investigação da escrita com o intuito de aparelhar o leitor acerca de seu valor enquanto ferramenta de inserção social e de reconfiguração mental, cujo aprendizado é priorizado nos anos iniciais do ensino fundamental e cuja falta de domínio acarreta prejuízos cognitivos e sociais. Para tanto, percorrem-se as origens das linguagens oral e escrita - a fim de reportar o que acarretou seus surgimentos e quais foram suas repercussões na vida e na natureza da espécie humana – e os tipos de ortografia decodificados pelo cérebro humano.

Em seguida, o modo como o cérebro produz e compreende a linguagem oral e escrita é discutido. Processos cognitivos e modelos teóricos relacionados à compreensão e a produção da escrita são abordados, com o objetivo de evidenciar suas convergências e divergências.

Apresenta-se e discute-se, então, o transtorno de aprendizagem denominado dislexia, a partir de suas características comportamentais, neurológicas e de dados estatísticos, capazes de ilustrar a abrangência do transtorno, que resulta em percalços de ordem cognitiva e social. Inicialmente o funcionamento cerebral de disléxicos, ao serem confrontados com o código escrito, é caracterizado; em seguida, o aprendizado e a expressão da escrita deste público são descritos e comentados.

Uma sessão dedicada à discussão sobre do modo como a interface entre Ciência da Computação e Linguística pode ser produtora e importante para o avanço de investigações correntes no âmbito da Psicolinguística finaliza a revisão de literatura empreendida.

Especificamente, discorre-se acerca do potencial da Teoria dos Grafos como ferramenta teórico-analítica, capaz de elucidar padrões de conectividade textual característicos de populações de estudo, aprimorando investigações científicas e complementando diagnósticos clínicos. Exemplos de grafos são apresentadas ao final da sessão, definindo seus componentes e caracterizando aqueles atributos que serão mensurados na presente análise.

O capítulo seguinte, denominado MÉTODOS, reúne informações a respeito dos participantes do estudo, do recrutamento realizado, dos instrumentos utilizados na coleta de dados, dos procedimentos que viabilizaram e conduziram a obtenção dos dados e das análises empreendidas – estatística e por Aprendizado de Máquina.

Após, no capítulo ANÁLISE E DISCUSSÃO DE RESULTADOS, os resultados das análises são apresentados e discutidos nesta penúltima sessão e, em seguida, a conclusão a respeito do estudo é explicitada, com base nas discussões prévias a respeito dos resultados das análises estatísticas e da categorização promovida pelos classificadores utilizados.

Por fim, entende-se que um texto escrito se estrutura como um tecido repleto de fios e pontos que determinam sua rigidez ou maleabilidade. Assim como um tecido é o produto do entrelaçamento dos seus fios e componentes têxteis, que interferem na sua aparência e textura, textos são produtos do arranjo de itens lexicais que determinam sua aparência e organização interna, determinantes para sua inteligibilidade. Escrever, portanto, é tecer conceitos ou ideias a partir de uma combinação própria de materialidades linguísticas, cujas distribuições, organizações e qualidades parecem ser, como se discute neste trabalho, passíveis de mensuração quantitativa e, sobretudo, preditivas de fluências de leitura.

1 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, apresenta-se o aporte teórico que sustenta a investigação aqui reportada e subjaz as análises e discussões dos dados comportamentais reunidos.

Dado o interesse desde trabalho em construir uma concepção ampla da dislexia e da compreensão leitora, de modo compreender e projetar seus possíveis reflexos na produção textual da população deste estudo, constrói-se aqui um panorama do desenvolvimento da linguagem verbal. Tal panorama abrange sua evolução, seus desdobramentos cognitivos – isto é, os correlatos neuronais, componentes e processos cognitivos relacionados à compreensão e à produção da linguagem verbal, oral e escrita - e as trajetórias de pesquisa que se dedicaram a teorizar sobre questões de caráter psicolinguístico associadas, especificamente, à expressão da escrita.

1.1 LINGUAGEM: DA ORALIDADE AO ARTEFATO CULTURAL DA ESCRITA

Por natureza, o homem é um animal racional tanto quanto é um ser social, que se comunica por meio da linguagem. Inicialmente, essa comunicação era totalmente dependente da oralidade - cujo surgimento e motivação ainda geram discussão.

A linguagem verbal é um produto da interação entre símbolos finitos, que transmitem sentido, e regras de combinação capazes de gerar infinitos arranjos desses símbolos, dotados de significado (PINKER, 1995). As interações entre símbolos e regras são materializadas nas modalidades oral e escrita da linguagem, e apresentam diferenças e semelhanças que se estendem de seus surgimentos até seus substratos neurais.

Estudos de caráter antropológico sugerem que as primeiras tentativas de fala ocorreram entre 1,8 e 2,5 milhões de anos atrás. Tais experiências teriam sido ocasionadas pela necessidade dos homens primitivos de compartilharem seus primeiros esforços para a fabricação de ferramentas (MORGAN et al., 2015), o que haveria favorecido o desenvolvimento de sua cognição (HOVERS, 2012), a ponto de torná-los aptos a falar sobre artefatos e, por consequência, transmitir o que sabiam, compartilhar esse conhecimento, ensinar seus semelhantes (BICKERTON, 2009).

Ao mesmo tempo, estudos em Psicolinguística, como o de Berwick et al. (2013), interessados sobretudo no desenvolvimento da linguagem humana moderna, argumentam que ela se trata de um fenômeno recente, tendo surgido entre 50 e 100.000 anos atrás, fruto de uma mutação genética pontual. Segundo a hipótese, os neandertais, subespécie do *Homo Sapiens* (chamada de *Homo sapiens neanderthalensis*) e com a qual compartilhamos mais de 90% de nosso DNA, não dispunham da linguagem da qual nos servimos hoje, por sua extinção ser anterior a esse período. No entanto, uma perspectiva diferente da citada, representada aqui por Dediu et al. (2013), contra-argumenta: o domínio da linguagem moderna se deve a um acúmulo gradual de inovações biológicas, sendo que o surgimento da linguagem da qual nos servimos hoje remete a cerca de um milhão de anos atrás – entre as origens do gênero *Homo* e o surgimento do *Homo heidelbergensis*, antepassado direto do *Homo sapiens neanderthalensis*.

Nuances de pesquisa à parte, há o consenso de que o desenvolvimento da linguagem permitiu o progresso da conquista do homem sobre si mesmo, sobre o que o cercava, uma vez que os indivíduos passaram a nomear o mundo e, por isso, a compreendê-lo com mais profundidade. A conquista da linguagem foi, possivelmente, a mais revolucionária de todas as vividas pela humanidade, uma vez que, ao atribuir sentido, categorizar e compartilhar o que via, sentia e criava, o homem se fez diferente das demais espécies.

A relação entre som e significado, obtida e compartilhada no exercício da linguagem oral, mostrou-se capaz de sofisticar, assim, a consciência do homem sobre si e sobre o mundo (DEHAENE, 2012). Contudo, a evolução tornou a espécie mais sedentária e, por isso, interessada em meios de garantir a visibilidade, a transmissão e a conservação da linguagem, independente da sobrevivência dos indivíduos e de suas limitações naturais. Diante dessa necessidade de uma modalidade de linguagem permanente, o homem primitivo recorreu, inicialmente, a arranjos de objetos, a sinais materiais, nós e desenhos para expressar, compartilhar e materializar o pensamento (HIGOUNET, 2014). Surgiu, assim, o estágio embrionário da escrita.

Há cerca de 5.000 anos, houve as primeiras tentativas de representação gráfica dos sentidos veiculados pela oralidade. Desde então, entre as iniciativas primitivas de criação de códigos escritos e o sistema alfabético atual, escritas sintéticas, analíticas e fonéticas foram desenvolvidas. As primeiras – sintéticas - consistiam em representações gráficas baseadas em sinais ou grupos de sinais limitados, que juntos expressavam uma ideia ou uma frase, tais como as pinturas rupestres de sítios pré-históricos da Península Ibérica. As analíticas representaram um avanço considerável e o surgimento, de fato, da escrita, visto que foram resultado da decomposição da frase em palavras; cada sinal designava uma palavra - e ainda designa, no

caso das línguas egípcia e chinesa. Por fim, quando o homem saltou da notação das palavras para a notação dos sons, houve a criação da escrita fonética, mais econômica e extremamente produtiva, comum à grande parte da civilização ocidental que é herdeira da escrita latina, por exemplo.

A sofisticação da escrita sintética em analítica e fonética representou o advento da escrita, com a qual foi viabilizado o acesso à materialização da linguagem oral e ao conhecimento que ela propagava, antes restrito a circunstâncias espaciais e temporais. Surgiu, assim, a ferramenta simbólica que conhecemos hoje, um artefato cultural aprendido, portanto não inato, dependente de instrução formal e resultante de um processo laborioso e vagaroso de apropriação de um sistema simbólico (LENT, 2010). Por ser tão constitutiva da civilização, a linguagem escrita serve precisamente de definição dela própria, pois revolucionou a vida do homem e, por isso, consistiu em um divisor de águas na história da humanidade: “a lei escrita substituiu a lei oral, o contrato escrito substituiu a convenção verbal, a religião escrita se seguiu à tradição lendária” (HIGOUNET, 2014, p.10). A representação gráfica da linguagem verbal recebeu o nome de ortografia.

1.1.1 Ortografias: transparentes e opacas, rasas e profundas

Ortografias são sequências ou agrupamentos de símbolos escritos que designam sentidos expressos oralmente, mapeando a linguagem falada por meio da escrita, de modo que diferentes línguas possuem diferentes ortografias ou mapeamentos (MOLL et al., 2014). Atualmente, há línguas de escrita alfabética – nas quais há correspondência entre grafemas e fonemas –, como o francês e o finlandês, assim como existem também aquelas não alfabéticas – caracterizadas por ortografias que não simbolizam diretamente os sons da fala –, como o hebraico e o sânscrito.

A invenção da escrita ocasionou o aprendizado da leitura, sempre caracterizado pelas circunstâncias ortográficas de cada língua; assim, escrita e leitura são faces de um mesmo artefato cultural. Sistemas ortográficos são dotados de diferentes níveis de profundidade, por exemplo, que interferem diretamente no processo de compreensão do código escrito pelos indivíduos leitores. A profundidade ou transparência de uma língua reflete quão pareados são os símbolos e os sons de seu sistema; quanto mais pareados são, mais transparente é a ortografia da língua; quanto menos pareados, maior a opacidade do sistema - considerando que a correspondência de diferentes sons para um única expressão gráfica exige que o leitor, na decodificação da linguagem escrita, selecione o fonema ou som adequado para o contexto da

leitura, dentre uma série de alternativas disponíveis para o que é decodificado (LUZ; BUCHWEITZ, 2016). Em resumo, são “transparentes” ou “rasas” línguas alfabéticas cuja correspondência entre grafema e fonema, majoritariamente, é biunívoca, isto é, para cada grafema há um fonema correspondente, e vice-versa. Aquelas línguas cujo sistema ortográfico alfabético não garante essa relação grafo-fonêmica direta são as chamadas línguas “opacas” ou “profundas”, e a profusão de arbitrariedades que as caracteriza, por vezes, torna a alfabetização nessas línguas mais laboriosa e demorada (DEHAENE, 2012; OJANEN et al., 2015).

A profundidade dos sistemas linguísticos tem sido atribuída aos diferentes graus de informações morfêmicas, históricas e etimológicas que caracterizam as línguas (FROST et al., 2009). As especificidades de cada sistema ortográfico colaboram para que seja possível distribuí-los em um contínuo de opacidade, de modo que em um extremo deste contínuo se encontre o finlandês, o alemão, o espanhol e o italiano – consideradas línguas transparentes ou rasas – e, no outro, o inglês, o francês, o hebraico e o mandarim – todos sistemas de conhecida opacidade. O português, língua materna dos participantes de pesquisa do presente estudo, estaria numa faixa intermediária deste contínuo, tal como ilustra a Figura 1:

Figura 1: Distribuição das línguas europeias quanto à profundidade ortográfica

Rasa	+	+	+	Profunda
Finlandês	Grego	Português	Francês	Inglês
	Italiano	Holandês	Dinamarquês	
	Espanhol	Sueco		
	Alemão			
	Norueguês			
	Islandês			

Fonte: Adaptado de Seymour et al. (2003)

Teorizar sobre o aprendizado da leitura, para Dehaene (2012, p.17) implica questionar o que ele denomina de paradoxo da leitura: “como pode ser que nosso cérebro de *Homo sapiens* pareça adaptado à leitura, quando esta atividade, inventada em todos seus componentes, não existe senão há alguns milhares de anos?”. Considerando que o surgimento da escrita remete há aproximadamente 5.400 anos, entre os babilônios, e que o alfabeto propriamente dito não

tem mais que 3.800 anos (ibid.), conclui-se que a relação do homem com a escrita é muito recente em nossa história evolutiva.

Desse modo, o cérebro dos leitores de hoje se desenvolve e opera, atualmente, sob a orientação das mesmas instruções genéticas de seus ancestrais, sem ter se reprogramado geneticamente para decodificar sequências de códigos linguísticos através do olhar. Dada a escala de tempo que compreende o processo evolutivo da humanidade, é possível afirmar que a escrita e a leitura são habilidades humanas recentes, que ocorreram a partir da adaptação do cérebro pra tais fins. Mas como a capacidade de ler se faz possível e em que medida ela se serve do aparato cerebral subjacente à oralidade? Discute-se esta questão a seguir.

1.2 O CÉREBRO HUMANO E A LINGUAGEM ORAL E ESCRITA

Conhecer as regiões cerebrais atuantes nos processos relacionados à compreensão e à produção da linguagem favorece o entendimento de distúrbios, transtornos e dificuldades de aprendizagem recorrentes entre falantes e aprendizes de línguas (ROCHA, 2015). Isso porque, como se explicita a seguir, aprender a ler consiste em acessar, por meio da visão, áreas da linguagem falada (DEHAENE, 2013). Desse modo, entender quais são as bases neurocognitivas do processamento da linguagem verbal - seja ela oral ou escrita - é pertinente quando discutimos distúrbios, dificuldades e transtornos de aprendizagem que possam comprometer a compreensão e a produção da linguagem escrita (HILLIS; RAPP, 2004).

1.2.1 O cérebro que compreende e produz a linguagem oral

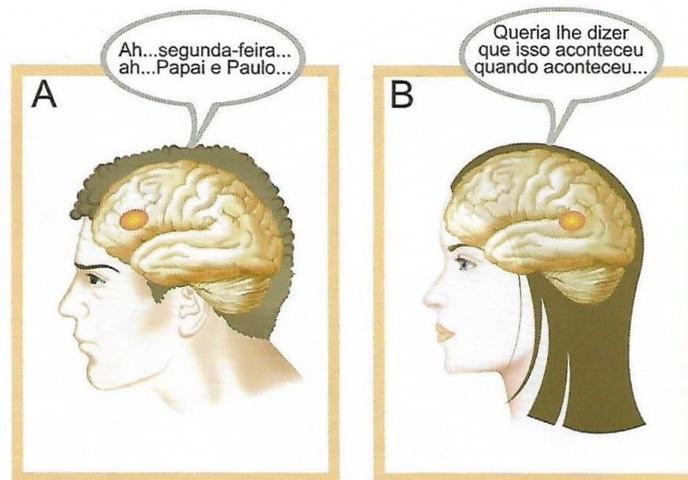
Indagações de caráter científico acerca da participação do cérebro na linguagem remetem aos primeiros estudos sobre patologias da fala relacionadas à compreensão e à produção da linguagem oral, em especial às colaborações de Pierre-Paul Broca e Carl Wernicke. Ambos foram precursores na investigação sobre afasias – distúrbios da linguagem falada ocasionados por acidentes vasculares cerebrais e lesões em regiões do cérebro diretamente relacionadas ao processamento linguístico (LENT, 2010).

Pierre-Paul Broca sugeriu que a linguagem estaria relacionada a regiões específicas do encéfalo (ibid.). Após a morte de pacientes que haviam perdido a capacidade de falar – portanto, afásicos -, o referido cientista analisou os seus cérebros e neles encontrou indícios de lesão em uma região do hemisfério cerebral esquerdo, situada na face lateral do lobo frontal, hoje

designada como Área de Broca (Figura 2). Desde então, estudos experimentais têm refinado a concepção de afasia de expressão (ou afasia de Broca), na qual indivíduos se mostram incapazes de falar, ou apresentam uma fala não fluente, caracterizada por poucas sílabas, palavras curtas e ausência de verbos.

Carl Wernicke, por sua vez, foi quem primeiro descreveu a afasia de compreensão (também chamada de afasia de Wernicke), relacionada a lesões na região córtex posterior esquerdo, próxima à ponta do sulco lateral de Sylvius (Figura 2). Pessoas com comprometimento dessa região, quer seja por conta de lesões ou acidentes vasculares cerebrais, comumente apresentam uma compreensão deficitária da fala de interlocutores, emitindo respostas verbais por vezes fluentes, porém desconexas e sem sentido.

Figura 2: Afasias de Broca e Wernicke

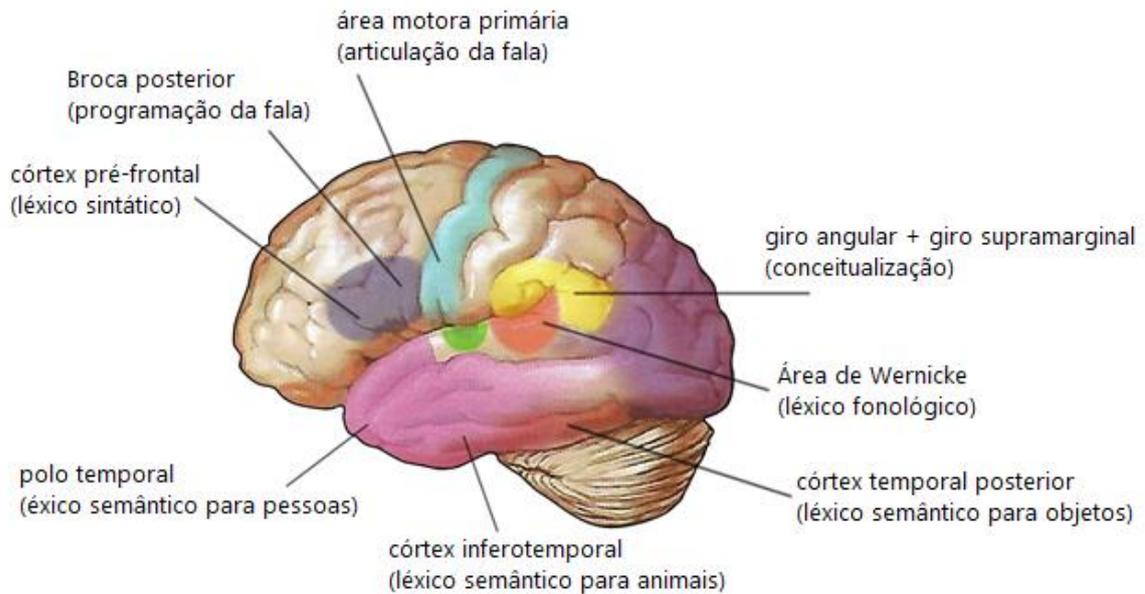


Fonte: Lent (2010)

Wernicke também propôs o primeiro modelo do processamento da linguagem oral (ibid.), que previa a conexão entre a Área de Broca e a Área de Wernicke por meio de um feixe, o chamado feixe arqueado, cuja lesão acarretaria outro tipo de afasia – a afasia de condução –, dificultando o entendimento de indivíduos acerca da própria fala e, igualmente, comprometendo suas respostas ao que lhes é dito. Tal modelo serviu de base para o atual modelo neuroanatômico conexionista da linguagem falada, concebido a partir de estudos em psicolinguística, que relativiza a anterior restrição da compreensão e da produção da linguagem oral às duas áreas citadas, uma vez que acidentes vasculares cerebrais não demonstram se circunscrever a tais áreas estritamente, mas sim atingir regiões diversas associadas à conceitualização, à articulação

da fala, à programação da fala e aos léxicos sintático, semântico e fonológico, como retrata a Figura 3:

Figura 3: Modelo conexionista da linguagem falada



Fonte: Adaptado de Lent (2010)

Sabe-se, entretanto, que o interesse das ciências pela linguagem não se restringe à sua modalidade verbal oral como objeto de estudo. Uma vez que fala e escrita representam, possivelmente, as duas vias de interação mais recorrentes na comunicação humana – e, portanto, cruciais para que o indivíduo entenda o mundo, o outro e a si -, compreende-se o porquê da vasta e diversificada investigação científica a respeito de ambas. A seguir, o modo como o cérebro compreende e produz a linguagem escrita passa a ser discutido com mais vagar, uma vez que este é o foco do presente estudo.

A oralidade e a escrita são manifestações da linguagem verbal que diferem entre si em diferentes aspectos. Uma delas é o tipo de informação sensorial e motricidade com os quais cada uma trabalha: a linguagem oral permite que o falante emita e compreenda sons produzidos por meio da motricidade orofaríngea, enquanto que a linguagem escrita o leva a decodificar imagens geradas a partir da motricidade manual (ROCHA, 2015). Sabe-se também que a oralidade é anterior à escrita; graças à remodelagem neuronal do cérebro - que, já na pré-história, era bem semelhante ao revelado pelas neuroimagens de hoje -, o homem atualmente

domina um conhecimento incomum a qualquer outra espécie e relativamente recente para a sua própria: o de extrair significado de manchas simbólicas. Contudo, a capacidade humana de atribuir sentido a ícones gráficos nem sempre foi compreendida por este viés, e é sobre esse assunto que se discute a seguir.

1.2.2 O cérebro que compreende e produz a linguagem escrita

A teoria da plasticidade generalizada - também conhecida como a perspectiva do relativismo cultural - remonta as ideias dos empiristas John Locke, David Hume e George Berkeley, e parte da premissa de que a natureza humana não é imposta biologicamente, mas flexível e maleável, construída a partir da imersão em culturas. Em outras palavras, a capacidade de aprender seria um denominador comum entre as pessoas, mas a irrestrita plasticidade cerebral permitiria o aprendizado de diferentes atividades e comportamentos, tantos quantos fossem as culturas nas quais os seres humanos se inserissem. Assim, o cérebro do homem seria um espaço onde se imprimiriam traços dos ambientes natural e cultural (DEHAENE, 2012), alheio a limitações de ordem biológica. No âmbito das ciências sociais – do qual advém essa perspectiva -, não é recorrente o estudo sobre as correspondências entre invenções culturais (como a escrita) e substratos cerebrais, uma vez que pouca relevância é conferida à biologia do cérebro e à teoria da evolução. O descolamento entre ambiente e (neuro)biologia, presente em tais teorias, parece conduzi-las a um afastamento do meio neurocientífico, dada sua falta de pertinência e compatibilidade com as premissas das neurociências. No âmbito dos estudos cognitivos, compreende-se que é ingênuo e restritivo discutir o que é mais importante para a realização da linguagem, se é o ambiente ou o biológico/genético, visto que ambos já demonstraram determinar a manifestação do fenômeno.

Uma oposição ao relativismo cultural começou a ser traçada quando a psicologia cognitiva passou a estabelecer relações entre ambiente e aparato neurobiológico. A exemplo de tal contraponto, tem-se a proposta do modelo de estágios de habilidades de leitura (FRITH, 1985). Segundo tal perspectiva, há três fases no desenvolvimento da capacidade de compreender e produzir a linguagem escrita – estágios estes que, posteriormente, comportam-se como estratégias, uma vez que permanecem disponíveis para a decodificação da escrita, de acordo com as especificidades de cada texto e das necessidades do falante. A primeira, denominada logográfica, foi definida como uma fase em que a criança identifica palavras com base em um padrão visual, ao qual atribui um significado por memorização, sem estabelecer correspondências entre letras, grafemas e fonemas. A segunda, chamada alfabética, caracteriza-

se pela identificação de palavras escritas a partir da atribuição de sons às letras ou aos grupos de letras, isto é, a partir do desenvolvimento da rota fonológica, um percurso cerebral associado à conversão de letras em sons - e vice-versa. O último estágio, conhecido como ortográfico, corresponde ao período em que a criança já reconhece a existência de palavras dotadas de irregularidades, que violam as relações até então estabelecidas entre grafemas e fonemas; assim, ela aprende a pronúncia e a ortografia da palavra por memorização. Neste estágio, as habilidades de leitura e escrita se mostram maduras, e a leitura de palavras familiares é fluente graças à consolidação da rota lexical, outro fluxo de atividade cerebral associado a um reconhecimento visual direto das palavras.

Observa-se, assim, que a oposição entre o relativismo cultural e o modelo de estágios de leitura reside, sobretudo, na concepção do desenvolvimento da compreensão e da produção da escrita como sendo processual, gradativo e definido pelas suas subjacências neurobiológicas. Habilidades sucessivas relacionadas à escrita, cujo domínio depende de processos cognitivos, passaram, assim, a serem exploradas.

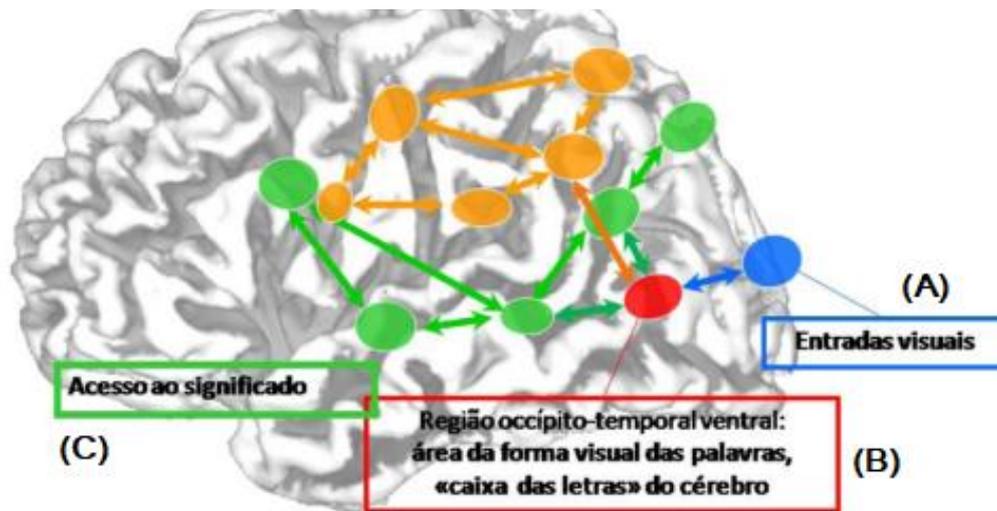
A perspectiva de Frith (1985) foi reforçada e refinada por contribuições da neurociência cognitiva. Especificamente, com o advento da neuroimagem funcional, perguntas tais como “o que é comum e diferente na rede cerebral da leitura entre indivíduos leitores?” se tornaram possíveis de serem investigadas *in vivo*, e não mais dependentes de autópsias. E as respostas revelaram, em meio a outras evidências, a recorrência de uma ativação cerebral em todos os indivíduos leitores, independente da direção da escrita da ortografia das línguas, correspondente à área occípito-temporal do hemisfério esquerdo (SCLIAR-CABRAL, 2010). Mas em que contexto se deu tal descoberta?

Evidências sugerindo a existência de correlatos neurais pré-existentes à leitura e à escrita, porém reciclados para tais finalidades embasaram o modelo da reciclagem neuronal. Tal modelo prevê uma arquitetura cerebral determinada geneticamente, mas que adquire novos contornos graças a sua plasticidade e adaptabilidade a novas aprendizagens, incluindo a leitura. Dehaene (2012, p. 20) assim explica:

Formulo, então, a hipótese de que as invenções culturais como a leitura se inserem nesta margem de plasticidade. Nosso cérebro se adapta ao ambiente cultural, não absorvendo cegamente tudo o que lhe é apresentado em circuitos virgens hipotéticos, mas convertendo a outro uso as predisposições cerebrais já presentes. Nosso cérebro não é uma tabula rasa onde se acumulam construções culturais: é um órgão fortemente estruturado que faz o novo com o velho.

Segundo a hipótese da reciclagem neuronal, a arquitetura do cérebro humano é caracterizada pela articulação de diferentes subsistemas, cujas funções e regiões cerebrais são, por vezes, recicladas para o desempenho de tarefas específicas (DEHAENE, 2013). No caso da leitura, há a articulação entre áreas envolvidas no processamento da linguagem falada e regiões do processamento visual, tal como retrata a Figura 4:

Figura 4: A compreensão leitora no cérebro



Fonte: Adaptado de Dehaene (2013)

Observa-se que há o envolvimento dos seguintes substratos e circuitos: (A) área das entradas visuais, na região occipital, onde ocorre o processamento de sinais luminosos; (B) a área responsável pelo reconhecimento da palavra escrita, localizada na região occipito-temporal ventral do hemisfério esquerdo do cérebro – denominada por Dehaene (2012) de *Visual Word Form Area* (em português, “Área da Forma Visual das Palavras” (AFVP); (C) circuitos encarregados do processamento da imagem acústica da fala até as áreas que processam o significado; (C) circuitos que processam as informações correspondentes aos gestos motores relacionados à produção da fala.

Tal mapeamento do cérebro em atividade de leitura foi favorecido por um estudo experimental que não só investigou o cérebro leitor, como também buscou identificar as funções de regiões cerebrais envolvidas no ato de ler antes de serem orquestradas para tal tarefa

(DEHAENE et al., 2010). A partir de ressonância magnética funcional (RMf), foram comparadas as atividades cerebrais de três grupos de participantes antes e após serem preparados para uma tarefa de leitura: analfabetos brasileiros sem vivências em ambiente escolar, ex-analfabetos (alfabetizados quando adultos) e indivíduos alfabetizados. Em pouco tempo de treinamento, todos aprenderam a ler, exceto os analfabetos. O que diferenciou as performances daqueles que leram foi a velocidade de leitura, uma vez que os mais rápidos foram os que já eram leitores fluentes e os mais lentos, os ex-analfabetos. Subjacente a esta diferença, estava a AFVP, cuja ativação acompanhou a decodificação, por parte dos participantes, do texto escrito. A mesma região, no entanto, respondeu a rostos, artefatos e tabuleiros de xadrez em indivíduos analfabetos, corroborando a hipótese de que sua função é reciclada para a leitura.

Segundo tal hipótese, em leitores, uma região cortical originalmente disposta ao reconhecimento da forma dos rostos se modifica e passa a identificar a forma visual das letras e palavras, durante o processo de aprendizado da leitura e da escrita: a já citada AFVP. Azevedo (2016), em consonância com a representação da Figura 4, explica que se trata de uma região cortical associativa; isto é, conecta-se tanto a áreas que processam as características do estímulo visual quanto a áreas que se ocupam da forma fonológica das palavras, do seu significado e da sua produção oral.

A partir da alfabetização e da automatização de processos no nível da palavra, a Área da Forma Visual das Palavras se especializa e se adapta, especificamente, para processamento das palavras, reciclando-se. Logo, o desenvolvimento da capacidade de compreender e produzir a linguagem escrita depende da exposição sistemática e instruída ao código linguístico gráfico, pois a habilidade de decodificar, ao contrário da habilidade oral, não é instintiva. Enquanto a primeira é fruto de um processo de aprendizado, a segunda é resultante de uma aquisição geneticamente programada.

Em geral, a rede neural da linguagem verbal e da leitura é equivalente para falantes e leitores de diferentes línguas e culturas. Um estudo recente mostra esta sobreposição da assinatura neural da linguagem e da leitura em inglês, espanhol, chinês e hebraico (RUECKL et al., 2015). Salvo diferenças milimétricas, o processamento da linguagem escrita recruta sempre essa área na junção occipito-temporal – a AFVP -, quer esteja em inglês ou em chinês o fragmento simbólico processado, na identificação de palavras (DEHAENE, 2012). Mesmo quando as palavras são processadas pelo cérebro de maneira inconsciente – como quando são apresentadas rapidamente em vídeos ou em testes experimentais no computador -, a região é ativada (DEHAENE, 2010). A resposta da AFVP para letras e palavras é considerada um

marcador de fluência e automaticidade na leitura. Consequentemente, a hipoativação desta região é indicativa de dificuldades ou transtornos de aprendizagem.

A fim de extrair significado das manchas simbólicas da escrita, o cérebro mobiliza regiões específicas, que se associam para que produzam sentido daquilo que o indivíduo enxerga. A seguir, modelos de compreensão leitora são discutidos.

1.2.3 Modelos de Compreensão Leitora

Diferentes vertentes de pesquisas cognitivas dedicam-se ao estudo da compreensão e da produção textual, que originaram diferentes modelos do processamento do código escrito e que seguem elucidando o fenômeno da leitura, cada qual a sua maneira (VAN DEN BROEK; YOUNG; LINDERHOLM, 1999). Recebem destaque, no presente estudo, os trabalhos que se caracterizaram por atribuir aos elementos textuais importância basilar na manutenção da coerência interna do texto e aqueles que se voltaram, essencialmente, para o estudo do próprio processo da leitura.

Originalmente, estudos acerca da leitura eram motivados pelo entendimento de que a memória dos textos era sistemática, não aleatória – isto é, havia uma lógica subjacente ao que era lembrado, as lembranças não se davam ao acaso. Assim, tinham o propósito de desvendar o que caracterizava e definia quais eram as lembranças de leitura dos indivíduos, bem como teorizar sobre a natureza desse tipo de memória (MANDLER; JOHNSON, 1977; STEIN; GLENN, 1975) Estes trabalhos foram favorecidos pela consolidação da vertente de estudos dita *bottom-up*, cujo princípio é o de que cada elemento textual realiza a manutenção da coerência do texto ao se relacionar com os demais elementos. Os modelos oriundos deste período propunham representações mentais construídas pelos leitores e sistematizavam a edificação da coerência extraída da leitura, em relação aos seus referentes (KINTSCH; VAN DIJK, 1978) e à estrutura causal do texto (TRABASSO; VAN DEN BROEK, 1985). Van Den Broek, Young & Linderholm (1999) salientam que esta linha de pesquisa se serviu do advento de metodologias e instrumentos que passaram a mensurar atividades de leitura *online*, a fim de descrever e entender o que ocorre quando os leitores leem um texto. A atenção das pesquisas, portanto, havia mudado do produto da leitura e de sua memória para o processo presente no ato de ler. Os modelos se destinaram, então, a representar os processos cognitivos que ocorrem *online*: as inferências realizadas pelos leitores, a interação entre os limites de recursos de atenção e a necessidade de compreensão, etc. Em decorrência disso, surgiram, entre os outros, os modelos

denominados Modelo de Construção-Integração⁵ (KINTSCH, 1988), Modelo de Estratégia de Estado Atual⁶ (FLETCHER; BLOOM, 1988) e Criador de Inferências Causais⁷(VAN DEN BROEK, 1990).

O propósito da terceira geração, que se consolidou a partir de 1990, foi integrar os aspectos *online* da leitura - que já vinham sendo estudados e representados por diferentes modelos - a aspectos *offline*. Em outras palavras, a geração se dedicou a estudar e descrever a relação entre os processos cognitivos associados à compreensão leitora e à representação da memória (GOLDMAN; VARMA, 1995; TRABASSO; LANGSTON, 1998). Um exemplo expressivo de tal vertente é o de Kintsch e Van Dijk (1978), no qual propuseram um novo entendimento da compreensão leitora, ao questionar o valor da palavra e da sentença enquanto unidades fundamentais de sentido de um texto. Essa concepção culminou em um novo modelo de processamento discursivo, cuja principal premissa era a da divisão do texto em unidades mínimas semânticas, as ditas proposições, cruciais na composição da compreensão textual (BUCHWEITZ, 2008). Uma análise panorâmica da proposta de Kintsch e van Dijk permite associar o modo como abordam o texto a uma espécie de janela para a mente humana, tal como se espera de um modelo mental. Os autores partem da materialidade textual para retratarem operações mentais abstratas, que se tornam tangíveis unicamente pela decodificação do texto; distinguem três processos de compreensão que transcorrem tanto paralela quanto sequencialmente, abdicando da concepção sequencial de processo de leitura, a saber: (a) organização das proposições em um agrupamento coerente, (b) redução do sentido do texto à sua essência e (c) geração de novos textos pelos leitores, a partir do rearranjo das informações apreendidas.

Como mencionado anteriormente, a compreensão textual se constrói enquanto processo cognitivo superior a partir do arranjo de diferentes fenômenos, tais como percepção das palavras, análises sintática e semântica e inferenciação. Logo, são muitos e complexos os fenômenos subjacentes ao entendimento de um texto. Contemporaneamente, tamanha complexidade costuma ser associada às interações entre múltiplos níveis – níveis lexical e sentencial, por exemplo – e sistemas da linguagem – como os sistemas sintático e semântico – da linguagem, que demandam bastante atenção e memória ao serem processados (GOLDMAN; VARMA, 1995; GRAESSER; SINGER; TRABASSO, 1994; VAN DIJK; KINTSCH, 1983;

⁵ Originalmente definido como *Construction-Integration Model* (KINTSCH, 1988)

⁶ Originalmente definido como *Current State Strategy* (FLETCHER; BLOOM, 1988)

⁷ Originalmente definido como *Causal Inference Maker* (VAN DEN BROEK, 1990).

VAN DEN BROEK, 1990). Diante de tal complexidade e diversidade, a profusão de modelos teóricos que se dedicaram e ainda se dedicam a explicar o processamento textual o fazem, comumente, a partir de dados comportamentais - especificamente, textos escritos - o que reflete a relação de interdependência entre compreensão leitora e produção textual.

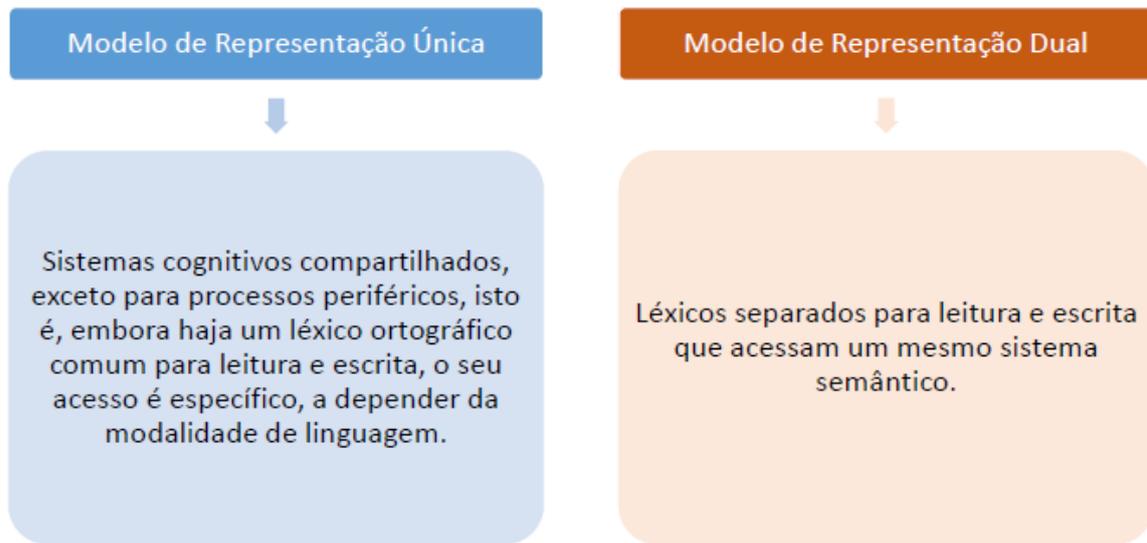
A seguir, convergências e divergências entre leitura e escrita são debatidas.

1.2.4 Convergências e divergências entre leitura e escrita

Compreender e produzir textos escritos são tarefas que dependem do estabelecimento de relações entre os diferentes componentes do sistema linguístico (elementos fonológicos, semânticos, sintáticos e pragmáticos) e da informação extratextual armazenada pela memória de longo prazo, recrutada nas inferências realizadas e na acomodação da nova informação produzida (SCHERER, 2009). Tais tarefas, em outras palavras, demandam a realização de processos cognitivos ditos superiores, que geram produtos elementares para o sucesso da vida escolar e social dos seres humanos: a leitura e a escrita.

Intuitivamente, produzir e compreender textos escritos parecem duas faces da mesma moeda, mas não há consenso na literatura quanto a isso. Objetivamente, há dois modelos que defendem configurações distintas subjacentes às habilidades de leitura e escrita – um que prevê a ocorrência de um único léxico, comum a ambas as habilidades, e outro que advoga a existência de léxicos distintos para produção e compreensão da escrita: o primeiro, intitulado Modelo de Representação Única, e o segundo, Modelo de Representação Dual (HILLIS; RAPP, 2004). Como sintetiza a figura abaixo, eles abordam as convergências e divergências entre leitura e escrita de maneiras distintas:

Figura 5: Modelo de Representação Única e Modelo de Representação Dual



Fonte: A autora

O Modelo de Representação Única concebe leitura e escrita como faces de uma mesma – “única” – moeda, que compartilha a conversão grafema-fonema, o buffer grafêmico, o léxico ortográfico, o sistema semântico e o léxico fonológico. Admite, inclusive, a existência de um léxico ortográfico comum para leitura e escrita, mas cujo acesso, considerado um processo periférico, é específico de cada modalidade de linguagem.

O Modelo de Representação Dual, por sua vez, postula que leitura e escrita compartilham apenas um mesmo sistema semântico, o que significa que a conversão ortográfico-fonológica bem como seus léxicos são particularizados, distintos entre si.

Ambas as perspectivas tem orientado estudos no campo da Psicologia Cognitiva, que tanto reportam déficits e sistemas cognitivos diferentes para leitura (CASTLES; COLTHEART, 1993) quanto advogam que leitura e escrita compartilham representações mentais (EHRI, 2005). Em tese, o que é aprendido através de uma habilidade favorece a outra, mas não necessariamente no mesmo grau (ANGELELLI; MARINELLI; ZOCCOLOTTI, 2010); embora representações lexicais ortográficas incompletas consigam sustentar a leitura, a escrita exige representações especificadas para um bom desempenho. Crianças e adultos com boa fluência leitora conseguem ler palavras que não podem soletrar corretamente porque muitas das representações subjacentes a suas leituras são imprecisas ou incompletas, mas contêm informações ortográficas suficientes para a realização da tarefa (FRITH, 1985).

O Modelo Dual encontra fortes evidências em estudos acerca de déficits de leitura e de escrita em pacientes com danos cerebrais, que reportam comprometimentos assimétricos

para produção e compreensão da escrita em adultos que, anteriormente, apresentavam bom desempenho para ambas as habilidades (ANGELELLI; MARINELLI; ZOCCOLOTTI, 2010). No entanto, associações entre comprometimentos em ambas as habilidades são mais comuns que dissociações. Déficits relacionados à leitura e à habilidade ortográfica, não raro, compartilham um padrão de prejuízo, independentemente do grau de profundidade da língua (WIMMER; MAYRINGER, 2002; ANGELELLI; MARINELLI; ZOCCOLOTTI, 2010). Por exemplo, estudos reportam crianças com disgrafia⁸ e dislexia de superfície, assim como crianças com dislexia e disgrafia fonológicas, indicando um paralelismo entre os tipos de prejuízos (HANLEY; HASTIE; KAY, 1992; CURTIN; MANIS; SEIDENBERG, 2001). Além disso, o diagnóstico da dislexia muitas vezes é acompanhado pela classificação quanto aos subtipos “fonológica” ou “de superfície”, com base nas habilidades ortográficas que apresentam, o que sugere uma convergência entre os domínios da leitura e da escrita.

Logo, não há consenso quando a convergências e divergências entre leitura e escrita, o que reforça a importância de se aprofundar esta discussão com novos estudos.

As digressões por entre teorias e modelos interessados nos processos cognitivos relacionados à leitura e à escrita foram elucidativas da importância de modelos de processamento da linguagem para o entendimento de como são complexas a compreensão e da produção textual. Aprender a ler e a escrever textos, portanto, são tarefas que demandam variados esforços cognitivos de uma criança, e nem todos os alunos conseguem dispor destes recursos cognitivos, de maneira que, por vezes, avançam na vida escolar sem dominar plenamente tais competências. A dificuldade vivenciada por estes jovens em suas experiências de alfabetização, não raro, é decorrente de um transtorno de aprendizagem, como se discute a seguir.

1.3 DISLEXIA: UM TRANSTORNO DE APRENDIZAGEM

Cerca de 40% dos estudantes em séries iniciais de alfabetização apresentam dificuldades de aprendizagem atribuídas a múltiplas causas: falta de oportunidade social, ambiente cultural pouco estimulante, limitações socioeconômicas, falhas no acesso ao ensino e

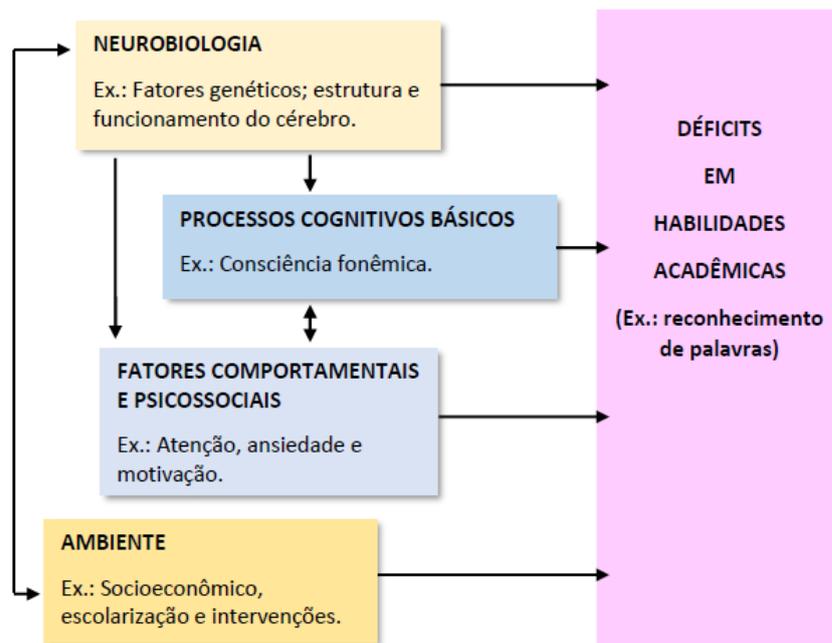
⁸A disgrafia é considerado um distúrbio da linguagem escrita comumente decorrente de problemas psicomotores, sem relação prejuízo, necessariamente, do desenvolvimento intelectual (FLETCHER et al., 2009)

a métodos pedagógicos adequados ou a algum tipo de distúrbios ou transtornos da aprendizagem (ARAÚJO, 2011; HOEFT et al., 2011; MEYLER et al., 2007; SHAYWITZ et al., 1998; SIGMAN et al., 2014).

Historicamente, transtornos de aprendizagem têm se mostrado constructos de difícil definição, portanto alvos de variadas iniciativas de descrição. Isto se deve à sua natureza, uma vez que eles se comportam como variáveis latentes não-observáveis, que não existem ou são percebidas senão quando mensuradas – como é o caso do quociente de inteligência (QI). A complexidade dos TEA advém dos diferentes fatores atuantes sobre os resultados acadêmicos das crianças, a saber: fatores neurobiológicos, processos cognitivos básicos, fatores comportamentais e psicossociais e fatores ambientais. O desempenho acadêmico dos alunos, portanto, atua como a principal manifestação dos transtornos neste público. Lesões cerebrais, consciência fonêmica deficitária, ansiedade, déficit de atenção e diversas privações comuns a estratos socioeconômicos próximos da linha da pobreza, por exemplo, são fatores capazes de delinear a performance dos alunos em atividades escolares e, assim, comprometer seus domínios da compreensão e da produção da escrita – entre outras habilidades acadêmicas.

A Figura 6 sistematiza a atuação destes fatores:

Figura 6: Fatores atuantes sobre o desempenho acadêmico infantil



Fonte: Adaptado de Fletcher et al. (2009)

Além disso, os traços que configuram os transtornos de aprendizagem se distribuem em um *continuum* de gravidade, e não entre categorias discretas ou dicotômicas. Por essa razão, como afirmam Fletcher et al. (2009, p. 41),

Essas iniciativas representam tentativas de avaliar o desempenho *inesperadamente baixo*, que geralmente é conceituado como indivíduos que não conseguem aprender as habilidades acadêmicas, mesmo na ausência de causas conhecidas para o desempenho inferior (transtorno sensorial, retardo mental, distúrbios emocionais, desvantagens econômicas, diversidade linguística, instrução inadequada). Muitas iniciativas de definição e de identificação foram tentativas de medir este atributo, que simboliza o constructo do TEA. A principal abordagem para a identificação de transtornos de aprendizagem é a mensuração da desigualdade entre o desenvolvimento acadêmico e o cognitivo como indicador do caráter “inesperado” dos TEAs, juntamente com a exclusão de outras causas que poderiam justificar um baixo desempenho.

A Figura 6 ilustra como TEAs se associam a déficits em habilidades acadêmicas específicas. Por essa razão, três níveis de análise baseados na avaliação das habilidades acadêmicas dos escolares são estabelecidos: o primeiro corresponde aos próprios déficits acadêmicos observados em atividades escolares, o segundo é orientado por fatores processuais-cognitivos básicos e comportamentais/psicossociais e o terceiro, por fatores neurobiológicos e ambientais. Processos cognitivos básicos, como a consciência fonológica, exercem uma determinação direta sobre a compreensão e a produção da escrita, por exemplo. Fatores comportamentais/psicossociais, por sua vez, estabelecem uma relação de mútua influência com os processos cognitivos básicos, uma vez que disfunções cognitivas podem acarretar, entre outros prejuízos, problemas de atenção. Fatores neurobiológicos determinam características neurológicas e genéticas da criança, enquanto que fatores ambientais correspondem a questões de ordem social e econômica que atuam sobre o desenvolvimento e o aprendizado da criança. Há uma interação entre ambos os domínios neurológicos e ambientais, como retrata a Figura 6. Fatores neurobiológicos também atuam sobre processos cognitivos básicos e fatores comportamentais/psicossociais, uma vez que especificidades neurológicas podem definir o sucesso ou a deficiência de processos cognitivos, bem como o comportamento de um indivíduo.

Uma vez compreendida a dinâmica de influências subjacente aos TEAs, resta direcionar esta discussão para o foco deste estudo: a dislexia. A falta de consenso quanto à sua definição entre pesquisadores e profissionais especializados costuma desfavorecer o diagnóstico precoce do transtorno (PUNT, 2010). No entanto, assume-se aqui a perspectiva de que a dislexia consiste em uma dificuldade de aprender a ler imprevista, apresentada por

crianças cujas condições de aprendizado se assemelham às normais, sem patologias ou déficits cognitivos que possam prejudicar seu aproveitamento escolar (APA, 2013; SHAYWITZ, 2005; SHAYWITZ, 2008).

Há dois tipos de dislexia, cujas naturezas são distintas: a dislexia do desenvolvimento e a dislexia adquirida (SALLES; PARENTE; MACHADO, 2004). A dislexia do desenvolvimento, também chamada de primária ou específica, caracteriza-se pelo fracasso constitucional na aquisição da capacidade de leitura; um fruto hereditário em primeira análise. No entanto, a hereditariedade, quando associada a fatores ambientais prejudiciais, tende a ter sua influência potencializada no desenvolvimento da dislexia (PENNINGTON, 2006).

A dislexia adquirida, também denominada de sintomática, consiste em um transtorno de leitura comumente decorrente da afasia (VUKOVIĆ; VUKOVIĆ; MILLER, 2016) e costuma comprometer as habilidades de fala, leitura e escrita que eram comuns ao indivíduo antes de ser acometido pela afasia. A literatura tem investigado questões relacionadas à dislexia adquirida a partir de estudos que analisam, por exemplo, o papel do hemisfério direito na recuperação de afásicos (COCQUYT, 2017) ou padrões de dislexia adquirida em sérvios afásicos, comparando perfis de grupos com afasia de Broca com outros com afasia de Wernicke (VUKOVIĆ; VUKOVIĆ; MILLER, 2016).

Há, ainda, subtipos de dislexia derivados da proposta do Modelo Cognitivo de Dupla Rota da leitura (COLTHEART, 2013): a dislexia de superfície e a dislexia fonológica (CASTLES; COLTHEART, 1993). O modelo em questão advoga que o processamento da ortografia envolve um sistema sublexical – que relaciona grafemas e fonemas a partir de regras fonológicas – e um sistema visual ortográfico – que depreende o significado diretamente. Quando há um comprometimento mais circunscrito ao sistema sublexical, há deficiência no processamento fonológico e configura-se um quadro de dislexia fonológica. Quanto o sistema lexical é o comprometido em maior medida, o componente ortográfico da leitura é prejudicado, configurando um caso de dislexia de superfície. Na prática, a dislexia fonológica dificulta a leitura de pseudopalavras, enquanto que a dislexia de superfície compromete mais seriamente a leitura de palavras irregulares (FLETCHER et al., 2009).

Entende-se que a competência na leitura envolve fluência e automatismo (mensurados através da precisão e da rapidez do sujeito) no reconhecimento de palavras e compreensão do material lido (SALLES; PARENTE; MACHADO, 2004). Se comparados a leitores típicos, indivíduos com dislexia apresentam baixa precisão e rapidez de leitura; o principal déficit observado entre disléxicos, quando analisadas suas habilidades acadêmicas, consiste na decodificação de palavras isoladas que, por consequência, implica em uma dificuldade

substantial na compreensão de textos escritos (SHAYWITZ, 2008). Isto porque a atribuição de sentido à escrita é favorecida pela decodificação rápida e pelo reconhecimento automatizado de palavras isoladas; quando esta automatização não ocorre, a compreensão fica comprometida⁹.

A capacidade leitora significativamente deficiente, como explicitado, costuma ser uma manifestação da dislexia. No entanto, é imprescindível elucidar a diferença entre leitores pouco fluentes (ou maus leitores¹⁰) e disléxicos. Enquanto os primeiros podem dispor de uma surdez mal remediada, um tipo de retardo mental, condições de educação insipientes ou simplesmente dificuldade em assimilar regras de ortografia complexas e arbitrárias (DEHAENE, 2012), crianças disléxicas, por sua vez, sofrem de limitações naturais e irreparáveis, apenas passíveis de atenuação.

Há especificidades no perfil de adultos e crianças disléxicos; na infância, as crianças com dislexia consideram difícil aprender os grafemas e suas correspondências com os fonemas da oralidade; na fase adulta, os indivíduos com dislexia apresentam déficit na precisão da leitura de palavras, o que afeta suas velocidades de leitura e suas produções escritas (SHAYWITZ, 2008).

Possivelmente, um dos aspectos primordiais a serem destacados a respeito da dislexia seja o fato de que não se trata de uma doença, mas sim de um transtorno de aprendizagem com implicações cognitivas e sociais na vida de disléxicos. Uma vez que não se trata de uma patologia, a dislexia não tem cura. Para aqueles que são disléxicos, a tarefa de decodificar um texto escrito sempre será árdua, e esse percalço constante remete a questões essenciais relacionadas às bases neurofisiológicas do transtorno, explicitadas a partir de agora.

1.3.1 O cérebro com dislexia

O cérebro humano, como mencionado anteriormente, não é programado para a ler, mas o faz graças a um processo de adaptação pelo qual passa, que altera circuitos da visão e os prepara para o processamento da escrita (DEHAENE, 2012; DEHAENE et al., 2010). Por essa razão, dedicar especial atenção às características neurobiológicas da dislexia mostra-se

⁹ Na sessão x, página x, esta questão será discutida mais detalhadamente.

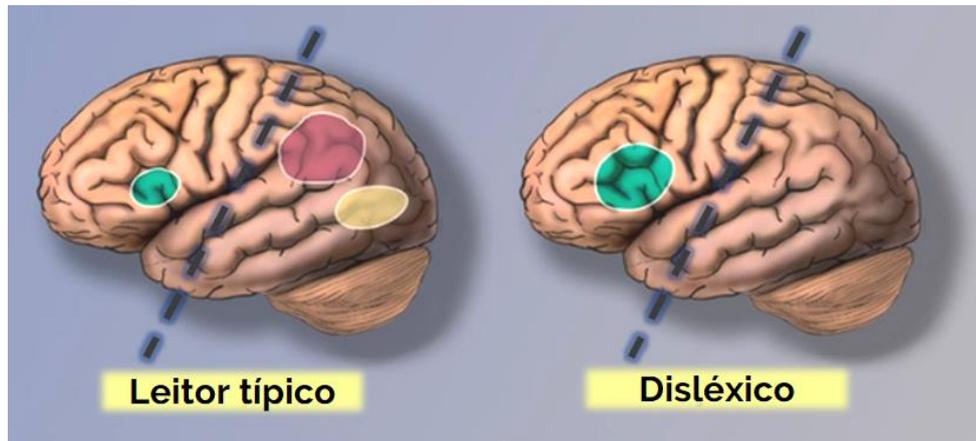
¹⁰ Adaptando a nomenclatura de Ganschow & Sparks (1986), utilizamos “maus leitores” para designar crianças com dificuldades de leitura que podem sofrer de uma dislexia não diagnosticada, e “bons leitores” como um sinônimo de aprendizes com desenvolvimento típico de leitura.

imprescindível para a compreensão do que torna a leitura de disléxicos mais laboriosa do que a das demais crianças.

A Área da Forma Visual das Palavras, discutida anteriormente, é modificada para processar a escrita; após o processamento visual primário, essa região identifica letras e tem início, então, a compreensão da leitura, que depende de um processamento que percorre o cérebro por meio das rotas fonológica e lexical. Em crianças disléxicas, este circuito não se forma da maneira como ocorre com as demais crianças, cujo aprendizado transcorre com relativa facilidade (BUCHWEITZ, 2016). A configuração cerebral subjacente à dislexia apresenta anomalias recorrentes, tais como desorganização da anomalia do lobo temporal, alteração da conectividade de regiões corticais e subativação do lobo temporal esquerdo no curso da leitura (GUARESI; ALMEIDA, 2015). Essa desordem cerebral implica em dificuldade no aprendizado da leitura e da escrita, uma vez que as associações sinápticas imprescindíveis para o processamento da leitura não ocorrem, já que a reciclagem de células nervosas especializadas é deficitária.

Em resumo, a dislexia é fruto de uma atividade funcional cerebral anômala; várias regiões recrutadas normalmente para a compreensão leitora são subativadas em disléxicos, enquanto outras são superativadas. Especificamente, pessoas com dislexia apresentam dificuldades atribuídas a um déficit no componente fonológico da linguagem (RAMUS; PIDGEON; FRITH, 2003; DEUSCHLE; CECHELLA, 2009). Por exemplo, em estudos de neuroimagem voltados para a investigação da dislexia, é recorrente uma subativação na região temporal posterior esquerda (Figura 7), associada ao processamento fonológico, tanto em adultos quanto em crianças disléxicas (SHAYWITZ, 1998). Igualmente comum entre estes sujeitos é a superativação do córtex frontal inferior esquerdo - região da Área de Broca – durante a leitura (DEHAENE, 2012; GEORGIEWA, 2002). Estes resultados fazem sentido se for considerado o fato de que, em leitores típicos ou fluentes, as áreas posteriores associadas à linguagem demonstram mais ativação que as áreas anteriores – uma vez automatizada a leitura, não há necessidade do intenso trabalho de regiões articulatórias. Estes seriam indícios de mecanismos de compensação do cérebro, motivados pela fraca atividade de regiões posteriores de decodificação; um esforço para viabilizar a experiência de leitura, mesmo sendo ela, nesse caso, controlada e consciente (DEHAENE, 2012).

Figura 7: Atividade cerebral de disléxicos e de leitores normais, durante a leitura



Fonte: Adaptado de Shaywitz (2003)

Sendo ou não normais suas funções cerebrais para a leitura, crianças brasileiras em fase de alfabetização são inseridas em um sistema educacional que, geralmente, adota estratégias de ensino e aprendizagem da leitura e da escrita que recorrem a duas possibilidades de abordagem: a global ou a fônica. Discutiremos, a seguir, a relação entre as abordagens de ensino adotadas e os resultados dessas escolhas, por vezes refletidos nas produções escritas de bons e maus leitores.

1.3.2 A alfabetização e a produção escrita de disléxicos

Como fora discutido brevemente no início deste trabalho, as línguas naturais conhecidas possuem diferentes ortografias, e muitas das dificuldades para ler textos escritos associam-se às características da ortografia da língua ensinada. Ojanen et al. (2015) sugerem que ortografias transparentes são aprendidas com mais rapidez se comparadas às opacas. Corroborando essa hipótese, pesquisas sobre fluência evidenciam essa diferença; em inglês, após um ano de alfabetização, crianças ainda cometem mais de 60% de erros de leitura; em espanhol e italiano – línguas mais transparentes que o inglês e, portanto, mais pareadas para sons e símbolos gráficos, sem grandes variações -, essa taxa cai para 5% (DEHAENE, 2012).

Cabe salientar que há uma relação causal e bidirecional entre o processamento fonológico e o desenvolvimento da capacidade de produzir textos escritos (NAVAS; PINTO; DELLISA, 2009). Isto se deve, essencialmente, a duas razões: (a) as habilidades de processamento fonológico atuam como alicerce para o aprendizado da linguagem escrita e (b) a competência em leitura e escrita conduz ao desenvolvimento de níveis mais refinados de

processamento fonológico, em uma relação de causalidade recíproca (CAPOVILLA; CAPOVILLA, 2004; CAPOVILLA; CAPOVILLA, 2007; AFFONSO et al., 2011).

É coerente atribuir à natureza da ortografia um papel importante no aprendizado da leitura e da escrita, que costuma ser orientado, no Brasil, pelos métodos fônico e global. O primeiro estrutura-se a partir da concepção de que a leitura, por ser diferente na fala, requer instrução formal do alfabeto e dos seus respectivos sons correspondentes, a fim de promover a consciência fonológica (ARAUJO, 2011; LUZ; BUCHWEITZ, 2016). O método global, em contrapartida, não adota os grafemas isolados como unidades de leitura, mas sim as palavras inteiras; ele promove o aprendizado a partir da memorização da pronúncia da palavra toda e não de uma parte dela, de modo que o ensino do som das letras do alfabeto não ocorre explicitamente.

Independentemente do método de ensino adotado, a leitura fluente é conquistada a partir da prática exaustiva; quando esta ocorre e o ato de ler se torna automatizado, os recursos cognitivos que antes eram destinados prioritariamente para a decodificação de palavras passam a ser utilizados, majoritariamente, para a compreensão leitora (RICHARDSON; LYYTINEN, 2014). Em outras palavras, a compreensão de textos escritos em crianças disléxicas é prejudicada porque estas não conseguem automatizar o reconhecimento de palavras, mesmo após praticarem e tentarem exaustivamente se apropriar das letras do alfabeto e de seus fonemas correspondentes. Logo, os processos de reconhecimento de palavras demandam esforço cognitivo demasiado destas crianças, restando assim menos recursos cognitivos para os processos superiores que levam à integração e à compreensão de um texto (FLETCHER et al., 2009). Neste sentido, sabe-se que disléxicos são menos numerosos em nações cujas línguas têm ortografia transparente e beneficiam-se do método de ensino fônico, que não sobrecarrega sua memória e tem se mostrado mais eficaz na remediação da dislexia (ARAUJO, 2011).

Além de acarretar às crianças pouca precisão no reconhecimento de palavras e, conseqüentemente, comprometer a compreensão do texto escrito em graus variados, a dislexia se caracteriza por afetar habilidades ortográficas (APA, 2013). A tarefa de produzir um texto escrito pode ser substancialmente árdua para crianças e adultos disléxicos, considerando que se trata de uma atividade que demanda funções cognitivas sofisticadas e habilidades básicas de escrita. As alterações no cérebro e na relação com a escrita de disléxicos não se devem obrigatoriamente a fatores de ordem intelectual ou sensorial, nem mesmo a falhas na educação formal das crianças. A literatura disponível credita estes déficits, como mencionado anteriormente, a predisposições genéticas em especial (FUCHS et al., 2001; LISHMAN, 2003;

SHAYWITZ, 2008; LAW et al., 2014). Ao fim e ao cabo, tanto o déficit nas habilidades fonológicas quanto a menor exposição a circunstâncias de leitura - ambos decorrentes da dislexia - indicam que, apesar de se tratar de um distúrbio específico de leitura, a dislexia costuma ser acompanhada por problemas com a escrita (AFFONSO et al., 2011).

Dadas as circunstâncias neurológicas e comportamentais relacionadas ao transtorno e suas notórias implicações na escrita, resta a pergunta: especificamente, de que modo a produção escrita de pacientes diagnosticados com dislexia reflete o seu transtorno de aprendizagem? O que se pode inferir a partir da produção escrita sobre a fluência leitora de uma criança ou de um adulto? Historicamente, as pesquisas acerca da dislexia se concentraram mais na leitura do que na ortografia de estudantes disléxicos. Contudo, a literatura disponível sobre a segunda é elucidativa de meios de se abordar a questão, a partir de diferentes desenhos experimentais, ferramentas de análise e populações de estudo.

A começar pela escrita de adultos disléxicos, Tops et al. (2013) encontraram indícios de que o comprometimento do componente fonológico da linguagem torna mais difícil a associação de grafemas a fonemas, a velocidade e a acurácia na escrita de palavras em adultos disléxicos, na Bélgica. Compararam os textos escritos em holandês de participantes com dislexia e leitores fluentes com a mesma idade, cursando o ensino superior, a fim de comparar a escrita de jovens adultos com dislexia com a de adultos jovens saudáveis. Foram analisados a quantidade e a natureza dos erros ortográficos, a qualidade dos textos produzidos, o uso de palavras e a caligrafia, por meio de uma tarefa de produção de resumo de um texto informativo e por meio de outra tarefa de ditado de sentenças. Os resultados mostraram grande incidência de erros ortográficos e morfossintáticos nos textos de participantes com dislexia. A qualidade dos textos produzidos pelos participantes disléxicos se mostrou inferior à dos textos do grupo controle, assim como a estrutura e a coesão textuais, ainda que o número e os tipos de palavras utilizados pelos dois grupos tenham sido similares.

Destaca-se, também, dentre os estudos a respeito da produção escrita de adultos disléxicos, o de Basso et al. (2017), no qual foram comparados textos de adultos diagnosticados com dislexia a textos de leitores adultos típicos. Para cada sujeito de pesquisa disléxico, foi pareado um participante controle, levando em consideração sexo, idade, anos de escolaridade e ocupação. Cada participante redigiu um texto narrativo autobiográfico, a respeito de experiências pessoais associadas à vida escolar. Estes textos foram analisados quanto à complexidade estrutural, erros ortográficos e métricas textuais obtidas por meio da análise computadorizada realizada pelo software *Coh-matrix-Port*. Os resultados mais expressivos das características dos textos dos participantes com dislexia, se contrastados com os controles,

indicaram textos com frases mais longas, palavras mais curtas, alta frequência de adjetivos e baixa incidência de substantivos ambíguos. O uso recorrente de “e” também consistiu em uma característica relevante e correlacionada com a quantidade de erros ortográficos e QI, indicando uma dificuldade do grupo clínico em estabelecer conexões textuais no texto, por meio do uso variado e adequado de conectivos. Além disso, os textos dos disléxicos se revelaram pouco estruturados (distribuídos, em sua maioria, entre as categorias I e II de uma escala de cinco categorias de complexidade estrutural crescente) e falhas ortográficas foram mais numerosas entre os disléxicos, sugerindo que esta dificuldade, comum na infância de crianças com este transtorno de aprendizagem, persiste na fase adulta.

No que tange a escrita de crianças com dislexia, Nolan e Caramazza (1983) desenvolveram um estudo de caso com um paciente diagnosticado com dislexia profunda, a fim de melhor compreender sua capacidade de produzir linguagem escrita e falada. Por meio de uma bateria de testes, este paciente disléxico produziu erros semânticos, visuais e derivativos. Além disso, ele exibiu efeitos abstratos e discursivos na leitura oral, na nomeação oral e escrita e na escrita orientada por ditado. Este paciente corroborou, assim, a hipótese apresentada em Nolan e Caramazza (1982), segundo a qual os sintomas que caracterizam a dislexia profunda são possíveis de serem observados nas respostas a quaisquer tarefas que requeiram mediação lexical. Isto porque as respostas escritas do paciente em todas as tarefas, exceto para uma tarefa de cópia direta, foram caracterizadas por erros ortográficos que incluíram transposições, omissões, substituições e adições de letras. Com base nestes resultados, atribuíram tais equívocos ortográficos a uma interrupção no processo de conversão entre grafemas e fonemas – isto é, uma falha do processamento fonológico.

Affonso et al. (2011) avaliaram o padrão de resposta de crianças disléxicas para uma tarefa de nomeação de figuras por meio da escrita, considerando os tipos de erros ortográficos cometidos pelos participantes. No estudo, foram comparados os textos de crianças com dislexia (grupo dislexia), aos textos de dois grupos controles: o primeiro pareou os participantes por sexo, idade e tipo de escola, selecionando crianças com desempenho de leitura dentro da média esperada, com base na média total da série cursada; o segundo pareou os participantes por sexo, tipo de escola e nível de leitura, selecionando, assim, leitores mais novos. Os resultados revelaram uma semelhança entre o desempenho do grupo de participantes com dislexia e o do grupo controle pareado para o mesmo nível de fluência leitora, de modo que ambos os grupos não diferiram quanto ao número de acertos. No entanto, ambos acertaram menos que o grupo controle. Entre os disléxicos, houve mais erros associados à correspondência unívoca grafema-fonema, a omissão de segmentos e à correspondência fonema-grafema independente de regras.

A análise dos erros ortográficos, assim, mostrou-se pertinente para a compreensão das estratégias utilizadas pelas crianças e dos processos linguísticos subjacentes às dificuldades de escrita em indivíduos com dislexia.

Dias e Ávila (2008) avaliaram o modo como crianças registram, em suas produções escritas, a correspondência entre grafemas e fonemas e como identificam e analisam os erros que cometem na escrita. Para tanto, analisaram e compararam os equívocos produzidos por 56 escolares do 1º ao 4º ano do ensino fundamental de uma escola particular, ao realizarem um ditado de palavras de alta e baixa frequência e pseudopalavras. Os participantes foram divididos entre um grupo clínico, com dislexia, e um grupo controle, com crianças sem queixas ou alterações relativas à compreensão leitora, pareados por idade, sexo e série com o grupo clínico. As observações e justificativas das crianças a respeito dos erros e acertos de escrita que produziram foram gravadas e comparadas. Como esperado, os escolares do grupo pesquisa mostraram pior desempenho quando comparados com as crianças com transtorno de aprendizagem. Houve uma significativa diminuição dos erros com a progressão da escolaridade, sobretudo na escrita de palavras de alta frequência. Os escolares do grupo clínico demonstraram maior dificuldade em analisar os itens linguísticos escritos e reconhecer erros e acertos, o que indica que o transtorno específico de leitura favoreceu uma maior dificuldade em escrever e analisar a escrita, produzida em ortografia transparente. Nesse grupo de escolares, os erros diminuíram com a progressão das séries e variaram conforme o item linguístico em sua produção e identificação – quanto mais frequente, mais fácil foi a percepção do erro.

Dadas as considerações realizadas e a revisão de literatura empreendida, conclui-se que a exploração científica acerca da produção escrita de disléxicos tem se desdobrado, essencialmente, em investigações qualitativas a respeito da coesão textual e de aspectos linguísticos associados à ortografia e às classes de palavras empregadas, por exemplo. Estudos de caráter quantitativo focados em mensurar aspectos estruturais dos textos deste grupo clínico não são comuns. Por esta razão, discutir ferramentas, teorias e campos do saber complementares à Psicolinguística, capazes de desenvolver, juntos, estudos neste nicho investigativo, mostra-se pertinente. Um exemplo de tais campos do saber é a Ciência da Computação, sobre a qual se discorre a seguir.

1. 4 CIÊNCIA DA COMPUTAÇÃO E LINGUÍSTICA: UMA INTERFACE PRODUTIVA E PERTINENTE

A investigação de fenômenos relacionados à linguagem a partir do território da Linguística pode suscitar variadas questões de pesquisa, tantas quantas são os contatos que a área estabelece com outros campos do saber. O diálogo entre a Ciência da Computação e a Linguística, por exemplo, tem se mostrado fértil neste sentido (FERRARI, 2004) e favorecido o desenvolvimento da vertente de estudos linguísticos conhecida como Linguística Computacional (LC), comumente associada ao campo Processamento da Linguagem Natural (PNL). Na literatura, não há consenso quanto às definições, diferenças e propósitos caracterizam ambos os campos. Adotam-se, portanto, as perspectivas a seguir, a fim de construir um breve e particular panorama de como os campos da Inteligência Artificial, do Aprendizado de Máquina, do Processamento de Linguagem Natural e da Linguística Computacional se aproximam da análise textual aqui proposta.

Basicamente, a Linguística Computacional consiste em uma abordagem computacional da linguagem cuja finalidade é identificar meios de dotar máquinas de linguagem não-natural (PARDELLI; BIAGIONI, 2013). Trata-se de uma disciplina híbrida, que mobiliza interesses e conhecimentos da Linguística e da Ciência da Computação (OTHERO, 2006) relacionados ao desenvolvimento de sistemas com capacidade de reconhecer e produzir informação apresentada em linguagem verbal - oral ou escrita (VIEIRA; LIMA, 2001). Para tanto, descrevem, explicam e propõem formalizações de aspectos lógico-formais da linguagem humana, de modo a instruir uma máquina a compreender e gerar línguas naturais. Corretores ortográficos e gramaticais, tradutores automáticos e *parsers*¹¹ são alguns exemplos de produtos que se utilizam dos estudos em Linguística Computacional.

Para Othero (2006), trabalhos assentados nos terrenos da Linguística de Corpus e do Processamento de Linguagem Natural (PLN), geralmente, confluem para o que se entende por Linguística Computacional. Segundo tal perspectiva, trabalhar com Linguística de Corpus é dedicar-se à exploração da linguagem a partir de evidências empíricas, reunidas, manipuladas e exploradas por meio de computador (SARDINHA, 2000). É sabido que a noção de corpora

¹¹ Um *parser* consiste em um analisador sintático cuja função é processar uma entrada de dados e dela extrair regras sintáticas específicas. Uma vez identificadas estas regras, o parser monta uma estrutura de como é a composição da entrada de dados em questão.

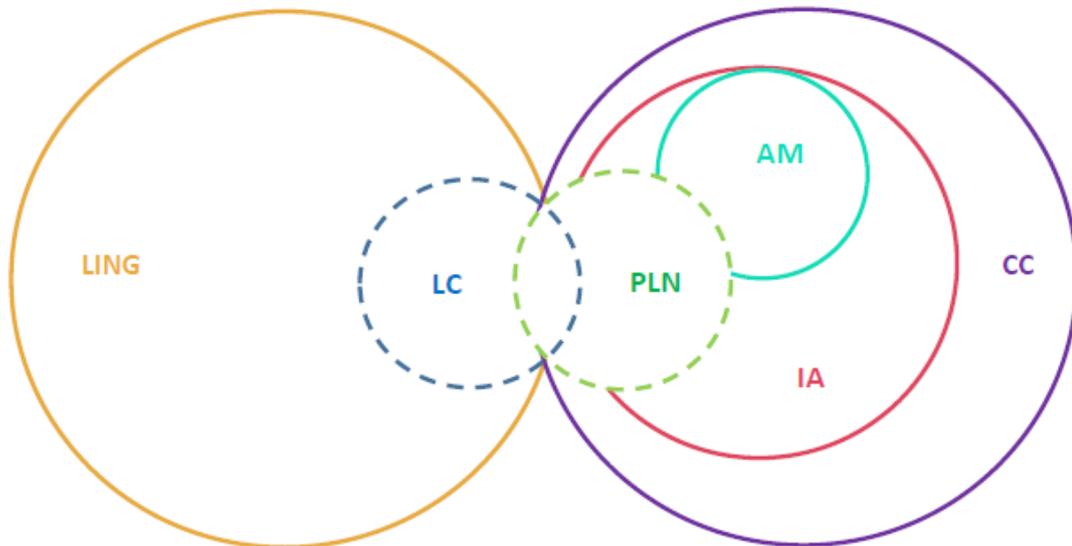
(ou de conjunto de corpus, que são agrupamentos de documentos) é anterior à criação do computador, se pensarmos que, por exemplo, Alexandre, o Grande definiu o Corpus Helenístico na Grécia Antiga, assim como as citações da Bíblia motivaram a criação de corpora na Antiguidade e na Idade Média (ibid.). No entanto, o computador viabilizou compilações e armazenamentos eletrônicos de conjuntos de dados de caráter linguístico, reunidos com o propósito de representarem uma variedade linguística para estudo. Isso porque um corpus linguístico de base computacional consiste em coleções de textos obtidos de situações de uso natural da língua, organizadas sistematicamente, de modo a servirem para a extração de novas informações acerca da língua que representam (BIBER, 1995). Como resultado, linguistas passaram a contar com circunstâncias de consulta e exploração de dados textuais empíricos automatizadas, mais econômicas e inteligentes, o que confirma a relevância atribuída à interface reportada entre Linguística e Ciência da Computação.

O Processamento de Linguagem Natural, contudo, tem sido associado tanto a esta condição de subárea da Linguística Computacional quanto à de linha de pesquisa da Inteligência Artificial (IA) (OTHERO, 2006). O interesse de cientistas da computação em formalizar e automatizar competências comumente associadas à natureza humana – como a capacidade de se comunicar por meio da linguagem verbal -, com frequência, resulta em trabalhos no campo da Inteligência Artificial (BELLMAN, 1978). Assim, estes profissionais estudam e desenvolvem meios de reproduzir faculdades mentais a partir da utilização de modelos computacionais (RUSSELL; NORVIG; INTELLIGENCE, 1995), capazes de representar e prever comportamentos humanos. No entanto, muitos estudos em IA investigam questões da ordem da linguagem verbal sem, necessariamente, reproduzir tarefas desempenhadas por pessoas ou animais, uma vez que métodos e estratégias de resolução de problemas que envolvem capacidades de computação altíssimas, superiores às da natureza humana, já são possíveis de serem desempenhados por máquinas (MCCARTHY, 2007). A exemplo de tais métodos, a aprendizagem profunda¹² viabilizou modelos computacionais que aprendem representações de dados com múltiplos níveis de abstração, o que tem tido grande êxito no reconhecimento de fala, por exemplo (LECUN; BENGIO; HINTON, 2015). Logo, é plausível resumir o propósito da Inteligência Artificial ao de instrumentalizar máquinas para que solucionem problemas apresentados pelo mundo, sejam eles compreendidos a partir da observação do comportamento de pessoas ou não.

¹² Comumente referida em inglês como *Deep learning*.

O esquema a seguir elucida o encontro interdisciplinar entre os campos mencionados:

Figura 8: Relações interdisciplinares



Fonte: A autora

Basicamente, o presente estudo assume Linguística Computacional (LC) e Processamento da Linguagem Natural (PLN) como campos porosos, que se invadem ao compartilharem interesses. Enquanto a LC situa-se, essencialmente, no território da Linguística (LING), o PLN insere-se, sobretudo, no domínio da Inteligência Artificial (IA), serve-se do Aprendizado de Máquina (AM) e, assim, concentra seus estudos, majoritariamente, no território da Ciência da Computação (CC).

Definir o PLN como um campo comum à Inteligência Artificial implica concebê-lo como uma tentativa de criar mecanismos de interpretação e geração de línguas naturais, como o português e o inglês, quer seja na modalidade oral ou escrita (MONARD; BARANAUSKAS, 2003). Mais do que isso, implica afirmá-lo como uma das muitas aplicações com as quais trabalha o *Machine Learning* (ML) – ou Aprendizado de Máquina, discutido a seguir.

1.4.1 Aprendizado de Máquina aplicado a dados linguísticos: um encontro em evidência

O Aprendizado de Máquina é um subcampo da Ciência da Computação que se desenvolveu a partir do reconhecimento de padrões e da inteligência artificial. Fundamentalmente, ele opera a partir de sistemas capazes de melhorar o seu desempenho para tarefas específicas a partir do processamento de dados. Tais sistemas são conhecidos como

classificadores, e são subjacentes ao desempenho da máquina diante da realização de uma tarefa – como filtrar spams em contas de e-mail, reconhecer imagens ou analisar sentimentos a partir de textos. A realização desta tarefa é resultante da obtenção de conclusões genéricas a partir de um conjunto de exemplos; um conceito é aprendido a partir de inferências indutivas acerca de um pacote de dados. Quando exposta a novos dados de mesmas características, a máquina ou o classificador busca generalizar o conhecimento adquirido previamente para a nova amostra. Em resumo, classificadores aprendem por meio de atributos de entrada a produzirem decisões e resultados plausíveis – estendendo padrões observados e armazenados previamente.

O aprendizado indutivo promovido pelo desenvolvimento de classificadores pode ser dividido entre o dito supervisionado e o não-supervisionado (RUSSELL; NORVIG; INTELLIGENCE, 1995). O primeiro resolve problemas preditivos, uma vez que trabalha com rótulos de classes discretos, isto é, com classes definidas previamente, entre as quais já se sabe que os objetos de análise do estudo se distribuem – bons e maus leitores, por exemplo. Já o segundo busca distribuir ou agrupar a amostra conforme semelhanças percebidas entre elas, uma vez que não há classes pré-definidas. Resumidamente, aprendizados supervisionados são da ordem da classificação, enquanto que os não-supervisionados objetivam a descrição.

Além disso, é possível identificar diferentes paradigmas utilizados em estudos que se servem do Aprendizado de Máquina, que atendem a diferentes propósitos. Destaca-se, a seguir, três deles:

Tabela 1: Exemplos de paradigmas de Aprendizado de Máquina

Paradigma	Descrição	Exemplo
Simbólico	Representações simbólicas de um problema por meio da análise de exemplos e contra-exemplos	Algoritmos de árvore de decisão como ID3
Estatístico	Modelos estatísticos que oferecem uma referência para encontrar aproximações do conceito induzido	Máquina de Vetores de Suporte ¹³ (<i>Support Vector Machines</i> - SVM) e Aprendizado Bayesiano
Conexionista	Inspirada no modelo biológico do sistema nervoso	Redes Neurais

¹³ Será dada preferência à sigla em inglês a partir de agora, em detrimento da sua versão traduzida para o português, a fim de conservar a terminologia tipicamente utilizada nos trabalhos da Ciência da Computação.

Fonte: Adaptado de Monard e Baranauskas (2003)

No âmbito dos estudos em Linguística ou daqueles em Ciência da Computação que se dedicam a classificar objetos de caráter textual, alguns métodos típicos de classificação têm se mostrado especialmente eficientes e pertinentes, destacando-se o SVM e a Classificação Bayesiana (HAN; PEI; AMBER, 2011). Tais métodos se utilizam de algoritmos como o SVM e o Naïve Bayes respectivamente. O primeiro adota, como entrada para o aprendizado da máquina, um conjunto de dados e prediz à qual de duas possíveis classes cada entrada pertence. Ele representa as classes em um espaço, no qual os dados de entrada são distribuídos como pontos no espaço, de acordo com suas classes. Trata-se de um classificador linear binário não probabilístico. O segundo, por sua vez, consiste em um classificador probabilístico baseado no Teorema de Bayes, que confere alta independência aos seus atributos de entrada ao predizer a probabilidade de um dado pertencer a uma determinada classe a partir de uma pressuposição “ingênua” de que as classes dos dados são mutuamente exclusivas, independentes e representam uma enumeração exaustiva de todas as possibilidades.

Uma vez discutidos os métodos e algoritmos comumente adotados em estudos científicos focados na classificação de textos a partir do Aprendizado de Máquina, resta discutir seus contextos de pesquisa, a partir de exemplos da literatura. Recentemente, pesquisadores da Universidade Federal do Rio Grande do Norte criaram um software, baseado na Teoria dos Grafos e denominado *Speech Graphs*¹⁴, capaz de representar as relações entre palavras de discursos por meio de grafos, e assim revelar padrões característicos da conectividade discursiva de diferentes grupos ou enunciadorees (MOTA et al., 2014). Tais padrões são identificados a partir da distribuição dos itens lexicais no texto, das relações de adjacência que estabelecem entre si e de medidas quantitativas que caracterizam os grafos, também chamadas de atributos, como a quantidade de palavras que possuem os textos por eles representados. Uma vez quantificados os atributos, Mota e colegas descobriram que é possível ensinar uma máquina a processar fragmentos de linguagem natural, extrair características recorrentes destes trechos, que constituam seus padrões de conectividade discursiva, e, então, distribuir outros textos entre as categorias previamente identificadas, a partir de suas características quantitativas.

Logo, observa-se que Linguística, Ciência da Computação e Teoria dos Grafos (TG), quando associadas, podem ser produtoras no âmbito de investigações acerca da linguagem.

¹⁴ Disponível para download em: <http://www.neuro.ufrn.br/software/speechgraphs> .

1.4.2 A Teoria dos Grafos como ferramenta de análise textual

A Teoria dos Grafos consiste em uma ramificação da matemática que se ocupa da relação entre elementos de um conjunto (WILSON, 1996); costuma ser adotada por trabalhos que buscam descrever a distribuição de informações entre regiões, elementos ou pontos quaisquer inter-relacionados dentro de uma rede, na qual transita informação entre esses pontos, por meio das conexões que os ligam (ROCHA, 2015; TRUDEAU, 1976).

Em 1736, o matemático suíço Leonhard Euler iniciou seus estudos sobre grafos ao solucionar o enigma popularmente conhecido como O Problema das Pontes de Königsberg. Cidade da antiga Prússia, Königsberg possuía duas ilhas que, juntas com a parte continental da cidade, eram ligadas por sete pontes. Discutia-se, na cidade, a possibilidade de atravessar todas as pontes, sem repeti-las. Euler, em 1736, encontrou a solução para esta charada ao transformar os caminhos das pontes em retas e suas intersecções em pontos, criando, assim, um grafo (ALEXANDERSON, 2006). Desde então, a teoria tem surpreendido matemáticos pelos resultados teóricos de especial variedade e profundidade (ORE, 1990), obtidos através de trabalhos que se servem de representações de grafos. A aplicabilidade da Teoria dos Grafos em estudos científicos tem se mostrado ampla e diversificada, e isso se deve ao seu caráter matemático (relevante para as generalizações próprias da ciência) e a sua adaptabilidade e correspondência às diferentes esferas dos saberes do mundo, que se constroem a partir das relações entre seus componentes.

Questões de pesquisa variadas relacionadas à linguagem já se serviram de modelos de redes complexas oriundas da TG. I Cancho & Solé (2001) investigaram a correspondência entre a coocorrência de palavras em uma mesma frase e a organização lexical subjacente à superfície linguística do enunciado. Costa et al. (2009) examinaram a associação livre de palavras em um jogo online, com o intuito de verificar se essas consequências poderiam ser explicadas em termos de trajetórias em um grafo semântico. Hills et al. (2009) estudaram o fenômeno denominado de estrutura associativa da linguagem e investigaram se a aprendizagem precoce de palavras pode ser favorecida pela diversidade linguística do ambiente de aprendizagem. A literatura ainda inclui análises de conexões neurais (VITEVITCH, 2008) e conexões de discurso ou textuais baseadas na Teoria dos Grafos (MOTA, 2013; BERTOLA et al., 2014; MOTA et al., 2016). A razão disso decorre do modo como as regiões do cérebro interagem para que haja o processamento de informações, que é semelhante às topologias de rede descritas pela Matemática como grafos - o que justifica a utilização recorrente da TG em neurociências. O

mesmo se pode dizer de estudos sobre padrões de conexões discursivas ou textuais, que se servem da Teoria dos Grafos para ilustrarem as relações estabelecidas entre os elementos que integram um conjunto em forma de texto ou discurso.

O mesmo se pode dizer de trabalhos sobre padrões de conexões discursivas - ou textuais -, que se servem da Teoria dos Grafos para ilustrarem as relações estabelecidas entre os elementos que integram um conjunto em forma de texto ou discurso. Três estudos revisados a seguir, especificamente, ofereceram evidências de que análises discursivas quantitativas podem ser ferramentas pertinentes em contextos clínicos e escolares.

Mota (2013) reporta o uso da TG para evidenciar déficits cognitivos em pacientes com psicoses. Para tanto, desenvolveu o software *Speech Graphs*, a fim de obter representações gráficas de discursos que viabilizassem o levantamento de dados expressivos de diferenças de conectividade textual entre relatos verbais de pacientes com psicose e indivíduos normais. Considerando que o diagnóstico diferencial da psicose é baseado em avaliações subjetivas de profissionais especializados, uma avaliação quantitativa do discurso dos sujeitos de pesquisa poderia ser complementar ao diagnóstico em questão. Especificamente, a hipótese que orientou o estudo foi a de que a quantificação dos atributos de grafo dos diferentes grupos revelaria diferenças significativas entre eles, possíveis de serem mensuradas, categorizadas e quantificadas. A fim de se obter estes dados quantitativos acerca da psicose, entrevistas gravadas com os dois grupos clínicos e com o grupo controle foram realizadas, nas quais eles produziram relatos de sonho e de vigília. Essencialmente, os resultados indicaram que os relatos de sonhos dos pacientes psicóticos têm menos conectividade entre seus itens lexicais que relatos do grupo controle. As medidas de conectividade utilizadas tiveram correlação negativa com sintomas cognitivos negativos, sugerindo assim que quanto mais prejudicados cognitivamente os sujeitos de pesquisa, menor a conectividade de seus relatos de sonho. Posteriormente, a utilização de classificadores binários baseados nos atributos de grafos distinguiu esquizofrênicos de maníacos com até 93,8% de sensibilidade e 93,7% de especificidade. Quando contrastados tais resultados com os escores de suas escalas psiquiátricas padrão (BPRS e PANSS), notou-se que o potencial de classificação da escala – 62,5% de sensibilidade e especificidade - ficou aquém daquele observado no classificador. Conclui-se, portanto, que os fluxos de pensamento normal e disfuncional são retratados matematicamente no discurso dos sujeitos de pesquisa.

.Seguindo a mesma linha de estudos acerca da conectividade de itens discursivos, Bertola et. al (2014) fizeram uso da TG aplicada a relatos de memória para quantificar déficits cognitivos patológicos associados à demência. No caso da demência, a análise teórico-gráfica

do teste de fluência verbal levou a uma boa classificação entre pacientes com doença de Alzheimer e aqueles com déficits cognitivos leves. A deficiência cognitiva foi acompanhada pelo aumento da densidade do gráfico, diminuição do diâmetro e menor caminho médio mais curto. Basicamente, pessoas com doença de Alzheimer demonstraram produzir mais L3 que sujeitos controles em testes de fluência verbal que solicitam que sejam nomeados o máximo possível de animais diferentes dentro de um minuto. Menos recorrências de palavras foram associadas a um melhor e maior desenvolvimento da memória de trabalho¹⁵, uma vez que prejuízos na memória de trabalho levam a prejuízos cognitivos que podem interferir no desempenho acadêmico. Houve também uma correlação negativa entre as medidas cognitivas QI e Teoria da Mente e o atributo de recorrência PE, resultado que reforçou a ideia de que o desenvolvimento da memória de trabalho contribui para o bom desempenho acadêmico.

Mota et al. (2016) realizaram um estudo longitudinal pioneiro acerca da relação memória e desempenho cognitivo e acadêmico. Especificamente, investigaram a relação entre memória episódica¹⁶ e performance acadêmica no ensino fundamental, com 76 crianças entre 6 e 8 anos, em um contexto naturalístico, diferente dos tradicionais testes de rememoração de estímulos apresentados em computador. Tais crianças relataram memórias de eventos autobiográficos e memórias de imagens inéditas imediatamente após codificarem-nas. A hipótese que orientou o trabalho foi a de que atributos de grafo relacionados à conectividade diminuem quando associados com o declínio cognitivo; assim, eles aumentariam com o desenvolvimento na alfabetização. Além disso, QI e Teoria da Mente¹⁷ foram avaliados e apresentaram correlação positiva com a diversidade e a conectividade de palavras, e negativa com a recorrência de palavras. Em resumo, as crianças com relatos com maior número de palavras diferentes, com maior conexão entre elas e com menor repetição de adjacência entre itens lexicais tiveram melhor desempenho para QI e para Teoria da Mente.

A fim de contribuir para o avanço de investigações sobre medidas de grafo de textos produzidos por populações saudáveis e clínicas, o presente estudo se insere nesta corrente de

¹⁵ A memória de trabalho é também definida como memória operacional e é responsável pelo armazenamento e pela manipulação de informações por curto prazo de tempo, a fim de que tarefas cognitivas complexas possam ser executadas (LENT, 2010).

¹⁶ Memória que compreende lembranças de coisas e eventos relacionados a um tempo ou lugar em particular, portanto, a informações com contextos espaciais e temporais específicos (LENT, 2010).

¹⁷ É denominada Teoria da Mente a capacidade cognitiva de atribuir estados mentais a si e aos outros. A interação interpessoal, própria das comunidades linguísticas, é mediada por pensamentos e crenças acerca de si e dos outros nem sempre verbalizados, atingidos graças à Teoria da Mente (FLETCHER et al., 1995).

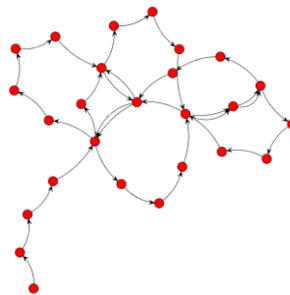
trabalhos que se servem da Teoria dos Grafos e, especificamente, do software *Speech Graphs*, para quantificar características textuais associadas a habilidades acadêmicas.

1.4.3 Definições e atributos de grafo

A compreensão da Teoria dos Grafos requer o domínio de sua linguagem gráfica de representação, bem como da nomenclatura referente a ela. A linguagem adotada pela literatura para reconstruir a TG não é consensual, de modo que, para garantir a consistência do presente trabalho, adota-se a terminologia empregada por Wilson (1996).

Grafo (G) é o nome dado à representação de um conjunto de elementos ligados entre si, para quaisquer que sejam as propriedades métricas (ibid.). Vejamos o exemplo:

Figura 9: Grafo



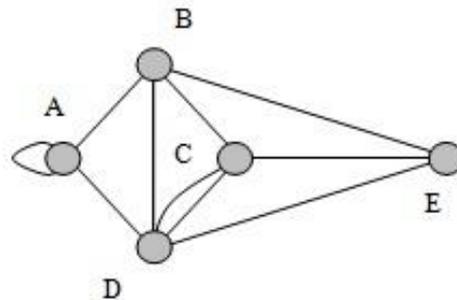
Fonte: Autora

Os **nós (N)** do grafo correspondem às bolinhas vermelhas e representam elementos de um mesmo conjunto; já os traços que ligam os nós são as **arestas (E)** do grafo e representam as relações entre os nós. Concebendo a representação dessa maneira, o grafo acima poderia ser a caracterização, por exemplo, das relações de amizade entre quatro jovens, da compatibilidade entre quatro dispositivos ou dos itens lexicais de um texto – considerando, sempre, a adjacência de cada um dos nós. Assim, é possível mensurar o grau de conectividade de um jovem, de um dispositivo ou das palavras que compõem um texto a partir do número total de arestas que chegam ou partem do nó que a(o) representa.

Logo, em linhas gerais, um grafo G que representa a relação de diferentes elementos entre si, integrantes de um conjunto, pode ser representado por $G(N,E)$, sendo N o conjunto de nodos (que simbolizam os elementos do conjunto) e E , o conjunto de arestas (representativas das relações dos elementos do conjunto).

Tendo isso em mente, considera-se o seguinte grafo:

Figura 10: Grafo com ciclos

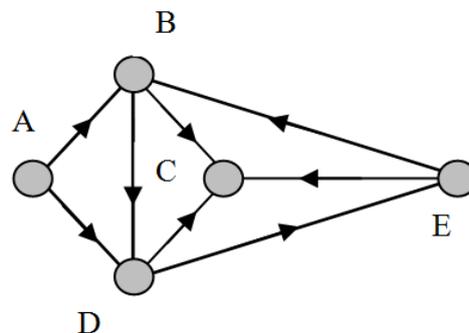


Fonte: Autora

A representação acima poderia simbolizar as vias ligando cinco pontos turísticos de uma cidade. O ponto A teria um acesso ao ponto B, um ao D e possuiria uma via de retorno para si mesmo, considerando o ciclo de um nó retratado no grafo. O ponto C, tendo em vista o grafo, teria um acesso ao ponto B, um acesso ao ponto E e dois acessos ao ponto D, dado o ciclo de dois nós presente na representação.

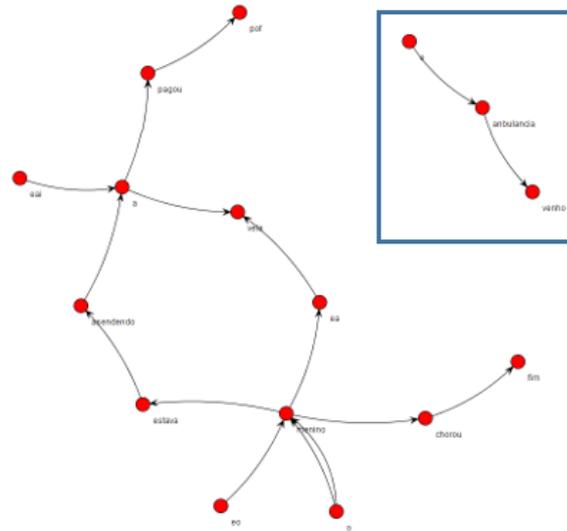
Mas e se a prefeitura da cidade determinasse que esses acessos seriam de mão única? Nesse caso, a representação abaixo, denominada **grafo direcionado**, não faria sentido:

Figura 11: Grafo direcionado



Fonte: Autora

Figura 13: Grafo desconectado do texto 58-2014-O-IE-LB



Fonte: Autora

Aqui, há um grafo $G(16,16)$ com dois componentes, sendo o menor deles formado pelos nós *e*, *anbulancia* e *venho*.

A análise empreendida no presente trabalho se utilizará da TG a fim de identificar e medir padrões de conectividade textual por meio da quantificação de atributos ou características de grafos, capazes de sinalizar quantitativamente as diferenças entre os textos analisados.

Como dito anteriormente, grafos são representações de composição variada. As características de cada composição revelam muito sobre os fenômenos representados pelos grafos, e são chamadas de atributos. Uma análise de grafos, portanto, implica a seleção de atributos pertinentes, que possam ser quantificados e que revelem nuances importantes do padrão de conectividade do fenômeno investigado.

Na tabela a seguir, com base no estudo de Mota et al. (2016), são descritas as doze medidas de grafo que orientaram a análise dos dados coletados, com suas respectivas definições matemáticas e interpretações psicolinguísticas, somadas à contagem de palavras¹⁸. No primeiro segmento da Tabela 2, duas medidas quantitativas expressivas de características gerais de cada grafo são descritas. No segundo, três atributos associados à conectividade de cada texto

¹⁸ À contagem de palavras não é atribuído o status de atributo de grafo. No entanto, trata-se de uma medida extraída pelo software *Speech Graphs* que tem se mostrado, na literatura, expressiva de status cognitivo geral e performance funcional (BERTOLA et al., 2014). Por esta razão, esta medida foi incluída nas análises.

(representado graficamente) são explicitados. Quatro atributos de recorrência, ou seja, medidas referentes à reincidência de itens lexicais no texto, são elencados no terceiro segmento. Por fim, três atributos globais, associados à dimensão total de cada grafo, estão discriminados, no quarto e último segmento da tabela.

Tabela 2: Atributos de grafo (Adaptado de Mota et al. (2016))

ATRIBUTOS GERAIS			
Sigla	Nomenclatura	Definição Matemática	Interpretação Psicolinguística
WC*	Contagem de palavras (<i>Word Count</i>)	Número total de elementos em conexão – contando reiterações de um mesmo elemento.	Expressa o número total de itens lexicais do texto; mede a profusão lexical.
N	Nós (<i>Nodes</i>)	Número total de elementos diferentes em conexão.	Expressa o número de itens lexicais distintos do texto; mede a diversidade lexical.
E	Arestas (<i>Edges</i>)	Número total de ligações entre elementos.	Número de links entre itens lexicais; mede a quantidade de relações que os itens lexicais estabelecem entre si.
ATRIBUTOS DE CONECTIVIDADE			
ATD	Grau Médio Total (<i>Average Total Degree</i>)	Grau médio total de todos os nós, considerando grau total a soma das arestas que chegam e que saem de cada nó.	Expressa a média de relações que os itens lexicais de um texto estabelecem entre si.
LCC	Maior Componente Conectado (<i>Largest Connected Component</i>)	Número total de nós que compreende o maior subgrafo em que cada par de nós estabelece ligação com outro.	Mede quão bem conectadas as palavras do texto estão.
LSC	Maior Componente Fortemente Conectado (<i>Largest Strongly Connected Component</i>)	Número total de nós que compõem o maior subgrafo no qual todos os nós estão ligados mutuamente.	Mede quão bem conectadas as palavras do texto estão.
ATRIBUTOS DE RECORRÊNCIA			
PE	Arestas paralelas (<i>Largest Strongly Connected</i>)	Número total de arestas ligando o mesmo par de nós mais de uma vez.	Mede a recorrência de ligação entre itens lexicais específicos.

<i>Component</i>			
L1	Ciclos de um nó (<i>1N/Loop</i>)	Ciclos de um só nó, nos quais o nó de saída e o de chegada são o mesmo.	Mede a reiteração imediata de itens lexicais específicos.
L2	Ciclos de dois nós (<i>2N/Loop</i>)	Ciclos de dois nós, nos quais o nó A se liga ao nó B que retorna ao A.	Mede a recorrência de ligação entre itens lexicais específicos.
L3	Ciclos de três nós (<i>3N/Loop</i>)	Ciclos de três nós, nos quais o nó A se liga ao nó B, que se liga ao nó C, que retorna ao nó A.	Mede a recorrência de ligação entre itens lexicais específicos.
ATRIBUTOS GLOBAIS DE GRAFO			
Dens.	Densidade (<i>Density</i>)	Razão do total de arestas do grafo sobre o total de arestas possíveis ($D = 2 \times E / N \times (N - 1)$)	Mede a espessura dos tecidos textuais, que acaba por determinar quão “maleáveis” eles são - no sentido de registrar quão condensados os nós se encontram.
Diam.	Diâmetro (<i>Diameter</i>)	Distância do maior caminho mais curto entre os pares de nós de uma rede	Mede a extensão dos tecidos textuais – quão estreito é o arranjo dos nós.
ASP	Menor caminho médio (<i>Average Shortest Path</i>)	Distância média do caminho mais curto entre os pares de nós de uma rede	Mede a extensão média dos tecidos textuais – quão estreito é o arranjo dos nós.

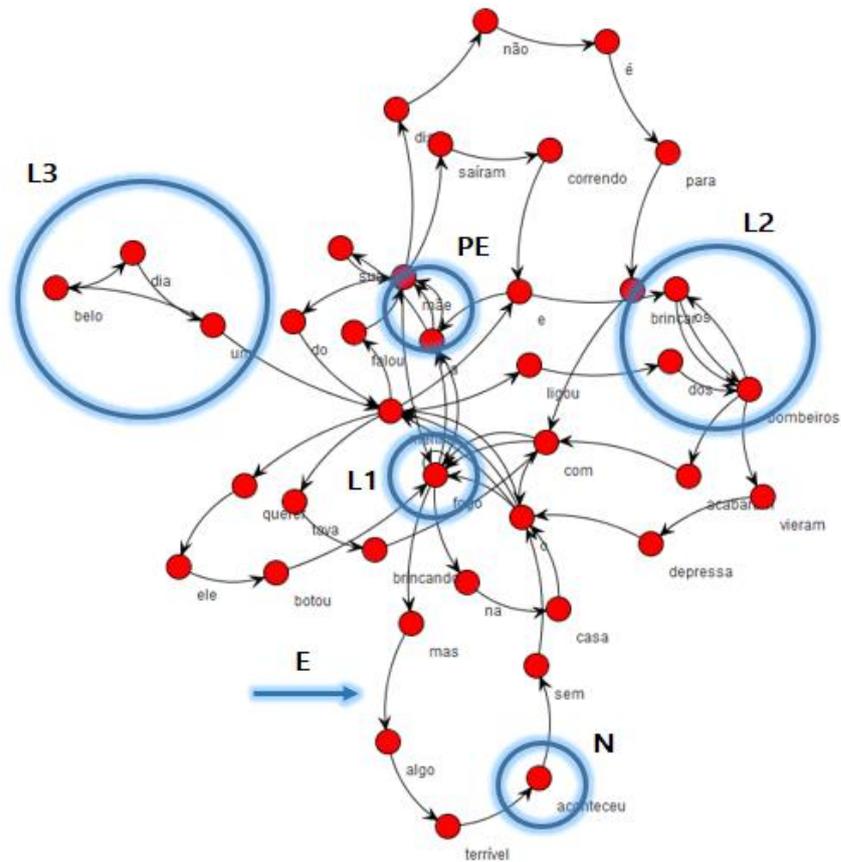
A fim de facilitar a compreensão das medidas que orientam a presente análise textual, a tabela abaixo reúne algumas evidências advindas da literatura acerca de como alguns atributos de grafo registram prejuízos cognitivos.

Tabela 3 – Atributos de grafo na literatura

ATRIBUTOS GERAIS	
Atributo	Evidências encontradas na literatura
WC*	Correlação positiva com status cognitivo geral e performance funcional (BERTOLA et al., 2014); Irrelevância na comparação entre adultos disléxicos e adultos com boa fluência de leitura (TOPS et al., 2013);
N	Correlação positiva com performance cognitiva (QI e Teoria da Mente) e LCC (MOTA et al., 2016); Correlação positiva com status cognitivo geral e performance funcional (BERTOLA et al., 2014);
E	Correlação positiva com status cognitivo geral e performance funcional (BERTOLA et al., 2014);
ATRIBUTOS DE CONECTIVIDADE	
ATD	Inexpressivo de diferença de status cognitivo geral e performance funcional entre grupos de idosos (ao lado de CC E LSC) (BERTOLA et al., 2014);
LCC	Correlação negativa com psicoses e, especificamente, com sintomas negativos (ex.: pouco contato visual, retração emocional e isolamento social)(MOTA et al., 2014); Correlação negativa com déficits cognitivos – especificamente, com dificuldade de entender conceitos abstratos (MOTA et al., 2014); Correlação positiva com performance de leitura (MOTA et al., 2016);
LSC	Correlação positiva com performance de leitura (MOTA et al., 2016);
ATRIBUTOS DE RECORRÊNCIA	
PE	Correlação negativa com performance cognitiva (QI e Teoria da Mente) (MOTA et al., 2016);
L3	Correlação negativa com a memória de trabalho em grupos de idosos (BERTOLA et al., 2014); Correlação negativa com performance cognitiva (QI) (MOTA et al., 2016);
ATRIBUTOS GLOBAIS DE GRAFO	
Dens.	Correlação negativa com status cognitivo geral e performance funcional entre grupos de idosos (BERTOLA et al., 2014);
Diam.	Correlação positiva com status cognitivo geral e performance funcional (BERTOLA et al., 2014);
ASP	Correlação positiva com status cognitivo geral e performance funcional (BERTOLA et al., 2014);

Grafos e fragmentos de grafos¹⁹ de textos da amostra, que contém atributos gerais, de conectividade e de recorrência listados acima, são apresentados a seguir:

Figura 14: Atributos visualizados no grafo



Fonte: A autora

¹⁹ Atributos globais, assim como ATD (de conectividade), por serem mais abstrados – médias, densidade - e abrangentes – diâmetro – não foram incluídos nos exemplos.

2 MÉTODOS

Neste capítulo, são pormenorizadas informações a respeito dos participantes recrutados para o estudo, dos instrumentos utilizados na coleta de dados, dos procedimentos que viabilizaram e conduziram a obtenção dos dados, da análise estatística empregada e das técnicas de aprendizado de máquina adotadas.

Com o intuito de tornar mais nítidas as motivações subjacentes à escolha dos métodos, são retomados aqui os objetivos e hipóteses desta investigação, mencionados previamente na introdução deste trabalho:

▪ **Objetivo**

Identificar padrões de conectividade textual recorrentes entre crianças e adolescentes com diferentes níveis de fluência de leitura – em especial disléxicas - e escolaridades, a partir da análise de grafos representativos de suas produções textuais e de técnicas de aprendizado de máquina.

Especificamente, este trabalho foi orientado pela expectativa de compreender:

- Se o tipo de transcrição textual impactaria as medidas de grafo extraídas dos textos, uma vez que o software utilizado para a extração dos atributos é sensível à quebra de linha e a correções na ortografia, componente textual expressivo de limitações impostas pela dislexia (BASSO et al., 2017);
- Se a normalização das medidas de grafo pelo número de palavras de cada texto seria igualmente distintiva de categorias de fluência de leitura e anos de escolaridade quanto as transcrições não normalizadas, uma vez que esta medida eliminaria o número de palavras de um texto, que consiste em um aspecto textual relevante em análises discursivas (MOTA et al., 2014; BERTOLA et al., 2014);
- Se leitores fluentes, leitores pouco fluentes e crianças diagnosticadas com dislexia apresentariam padrões específicos de conectividade textual, haja vista as correspondências, previstas na literatura, entre a compreensão e a produção escrita (CAPOVILLA; CAPOVILLA, 2007; NAVAS; PINTO; DELLISA, 2009; AFFONSO et al., 2011; BASSO et al., 2017);

- Se uma análise longitudinal evidenciaria diferenças significativas entre as produções textuais dos participantes das escolas que produziram textos ao longo dos três anos do ACERTA, de modo a registrar mudanças no desempenho acadêmico decorrentes do avanço escolar (CAPOVILLA; CAPOVILLA, 2004; CAPOVILLA; CAPOVILLA, 2007; AFFONSO et al., 2011; GÂNDARA; BEFI-LOPES, 2010);
- Se as médias dos atributos de grafo dos textos produzidos em 2014, pelo grupo Escolas, apresentariam proximidade com as dos textos escritos por pacientes do Ambulatório, uma vez que a produção escrita de disléxicos tende a ser substancialmente inferior à esperada para suas idades e escolaridades, e o grupo de disléxicos é bem heterogêneo quanto à idade e escolaridade (DEHAENE, 2012; (SHAYWITZ, 2008);
- Se o potencial preditivo da ferramenta seria, assim, corroborado com o uso de técnicas de aprendizado de máquina, dentro deste contexto de pesquisa.

As hipóteses que orientaram este estudo foram as seguintes:

- O tipo de transcrição impactaria as medidas de grafo extraídas dos textos, uma vez que o software utilizado para a extração dos atributos é sensível à quebra de linha e a correções ortográficas; textos corrigidos seriam processados e categorizados com mais acurácia pelo classificador uma vez que “normalizam” representações semânticas do discurso ao eliminarem erros ortográficos, que são “ruído” em análises discursivas;
- A normalização das medidas de grafo pelo número de palavras de cada texto, se produzisse atributos com diferença significativa entre os grupos experimentais, faria em menor escala que as transcrições não-normalizadas, uma vez que eliminaria o número de palavras que consiste em um aspecto relevante dos textos em análises discursivas (BERTOLA et al., 2014);
- Leitores fluentes, leitores pouco fluentes e crianças diagnosticadas com dislexia apresentariam padrões específicos de conectividade textual, haja vista as correspondências, previstas na literatura, entre a compreensão e a produção escrita (CAPOVILLA; CAPOVILLA, 2007; NAVAS; PINTO; DELLISA, 2009; AFFONSO et al., 2011; BASSO et al., 2017);
- Uma análise longitudinal evidenciaria que as produções textuais dos leitores fluentes das escolas que produziram textos ao longo dos três anos do ACERTA apresentariam

diferenças entre suas medidas de grafo correspondentes aos anos de 2014 e 2016, por conta do avanço da escolarização (CAPOVILLA; CAPOVILLA, 2004; CAPOVILLA; CAPOVILLA, 2007; AFFONSO et al., 2011; GÂNDARA; BEFI-LOPES, 2010);

- As médias dos atributos de grafo dos textos produzidos em 2014, pelos alunos das escolas envolvidas no ACERTA, apresentariam proximidade ou similaridade com as dos textos escritos por pacientes do ambulatório de aprendizagem, uma vez que a produção escrita de disléxicos tende a ser substancialmente inferior à esperada para suas idades e escolaridades, e o grupo de disléxicos é bem heterogêneo quanto à idade e escolaridade (DEHAENE, 2012; SHAYWITZ, 2008);
- O potencial preditivo da ferramenta seria corroborado com o uso de técnicas de aprendizado de máquina, que identificariam, com razoável acurácia, correspondências entre medidas de grafo dos textos e classificações de fluência de leitura prévias, obtidas por meio do diagnóstico dos pacientes do Ambulatório de Aprendizagem do ACERTA de escores de fluência de leitura adquiridos a partir de tarefas comportamentais.

2.1 PARTICIPANTES

A população de estudo do presente trabalho é composta por 181 participantes, todos brasileiros, com idade entre 8 e 16 anos – ao longo dos 4 anos de coleta de dados²⁰. Os participantes são provenientes das escolas participantes do ACERTA e do Ambulatório de Aprendizagem do projeto. Todos assinaram o termo de consentimento informado²¹, aprovado pelo Comitê de Ética da Pontifícia Universidade Católica do Rio Grande do Sul, CAAE número 30895614.5.0000.5336.

A tabela a seguir apresenta a distribuição da população de estudo quanto ao sexo dos participantes, a partir da observação dos pesquisadores:

²⁰ Alguns dados do ambulatório de aprendizagem foram coletados em 2017, estando os participantes dentro desta faixa etária. Os dados coletados nas escolas em 2014 foram produzidos por participantes mais jovens que aqueles coletados em 2016, por exemplo. A turma de jovens que compôs esta população foi, de modo geral, a mesma, com exceção de eventuais saídas e entradas de alunos nas escolas, ou mesmo reprovações. Repetentes foram mantidos no projeto e participaram da coleta de dados, quando permaneceram na escola.

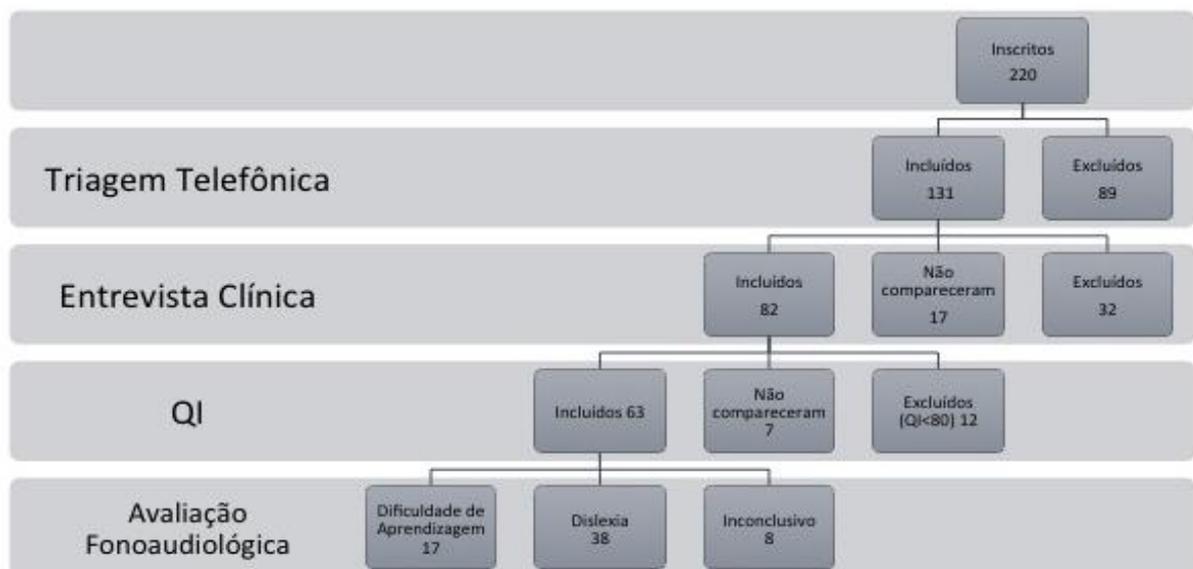
²¹ Disponível em APÊNDICES (APÊNDICE A).

Tabela 3: Características da população geral de estudo

Grupo	Idade média + DP	Subgrupo	Qtde	Total
Escolas	NC ²²	Total Escolas	129	181
		- Meninas	62	
		- Meninos	67	
Ambulatório	10,5 a. (DP = 1,4)	Total Ambu	52	
		- Meninas	16	
		- Meninos	36	

No que diz respeito aos participantes do Ambulatório de Aprendizagem do Projeto ACERTA, cabe mencionar as características amostrais apresentadas por Costa et. al. (2016), que qualificam a população avaliada entre novembro de 2013 e março de 2015, por meio do seguinte fluxograma:

Figura 17: Perfil dos pacientes do Ambulatório



Fonte: Costa et al. (2016)

²² Não foi controlada a variável idade média para o grupo Escolas. Foi critério de inclusão apenas ter entre 8 e 16 anos durante os três anos de coleta de dados nas escolas.

Segundo os autores, da amostra total (N=131), 65 (49,8%) eram do sexo masculino, com média de idade de 10,30 anos (DP = 1,76) e QI médio de 103,30 pontos (DP= 18,06; Mín. = 51 pontos; Máx. 149 pontos); 77 (59%) eram estudantes de escolas públicas, 21 (16%) eram alunos de escolas privadas e 33 (25%) não responderam; 45 (34%) nunca repetiram de ano, 31 (24%) repetiram 1 vez, 17 (13%) repetiram 2 vezes, 7 (5%) repetiram 3 vezes, 1 (0,8%) repetiu 4 vezes, 1 (0,8%) repetiu 5 vezes e 29 (22%) não informaram (COSTA et al., 2016, p. 155).

2.2 RECRUTAMENTO

Os participantes do presente estudo são sujeitos de pesquisa do Projeto ACERTA, oriundos de duas frentes de recrutamento: a primeira frente é composta pelos alunos do ensino fundamental de seis escolas²³ da rede pública de ensino de Porto Alegre (escolhidas em função da proximidade com a região central da cidade), e a segunda frente é integrada por pacientes do Ambulatório de Aprendizagem do Projeto ACERTA. Assim, dados de dois grandes grupos experimentais são analisados, aqui denominados de “Escolas” e “Ambulatório”.

Os critérios que conduziram o recrutamento dos participantes de cada grupo, cujos dados serão utilizados para análise, são descritos a seguir:

Tabela 4: Características e critérios de seleção da amostra

Grupo	Critérios
Ambulatório	<ul style="list-style-type: none"> ▪ Ter entre 8 e 16 anos; ▪ Ter solicitado atendimento no Ambulatório de Aprendizagem por encaminhamento da escola e/ou de outros profissionais, ou ainda através de divulgação nos veículos de imprensa; ▪ Ter sido diagnosticado(a) com transtorno de aprendizagem (dislexia) pela equipe de fonoaudiólogas do ACERTA; ▪ Ter entregue termo de consentimento assinado pelos responsáveis, que autorize o uso dos dados obtidos nas coletas para fins de pesquisa;
Escolas	<ul style="list-style-type: none"> ▪ Ter entre 8 e 16 anos ao longo da coleta de dados;

²³ Ver Introdução.

- Ser aluno(a) de qualquer uma das escolas da rede pública de ensino de Porto Alegre participantes do Projeto ACERTA, matriculado nos anos iniciais do ensino fundamental;
- Ter entregue termo de consentimento assinado pelos responsáveis, que autorize o uso dos dados obtidos nas coletas para fins de pesquisa;

Como está explícito na tabela acima, o fator estrato socioeconômico não foi controlado para ambos os grupos. A participação dos pais dos alunos das escolas – que poderia ser informativa e elucidativa de questões desta natureza - ficou restrita à assinatura do termo de consentimento. No entanto, todos os participantes do ambulatório de aprendizagem do ACERTA – e seus responsáveis – passaram por um protocolo de anamnese²⁴. Neste protocolo, constava um questionário baseado nos critérios ABEP/ABIPEME de classificação socioeconômica, desenvolvido pela Associação Brasileira de Institutos de Pesquisa de Mercado, com o propósito de distribuir a população em categorias segundo padrões ou potenciais de consumo²⁵; o questionário abandona a classificação da população em termos de classes sociais e opta por considerar seu poder de compra. Logo, a divisão definida abaixo é exclusivamente de estratos socioeconômicos. As respostas dos responsáveis pelos participantes ao questionário resultaram na seguinte divisão dos pacientes quanto aos seus estratos socioeconômicos:

Tabela 5: Distribuição dos participantes do ambulatório quanto à classe socioeconômica

Classe	N (pacientes)
A	0
B	14
C	35
D	3
E	0

Quanto ao tipo de escola, foi possível controlar a origem dos participantes das escolas participantes do projeto, uma vez que todas eram da rede pública de ensino de Porto Alegre.

²⁴ O questionário utilizado está disponível em ANEXOS (Anexo A).

²⁵ Esse critério cria uma escala ou classificação socioeconômica por intermédio da atribuição de pesos a um conjunto de itens de conforto doméstico, além do nível de escolaridade do chefe de família. O questionário ABIPEME, parte integrante do protocolo de anamnese, está disponível em ANEXOS (Anexo B).

No entanto, optou-se por não adotar tipo de escola como critério de exclusão prévio no recrutamento de participantes do ambulatório do ACERTA, uma vez que se supunha que a maior parte da amostra seria oriunda de escolas públicas (Tabela 6) e que não haveria diferença significativa entre os padrões de conectividade textual de alunos de escola pública e privada.

A distribuição do grupo Ambulatório entre os tipos de escola foi a seguinte:

Tabela 6: Tipo de escola de origem dos pacientes do Ambulatório

Tipo de escola	N (pacientes)
Pública	42
Privada	10

2.2.1 Procedimentos anteriores ao recrutamento

O critério de escolha empregado na seleção dos participantes do ambulatório de aprendizagem do ACERTA, como explicado acima, foi o diagnóstico dos participantes. Para a presente análise, só foram selecionados textos produzidos por pacientes com diagnóstico de dislexia. O diagnóstico foi conduzido pela equipe de fonoaudiologia do projeto, com base no protocolo apresentado da Tabela 7, e a partir de instrumentos específicos, descritos na Tabela 8:

Tabela 7: Protocolo de diagnóstico do Ambulatório

Etapa	Finalidade
1. Triagem telefônica	Rastrear critérios de exclusão da pesquisa, tais como alteração neurológica prévia ou quadro sindrômico e nível de escolarização inadequado.
2. Anamnese	Com a entrevista médica tradicional (dados de identificação, história médica pregressa e história familiar), é realizada uma revisão da história educacional da criança, a fim de descartar outras possíveis explicações para o baixo desempenho escolar. Há um rastreio para dificuldades de leitura e escrita para verificar se as queixas de aprendizagem são consistentes. Caso não se verifique critérios para exclusão, agenda-se a avaliação do quociente de inteligência estimado (QI) e a fonoaudiológica (bateria de avaliação das habilidades de leitura e escrita).
3. Questionários	Ao término da anamnese, a família recebe questionários autoaplicativos, destinados aos pais e outros aos professores; estes questionários visam rastrear questões econômicas, emocionais, cognitivas e escolares da criança. São eles: <ul style="list-style-type: none"> • <i>Lista de verificação comportamental para crianças/adolescentes de 4-18anos (CBCL)</i> (BORDIN; MARI; CAEIRO, 1995) • <i>SNAP IV</i>-escala validada para o português (MATTOS et al., 2006) • <i>Questionário de Capacidades e Dificuldades (SDQ-Por)</i> • <i>TCK-15</i>-escala de rastreamento das dificuldades de leitura e escrita para professores construída para este ambulatório • <i>Questionário socioeconômico ABIPEME</i> – criado pela Associação Brasileira de Empresas e Pesquisas de Mercado
4. Avaliação do Quociente Inteligência Estimado	Utiliza-se como instrumento de medida a versão reduzida do WISC III para medir o Quociente Intelectual (QI) estimado. Crianças com QI igual ou inferior a 80 são excluídas.
5. Avaliação Fonoaudiológica (tabela 6)	Composta por testes específicos que medem as habilidades de leitura e escrita. Para a avaliação da leitura, são empregadas tarefas que verificam a precisão na leitura de palavras e pseudopalavras, a velocidade ou fluência na leitura de textos e a compreensão. Para as habilidades de escrita, observa-se a qualidade da grafia, a precisão ortográfica e gramatical, a pontuação, além da clareza e da organização da expressão de ideias em forma de texto.

Fonte: Costa et al. (2016)

Tabela 8: Instrumentos utilizados na avaliação fonoaudiológica do Ambulatório

Habilidade	Instrumento	Score obtido
Precisão de leitura de palavras isoladas e pseudopalavras	Salles (2005)	Erros na leitura de palavras e pseudopalavras. Velocidade de leitura.
Fluência de leitura oral e silenciosa	Saraiva, Moojen & Munarski (2007)	Velocidade de leitura.
Compreensão leitora	Saraiva, Moojen & Munarski (2007)	Ideias relatadas espontaneamente e acertos no questionário dirigido.
Precisão ortográfica	Moojen (2009)	Erros e tipo de erros.
Velocidade de escrita	Cópia de texto	Letras por minuto.
Produção textual	Salles (2005)	Valor 0, 1 e 2

Fonte: Costa et al. (2016)

2.3 MATERIAIS E PROCEDIMENTOS

Como fora mencionado anteriormente, os dados comportamentais utilizados no presente estudo são provenientes da tarefa de produção escrita, baseada em Salles (2005), e na tarefa de compreensão leitora, baseada em Saraiva et al. (2006). A tarefa de produção escrita foi realizada nas sessões de coleta de dados do Projeto ACERTA nas escolas - durante o turno regular de estudo das crianças, na sala de aula, aplicada coletivamente - e no Ambulatório de Aprendizagem do ACERTA – aplicada individualmente. A tarefa de compreensão leitora foi aplicada individualmente, tanto nas escolas quanto no Ambulatório.

A seguir, o papel de cada uma dentro deste estudo é pormenorizado.

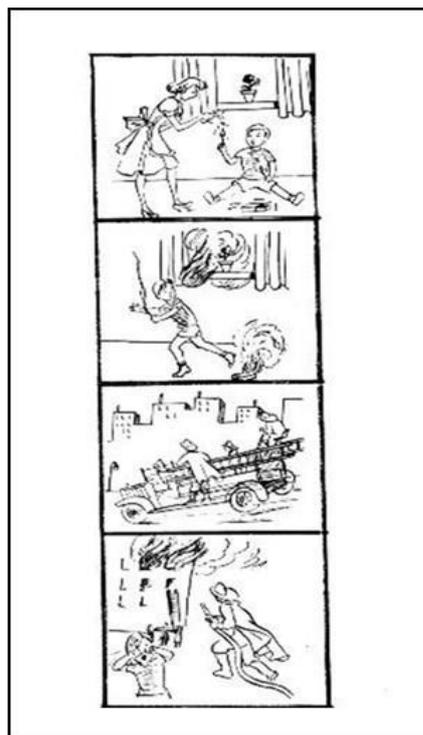
2.3.1 Tarefa de produção textual - (SALLES, 2005)

A análise da tarefa de produção textual empreendida aqui, ao contrário da proposta original, dispensou análises subjetivas e pautou-se, essencialmente, na Teoria dos Grafos, uma vez que pretendeu evidenciar a conectividade textual que caracteriza cada texto. O objetivo da análise textual realizada consistiu em revelar se estes arranjos de itens lexicais são recorrentes entre os participantes de um mesmo grupo experimental e se há, portanto, padrões de conectividade textual associados à fluência de leitura.

Para tanto, uma tirinha composta apenas por imagens que, em sequência, ilustram uma história com início, meio e fim (Figura 18) foi entregue às crianças por uma examinadora. Cada

participante recebeu a tirinha colada em uma folha A4, a fim de que pudesse observar a figura enquanto escrevia a história, evitando distrações. A examinadora, inicialmente, solicitou aos participantes que observassem as imagens por alguns instantes e, depois, escrevessem uma história que reportasse o que viram e entenderam da tirinha, sem restrições de extensão ou quantidade de linhas mínima ou máxima. O tempo não foi cronometrado, de modo a oferecer aos participantes liberdade de tempo para a conclusão da tarefa. A instrução dada a cada criança, como reporta Azevedo (2016, p. 121), foi a seguinte: “Olhe bem para estas figuras. Elas contam uma história. Agora você vai escrever esta história. O tamanho da história é você quem decide. Depois de terminar de escrever a história, você deve ler o que escreveu em voz baixa. Se quiser, pode fazer modificações na escrita durante ou depois de terminar de escrever”.

Figura 18: Modelo da tirinha da tarefa Produção de Texto



Fonte: Salles (2005)

Cada produção textual (Figura 19) foi transcrita em letra maiúscula, sem pontuação e etiquetada conforme o seguinte padrão: número de identificação do participante, ano de coleta, se original (O) ou corrigida (C), letras iniciais da escola e letras iniciais do nome do participante. Além disso, a produção textual de cada participante foi transcrita duas vezes em formato *.txt*,

de modo a se ter duas versões de cada texto: uma versão original, fiel à ortografia e às quebras de linha (Figura 20), e outra corrigida, com erros de ordem ortográfica reparados (Figura 21). O propósito que orientou esta escolha foi o de verificar se a quebra de linha e a falta de convergência de múltiplas versões ortográficas para uma mesma representação semântica²⁶ geram medidas diferentes de grafo (Figuras 22 e 23), capazes de conduzir a análise para resultados distintos. Questões de caráter gramatical – como concordância nominal e verbal, por exemplo – não foram alteradas nas versões corrigidas; apenas erros ortográficos, aglutinações e segmentações foram reparados.

Figura 19: Texto original escrito à mão, pelo participante 62-2014-IE-MD

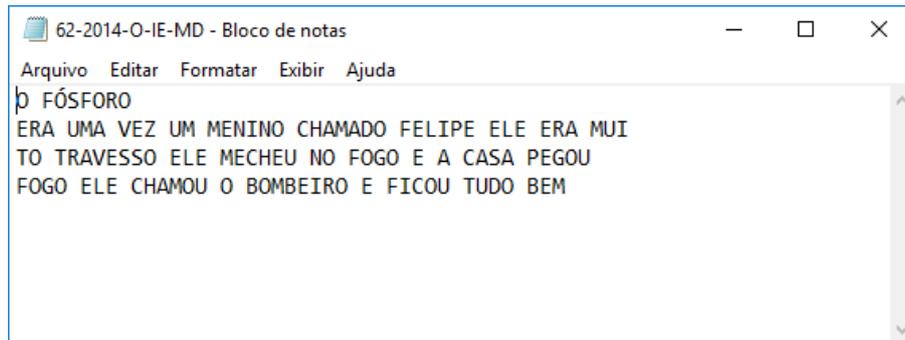
O FÓSFORO

Era uma vez um menino chamado Felipe ele era muito travesso ele mecheu no fogo e a casa pegou fogo ele chamou o bombeiro e ficou tudo bem.

Fonte: A autora

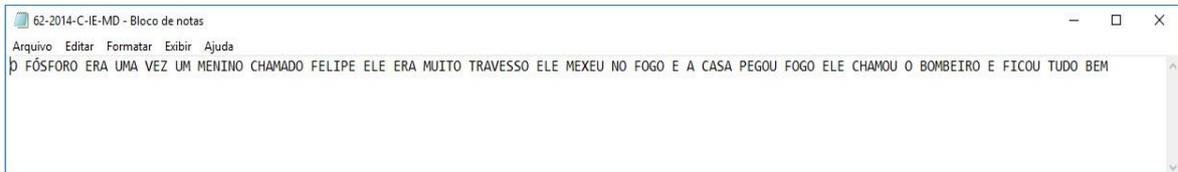
²⁶ Houve especial atenção à ortografia a fim de não comprometer a contagem e identificação de palavras realizada tanto pelo *Speech Graphs* quanto pelos classificadores SVM. Assume-se aqui que “caza” e “casa” convergem para uma mesma representação semântica. Logo, analisar uma versão corrigida, que assegure tal convergência, mostrou-se pertinente. A escolha por transcrever o texto de modo contínuo, sem quebra de linha, justifica-se pela possibilidade de visualizar grafos contínuos do texto, sem subgrafos ou agrupamentos de nós desconectados entre si.

Figura 20: Modelo de transcrição original do texto do participante 62-2014-IE-MD



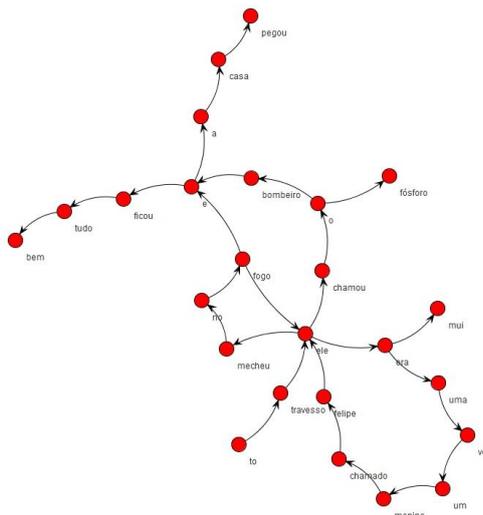
Fonte: A autora

Figura 21: Modelo de transcrição corrigida do texto do participante 62-2014-IE-MD



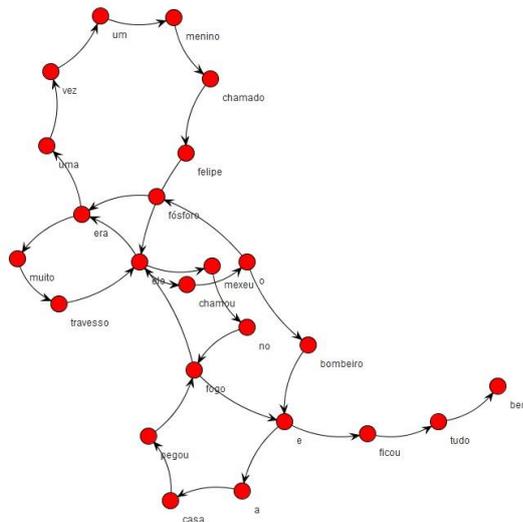
Fonte: A autora

Figura 22: Grafo da versão original do texto do participante 62-2014-IE-MD



Fonte: A autora

Figura 23: Grafo da versão corrigida do texto do participante 62-2014-IE-MD



Fonte: A autora

2.3.2 Avaliação de compreensão leitora de textos expositivos (SARAIVA, R., MOOJEN, S., MUNARSKI, 2007)

A fim de viabilizar a classificação dos participantes das escolas quanto à fluência de leitura - para, então, treinar os classificadores SVM destinados a identificar a classe de fluência de leitura dos autores dos textos analisados -, o desempenho dos participantes na avaliação de compreensão leitora de textos expositivos (SARAIVA, R., MOOJEN, S., MUNARSKI, 2007) foi adotado como critério de classificação. Basicamente, este material apresenta textos expositivos com tipos de organização diferenciados (dois textos para cada ano/série, a partir do 2º ano do ensino fundamental). Tem como propósito auxiliar psicopedagogos e fonoaudiólogos na avaliação da fluência e compreensão leitora de alunos com queixa de dificuldades de aprendizagem e leitura. Todos os textos são acompanhados de ilustrações que contribuem para que conhecimentos prévios dos alunos sejam acionados na leitura. No Projeto ACERTA, três textos²⁷ deste material foram utilizados para obter a velocidade de leitura silenciosa dos participantes (palavras por minuto) e a capacidade de compreensão leitora do indivíduo. Os textos foram escolhidos tendo em vista a idade e o ano do participante.

²⁷ Os três textos utilizados, ao longo dos três anos de coleta, estão disponíveis em ANEXOS (Anexo C).

A fim de se obter a fluência de leitura dos participantes das escolas, as respostas das crianças para o questionário de perguntas dirigidas²⁸ associado a cada um dos textos lidos foi computado. Assim, as categorias de leitura foram definidas pela razão entre o tempo de leitura dos textos (em segundos) e a pontuação nas perguntas dirigidas, com o intuito de garantir uma medida de compreensão leitora que levasse em consideração a compreensão dos participantes, não apenas os tempos de suas leituras; como a leitura dos textos foi silenciosa, então não havia como saber se as crianças haviam lido, de fato, os textos se este questionário não fosse respondido. Foram considerados bons leitores participantes cujas medidas de tempo/compreensão de leitura se situaram na faixa de menores valores da amostra (primeiro quartil). Em contrapartida, foram considerados maus leitores participantes com medidas opostas às dos bons leitores, correspondentes aos maiores valores de tempo/compreensão da amostra (quarto quartil).

²⁸ O questionário de perguntas dirigidas está disponível em ANEXOS (Anexo D).

3 ANÁLISE E DISCUSSÃO DE DADOS

Este capítulo discute os resultados obtidos ao passo que os apresenta. Organiza-se, assim, em duas grandes etapas. A primeira corresponde às análises estatísticas dos atributos de grafos representativos das produções textuais dos participantes. A segunda consiste em uma discussão a respeito dos resultados obtidos a partir da utilização de classificadores SVM programados para categorizar as produções textuais das crianças de acordo com a fluência leitora de cada uma, a partir dos atributos de grafo extraídos de seus textos.

3.1 ANÁLISES ESTATÍSTICAS DOS DADOS

Inicialmente, pretende-se elucidar os motivos subjacentes a prioridade conferida a alguns resultados em detrimento de outros, na discussão que se segue. Em vista disso, sete tabelas são apresentadas neste primeiro momento, reportando a significância dos atributos de grafo trabalhados, quando contrastadas as produções de diferentes subgrupos experimentais. Foram realizadas múltiplas rodadas de análises estatísticas – (I) análise de diferenças de grafo associadas aos estratos socioeconômicos dos pacientes do ambulatório, (II) análise de diferenças de grafo relacionadas ao tipo de escola dos participantes disléxicos, (III) análise longitudinal de produções textuais de crianças que produziram textos nos três anos de coleta e (III) análise de diferenças de grafo, por ano de coleta (2014, 2015 e 2016) e fluência leitora (bons leitores, leitores medianos, maus leitores e disléxicos) entre alunos das escolas e pacientes do Ambulatório. Cada uma destas rodadas, como dito anteriormente, foi conduzida a partir da análise de duas versões de cada texto – uma transcrição original e outra corrigida para ortografia – e de suas respectivas normalizações pelo número de palavras de cada texto, totalizando quatro análises de cada produção textual. Uma vez obtidos os resultados de significância apresentados nestas tabelas²⁹, optou-se por priorizar, na discussão subsequente, as análises baseadas nas transcrições originais e corrigidas não-normalizadas – para as quais os atributos registraram mais distinções entre os grupos.

Eis as tabelas:

²⁹ As análises longitudinais e transversais foram todas realizadas a partir da aplicação do teste Kruskal Wallis. Em seguida, houve a aplicação do teste Dunn sobre os resultados do Kruskal Wallis de múltipla comparação, com valores de significância ajustados para o método Holm.

Tabela 9: Análise longitudinal (Escolas)

Long.	WC*	N	E	PE	L1	L2	L3	LCC	LSC	ATD	Dens.	Diam.	ASP
C	0,0219	4,135x 10 ⁻³	0,0219	0,937	0,3677	0,113	0,2317	4,135x 10 ⁻³	3,748x 10 ⁻³	0,9282	6,302x 10 ⁻³	0,6945	0,5693
O	0,0108	3,171x 10 ⁻⁶	0,0182	0,6775	0,3677	0,0413	0,2019	3,697x 10 ⁻⁶	0,1225	0,6381	5,701x 10 ⁻⁶	0,0880	0,2019
C. N.	-	0,9535	0,0219	0,0185	0,3678	0,0557	0,9435	0,9535	0,0845	0,0034	7,557x 10 ⁻³	0,0647	0,0405
O. N.	-	0,4524	0,6683	0,4729	0,3678	0,0169	0,4201	0,9225	0,6061	2,174x 10 ⁻⁶	4,029x 10 ⁻⁶	0,0254	0,0302

Tabela 10: Análise transversal geral (Escolas e Ambulatório)

Trans.	WC*	N	E	PE	L1	L2	L3	LCC	LSC	ATD	Dens.	Diam.	ASP
C	5,054x 10 ⁻¹¹	6,595x 10 ⁻¹²	5,054x 10 ⁻¹¹	3,707x 10 ⁻⁷	0,1302	0,4424	0,1048	6,595x 10 ⁻¹²	1,038x 10 ⁻¹¹	1,835x 10 ⁻⁶	4,185x 10 ⁻¹¹	0,0496	5,489x 10 ⁻³
O	3,859x 10 ⁻¹²	1,445x 10 ⁻¹¹	6,951x 10 ⁻¹²	1,986x 10 ⁻⁶	0,9069	0,8138	0,1092	5,59x 10 ⁻¹³	1,711x 10 ⁻⁷	3,061x 10 ⁻⁶	2,102x 10 ⁻⁹	4,527x 10 ⁻⁶	2,054x 10 ⁻⁵
C. N.	-	0,00784	5,054x 10 ⁻¹¹	0,1363	0,1273	0,6835	0,8347	7,848x 10 ⁻⁶	0,0245	6,235x 10 ⁻¹²	1,326x 10 ⁻¹¹	1,045x 10 ⁻⁸	3,333x 10 ⁻⁹
O. N.	-	1,111x 10 ⁻⁶	0,0292	0,0161	0,923	0,9217	0,5436	0,3606	0,0203	5,125x 10 ⁻¹¹	3,49x 10 ⁻¹¹	5,141x 10 ⁻⁷	1,533x 10 ⁻⁹

Tabela 11: Análise transversal geral por ano de coleta e fluência leitora (2014)

2014	WC*	N	E	PE	L1	L2	L3	LCC	LSC	ATD	Dens.	Diam.	ASP
C	0,0728	0,0557	0,0459	0,0228	0,0558	0,4805	0,2387	0,0318	0,0253	0,149	0,0915	0,067	0,374
O	0,0389	0,0636	0,03	0,0375	0,1921	0,2635	0,1239	0,0714	0,0801	0,03817	0,114	0,709	0,462
C. N.	-	0,2028	0,0208	0,0401	0,0639	0,3842	0,177	0,480	0,247	0,0573	0,0684	0,303	0,208
O. N.	-	0,0572	0,165	0,2179	0,199	0,3761	0,2219	0,6104	0,1693	0,5168	0,0602	0,0213	0,4563

Tabela 12: Transversal geral por ano de coleta e fluência leitora (2015)

2015	WC*	N	E	PE	L1	L2	L3	LCC	LSC	ATD	Dens.	Diam.	ASP
C	0,0063	0,0084	0,0032	0,0138	0,509	0,7519	0,2878	0,0028	0,0039	0,0594	0,0085	0,067	0,0397
O	0,0027	0,0093	0,0015	0,0003	0,997	0,4552	0,1965	0,0025	0,0292	0,0038	0,0265	0,0145	0,0267
C. N.	-	0,1022	0,0084	0,186	0,482	0,668	0,786	0,331	0,453	0,0098	0,0095	0,1067	0,514
O. N.	-	0,002	0,125	0,0095	0,997	0,4681	0,466	0,626	0,202	0,0123	0,0812	0,2254	0,0107

Tabela 13: Transversal geral por ano de coleta e fluência leitora (2016)

2016	WC*	N	E	PE	L1	L2	L3	LCC	LSC	ATD	Dens.	Diam.	ASP
C	9,453x 10 ⁻⁷	1,157x 10 ⁻⁶	5,508x 10 ⁻⁷	2,271x 10 ⁻⁴	0,368	0,251	0,0292	3,137x 10 ⁻⁷	5,537x 10 ⁻⁸	6,483x 10 ⁻⁴	3,56x 10 ⁻⁶	0,0665	0,0033
O	3,038x 10 ⁻⁷	2,973x 10 ⁻⁶	2,098x 10 ⁻⁷	5,752x 10 ⁻⁶	0,675	0,774	0,0355	1,947x 10 ⁻⁷	3,55x 10 ⁻⁵	4,228x 10 ⁻⁶	4,866x 10 ⁻⁵	3,01x 10 ⁻⁴	1,381x 10
C. N.	-	2,215x 10 ⁻³	7,875x 10 ⁻⁸	0,2751	0,310	0,528	0,8231	0,0170	0,2742	2,119x 10 ⁻⁶	2,675x 10 ⁻⁶	1,253x 10 ⁻⁵	7,605x 10 ⁻⁶
O. N.	-	1,515x 10 ⁻⁶	0,042	0,01959	0,6695	0,9184	0,2649	0,2083	0,06528	4,736x 10 ⁻⁶	4,297x 10 ⁻⁶	3,804x 10 ⁻⁵	8,601x 10 ⁻⁷

A contabilização das recorrências de atributos com significância entre todas as análises revela que a alternativa mais expressiva de padrões de conectividade específicos por subgrupo experimental é a transcrição original não normalizada, com 63,07% (41/65) de significância para os atributos, entre as análises. A segunda é a transcrição corrigida não normalizada, com 58,4% (38/65) das medidas de grafo significativas. As versões normalizadas das originais e corrigidas tiveram baixa prevalência de distinção entre os grupos – em ambas, somente 43,3% (26/60) dos atributos tiveram significância. Por esta razão, os resultados destas análises foram desprezados nesta discussão³⁰.

A análise estatística dos dados que se inicia foi subdividida em duas partes. A primeira apresenta análises prévias realizadas a fim de confirmar a homogeneidade do grupo Ambulatório/Disléxicos e, se confirmada, estender os achados que indiquem pouca influência de fatores socioeconômicos não só sobre as habilidades de escrita do grupo Ambulatório, como também sobre as do grupo Escolas (cujos estratos socioeconômicos não foram controlados). A segunda reúne as principais análises rodadas intergrupos - comparando as produções textuais de Ambulatório e Escolas - e intragrupos - contrastando a evolução dos textos de trinta e três alunos das escolas ao longo de três anos.

³⁰ Após a defesa da dissertação, conclui-se ser importante salientar que, quando um grafo apresenta mais palavras que outro, o mais robusto tende a ter valores maiores para alguns atributos - como nós e arestas – pelo simples fato de possuir mais itens lexicais. Porém, isto não significa que haja, necessariamente, uma mudança estrutural no grafo. Além disso, quando os dados foram normalizados por WC (sob o uso de uma razão), assumiu-se que se tratava de uma relação linear, o que não é verdadeiro para todos os atributos. Tais questões serão corrigidas futuramente, nos artigos advindos desta dissertação.

3.1.1 Análises prévias - Ambulatório

Conforme reportado previamente, os participantes que integraram o grupo experimental Ambulatório compartilhavam o diagnóstico da dislexia do desenvolvimento. No entanto, diferiam entre si quanto à idade no momento de coleta dos dados – entre 8 e 16 anos -, ao tipo de escola que frequentavam – pública (N=42) ou privada (N=10) - e também quanto aos estratos socioeconômicos nos quais se inseriam – identificados como B (N=14), C (N=35) e D (N=3) a partir das respostas dos responsáveis pelos jovens ao questionário ABIPEME³¹, que compôs a anamnese.

A fim de verificar eventuais influências da variável tipo de escola na produção escrita do grupo Ambulatório – e, conseqüentemente, a necessidade de se trabalhar apenas com as crianças de escola pública (por serem mais numerosas e semelhantes, neste aspecto, aos participantes do grupo Escolas) -, foram feitas análises descritivas com medidas de posição e dispersão por categoria de escola. Posteriormente, foi aplicado o teste de Wilcoxon, com correção de continuidade, aos dados, uma vez que dois subgrupos estavam sendo comparados: Pública e Privada.

A Tabela 14 reporta os resultados de significância estatística obtidos quando foram contrastados os atributos de grafo de pacientes de escolas públicas com os de escolas privadas:

Tabela 14: Públicas x Privadas (Ambulatório)

Part.XPub.	WC*	N	E	PE	L1	L2	L3	LCC	LSC	ATD	Dens.	Diam.	ASP
C	0.807	0.617	0.807	0.887	0.094	0.276	0.075	0.617	0.569	0.269	0.471	0.217	0.094
O	0.802	0.624	0.829	0.404	0.222	0.39	0.198	0.925	0.925	0.227	0.2	0.17	0.197
C. N.	-	0.166	0.807	0.824	0.094	0.313	0.169	0.166	0.042	0.585	0.65	0.963	0.926
O. N.	-	0.203	0.512	0.454	0.222	0.441	0.265	0.359	0.02	0.293	0.231	0.279	0.941

Como se observa, os índices de significância dos atributos para cada uma das quatro análises se mantiveram maiores que 0,05 de modo geral, o que indicou substancial homogeneidade entre as produções dos dois grupos. LSC, contudo, quando analisado em corrigidas e originais normalizadas, revelou tendências para significância (em rosa, na Tabela 14). Como se sabe, o maior componente fortemente conectado, definido por Mota (2013) como

³¹ Disponível na sessão ANEXOS.

LSC, consiste em um atributo de conectividade que designa o maior agrupamento de nós que compõem o maior subgrafo no qual todos os nós estão ligados entre si, de modo que se pode ir do nó A para o nó B e, igualmente, do nó B para o nó A. Em Mota (ibid., p.27), LSC foi uma medida de grafo expressiva de diferenças significativas entre relatos de sonho e de vigília não normalizados de esquizofrênicos, bipolares e controles. Os esquizofrênicos deste estudo produziram grafos significativamente menos conectados – com menores índices de LCC, LSC e E – que bipolares e controles para relatos não normalizados. Entretanto, com a normalização dos dados, esta diferença aparentemente se perdeu.

No presente estudo e, especificamente, na análise em questão, atribui-se pouca relevância às duas tendências para significância encontradas para LSC, nas análises dos dados normalizados. A justificativa para tal decisão é a de que ambas as tendências perdem força em meio as outras quarenta e oito evidências de homogeneidade entre os grupos. Além disso, estes dois resultados são contrabalançados pelos outros dois valores para LSC em dados não normalizados, que indicam ser este atributo semelhante entre disléxicos oriundos de escolas públicas e privadas ($p = 0,529$ para corrigidas e $p = 0,925$ para originais). Assim, concluiu-se plausível a possibilidade de serem mantidos disléxicos advindos de escolas públicas e privadas em um mesmo grupo experimental – os seja, não foram descartados os dados dos estudantes de escola particular.

Quanto às diferenças entre os atributos de grafo de participantes disléxicos oriundos de três estratos socioeconômicos diferentes, após as análises descritivas com medidas de posição e dispersão por estrato socioeconômico, aplicou-se o teste não-paramétrico de comparação múltipla Kruskal-Wallis, que se presta a comparações entre três grupos. Os valores de significância, novamente, sugeriram homogeneidade entre as produções escritas dos pacientes do ambulatório – ainda que fossem oriundos de diferentes estratos socioeconômicos.

A Tabela 15 reúne os resultados obtidos para a significância estatística de cada atributo, em cada uma das quatro análises:

Tabela 15: Estratos econômicos – B x C x D (Ambulatório)

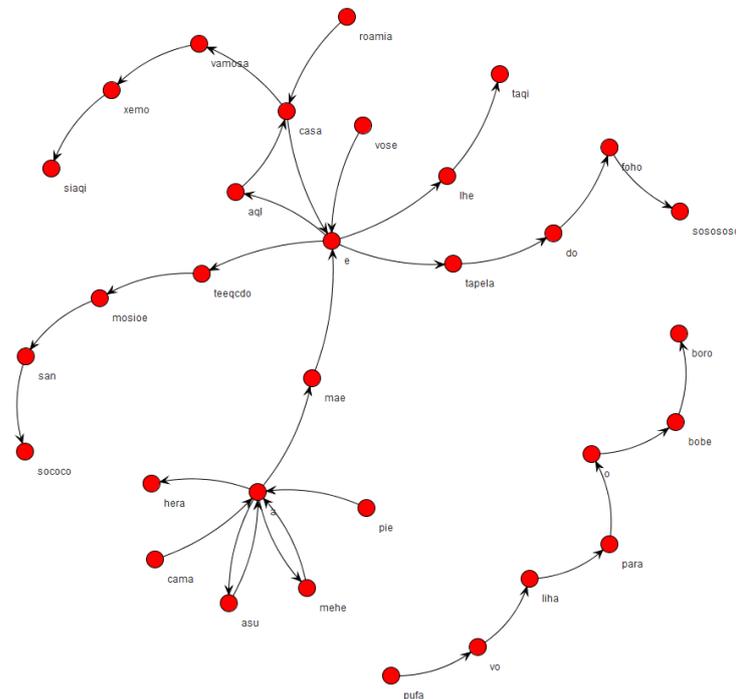
Estratos	WC*	N	E	PE	L1	L2	L3	LCC	LSC	ATD	Dens.	Diam.	ASP
C	0.123	0.107	0.123	0.416	0.035	0.405	0.159	0.107	0.124	0.492	0.058	0.135	0.056
O	0.165	0.209	0.140	0.462	0.286	0.002	0.801	0.231	0.426	0.223	0.363	0.702	0.612
C. N.	-	0.576	0.123	0.791	0.041	0.309	0.097	0.576	0.657	0.1	0.092	0.379	0.348
O. N.	-	0.356	0.104	0.132	0.284	0.005	0.790	0.491	0.895	0.334	0.224	0.081	0.082

Como se observa, as quatro análises revelaram pouquíssimo descompasso entre os textos dos participantes, corroborando a hipótese subjacente a este trabalho de que a dislexia do desenvolvimento, ainda que seja favorecida por questões de ordem ambiental e social (FLETCHER et al., 2009), depende, geralmente, de fatores outros, de ordem neurológica e psicológica, que afetam crianças independentemente de suas condições socioeconômicas e seus contextos de vivência (SHAYWITZ, 2008). Contudo, para dois atributos (cujos resultados estão grifados em azul na Tabela 15), houve diferenças entre as produções textuais do grupo Ambulatório: L1 e L2. Uma breve reflexão acerca das características de cada um, a partir de exemplos da amostra em questão, é capaz de auxiliar não só a compreensão destes resultados, como também o entendimento dos demais, que indicam que ciclos de um, dois e três nós são pouco distintivos de subgrupos experimentais neste estudo.

Como se sabe, L1 representa ciclos de um nó, isto é, quando um mesmo item lexical retorna para ele mesmo, quer seja porque houve uma falta de atenção por parte autor do texto - que causou a repetição da palavra -, uma intenção de dar ênfase à determinada palavra, uma tentativa de reproduzir determinado som ou simplesmente um encontro de itens lexicais iguais devido à eliminação da pontuação. Já L2 ou ciclos de dois nós são atributos mais complexos que os ciclos de um nó, dependentes de dois segmentos que encerram, juntos, um ciclo, por exemplo: “... era **uma vez uma**...”. Mas de que forma o tipo de transcrição pode impactar a identificação de ciclos de um ou dois nós?

Tem-se, como exemplos, as transcrições corrigida (Figura 24) e original (Figura 25) da produção textual do participante 87-A-C-2014-LCM, aluno de escola pública cujo estrato socioeconômico reportado foi o D:

Figura 27: Grafo da transcrição original do texto do participante 87-A-C-2014-LCM

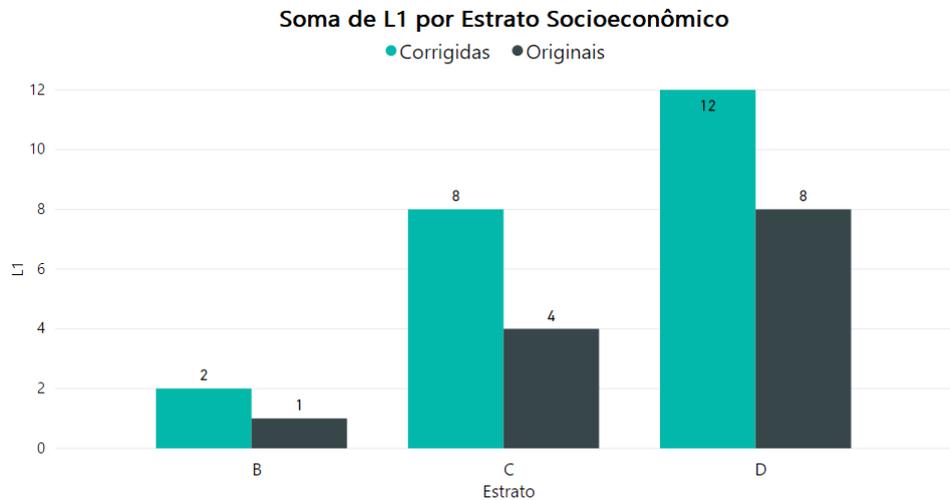


Fonte: A autora

A observação de ambas as transcrições e de seus respectivos grafos conduz ao entendimento de que cada versão registra de modo diferente itens lexicais incompreensíveis ou equivocados. A versão corrigida inclui dois ciclos de um nó – o primeiro retomando “SO” e o segundo, “A” - enquanto que a versão original não apresenta L1, tal como mostram os grafos acima. Tanto quando “SOSOSOSO” é tratado como um L1 (versão corrigida) como quando é considerado um único item (versão original), há inconsistências ortográficas e gramaticais, de modo que nenhuma das opções assegura uma compreensão satisfatória daquele trecho. O restante do texto, incluindo o L1 em “A” da versão corrigida, é repleto de casos como estes, em que nem mesmo a correção consegue elucidar muitas das palavras empregadas pelo autor. Logo, defende-se que não se trata de falta de atenção ou intenção de criar ênfase sobre aqueles itens específicos, tal como foram escritos. Entende-se que o texto, como um todo, caracteriza-se por imprecisões de ordem ortográfica e gramatical, comuns às produções textuais de disléxicos (TOPS et al., 2013; BASSO et al., 2017). Ciclos de um nó percebidos nas versões corrigidas são apenas algumas destas inconsistências, cuja identificação é favorecida pelo modo de transcrição corrigido. Isto porque, em casos como estes, a correção segmenta algumas “palavras” de modo a padronizar itens lexicais “supostamente recorrentes”, de acordo com o

entendimento do pesquisador. Como resultado, a incidência de L1s nas transcrições corrigidas foi superior à incidência destas nas originais, como retrata a Figura 28:

Figura 28: Soma de L1 Estrato Socioeconômico

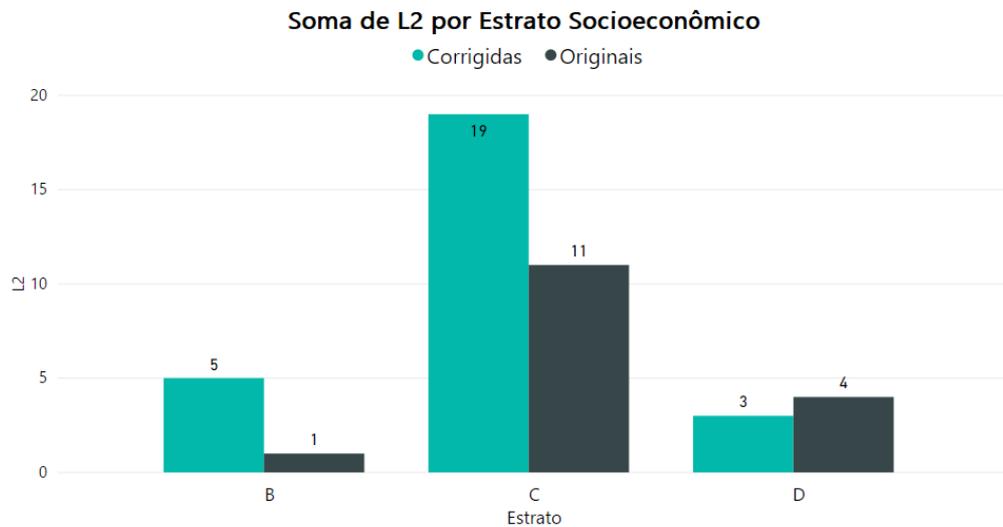


Fonte: A autora

O gráfico acima também revela que o grupo cuja produção de ciclos de um nó foi maior foi o composto por participantes do estrato D. Quanto a este resultado, afirma-se não ser possível generalizar os resultados para todo o estrato D porque havia apenas três participantes compondo este grupo, de modo que suas produções talvez não reflitam o padrão de conectividade do estrato, mas sim singularidades destes indivíduos. Se fosse mais robusta a amostra e os grupos mais equânimes nas suas quantidades de participantes, supõe-se que os resultados seriam mais parecidos, alimentando a suspeita de que fatores de natureza socioeconômica não são determinantes da fluência leitora.

Pode-se dizer que a maior incidência de L2 nas versões corrigidas se manteve parcialmente, como é possível observar no gráfico a seguir:

Figura 29: Soma de L2 por Estrato Socioeconômico



Fonte: A autora

Se comparado com o que ocorreu para L1, nota-se que a incidência de L2 não só foi menor para grupo D, como foi ligeiramente menor entre seus textos corrigidos, quando contrastados com os originais. A fim de verificar se o tamanho pequeno amostral contribuiu para que esta diferença entre os atributos - em princípio inexistente - fosse encontrada, aplicou-se um teste post-hoc Dunn aos dados, com o intuito de compreender a origem destes resultados e, com sorte, não ser preciso excluir participantes da amostra. Constatou-se, assim, que tanto para L1³² quanto para L2³³ as diferenças significativas se deram entre o grupo D e os demais, e nunca entre B e C. A constatação de que o subgrupo D foi crucial para estes dois resultados reforçou a ideia de que, se fosse um subgrupo experimental mais robusto, essa diferença não seria obtida. Logo, nenhum participante do subgrupo D foi excluído da amostra total do grupo Ambulatório.

A seguir, inicia-se a discussão dos resultados das análises intra e intergrupos.

³² B – C: p adj. = 0.78117164; B – D: p adj. = 0.03698105; C – D: p adj.: 0.02479270

³³ B – C: p adj.: 0.133617650; B – D: p adj. = 0.001827009; C – D: p adj. 0.009126064

3.1.2 Análises estatísticas intra e intergrupos – Escolas e Ambulatório

Houve uma análise intragrupos longitudinal que contemplou as produções das crianças das escolas que produziram textos nos três anos de coleta de dados do ACERTA. A seguir, são apresentados os dados e a discussão a respeito destes resultados.

3.1.3 Análise Longitudinal - Escolas

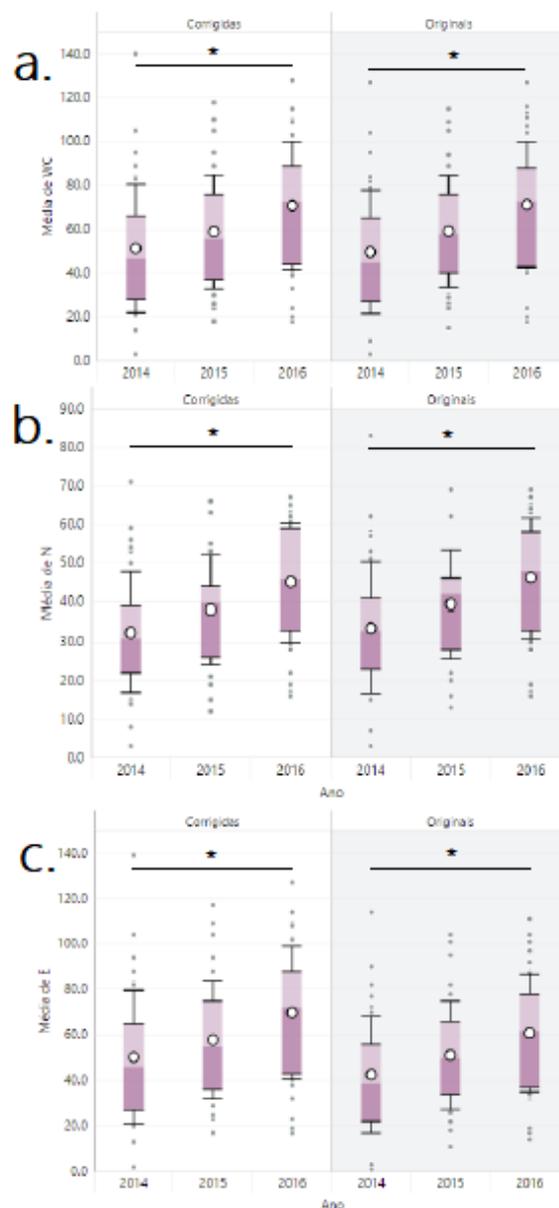
Inicia-se agora a inspeção das médias, desvios padrão e índices de diferença significativa entre as produções textuais de 33 alunos do grupo Escolas (23 meninas e 10 meninos), para cada um dos três anos de coleta de dados. Infere-se, a partir destes dados, que o tipo de transcrição não foi crucial na distinção dos grupos (se consideradas corrigida e original). Exceto para dois atributos - L2 e LSC - para os quais foi obtida diferença significativa apenas na análise das produções originais e corrigidas respectivamente, os demais resultados foram similares para ambos os tipos de transcrição. Além disso, o teste Dunn indicou diferença significativa para os atributos de grafo apenas entre 2014 e 2016.

A começar pelas médias das medidas gerais de grafos e pela contagem de palavras (Figura 30, a.), nas versões corrigidas, houve diferença significativa ($p = 0.02193$), especificamente, entre o primeiro e o último ano de coleta ($p \text{ adj.} = 0.01777975$). Diferença ainda maior ($p = 0.01081$) foi constatada na análise das versões originais, novamente entre os anos de 2014 e 2016 ($p \text{ adj.} = 0.007970434$). Estes resultados dialogam diretamente com os obtidos para as médias de nós de grafo (Figura 30, b.). A diferença significativa se manteve maior na análise das produções originais (corrigidas $p = 0.004135$; originais $p = 0.003171$) e unicamente entre 2014 e 2016 (corrigidas $p \text{ adj.} = 0.002826689$; originais $p \text{ adj.} = 0.002105771$).

O aumento das médias de nós e números de palavras ao longo do tempo parece refletir o incremento e a profusão lexicais típicas de crianças sem prejuízos cognitivos, resultante dos seus avanços na vida escolar. Se considerados os dados advindos das transcrições originais, conclui-se que a progressão constatada ao longo do tempo nas médias dos três grupos, para cada um destes atributos (2014 – WC= 49,67; N= 33,52; 2015 – WC = 59,15; N = 39,42; 2016 – WC = 71,18; N=; 46,18) indica não só que os escolares passaram a produzir textos com mais palavras gradativamente, mas também que o fizeram aumentando a incidência das mesmas palavras nas suas produções textuais, uma vez que o percentual da quantidade média de nós sobre a média de total de palavras diminuiu – 67,48%; 66,64% e 64,87% respectivamente. Este

dado, em certa medida, contraria a perspectiva de que a progressão de escolaridade contribui para o incremento de vocabulário (GÂNDARA; BEFI-LOPES, 2010). No entanto, sabe-se que os picos de desenvolvimento lexical acelerado durante a vida escolar estão sujeitos à interferência de fatores sociais - circunscritos ou não ao contexto escolar (BENEDICT, 1979), tais como estímulos promovidos pelos pais e responsáveis. Logo, entende-se que estes resultados possam refletir a interferência de contextos de pouco estímulo linguístico – comuns a muitos estudantes da rede pública de ensino brasileira.

Figura 30: Atributos gerais de grafo (N e E) e contagem de palavras (WC) (Longitudinal)



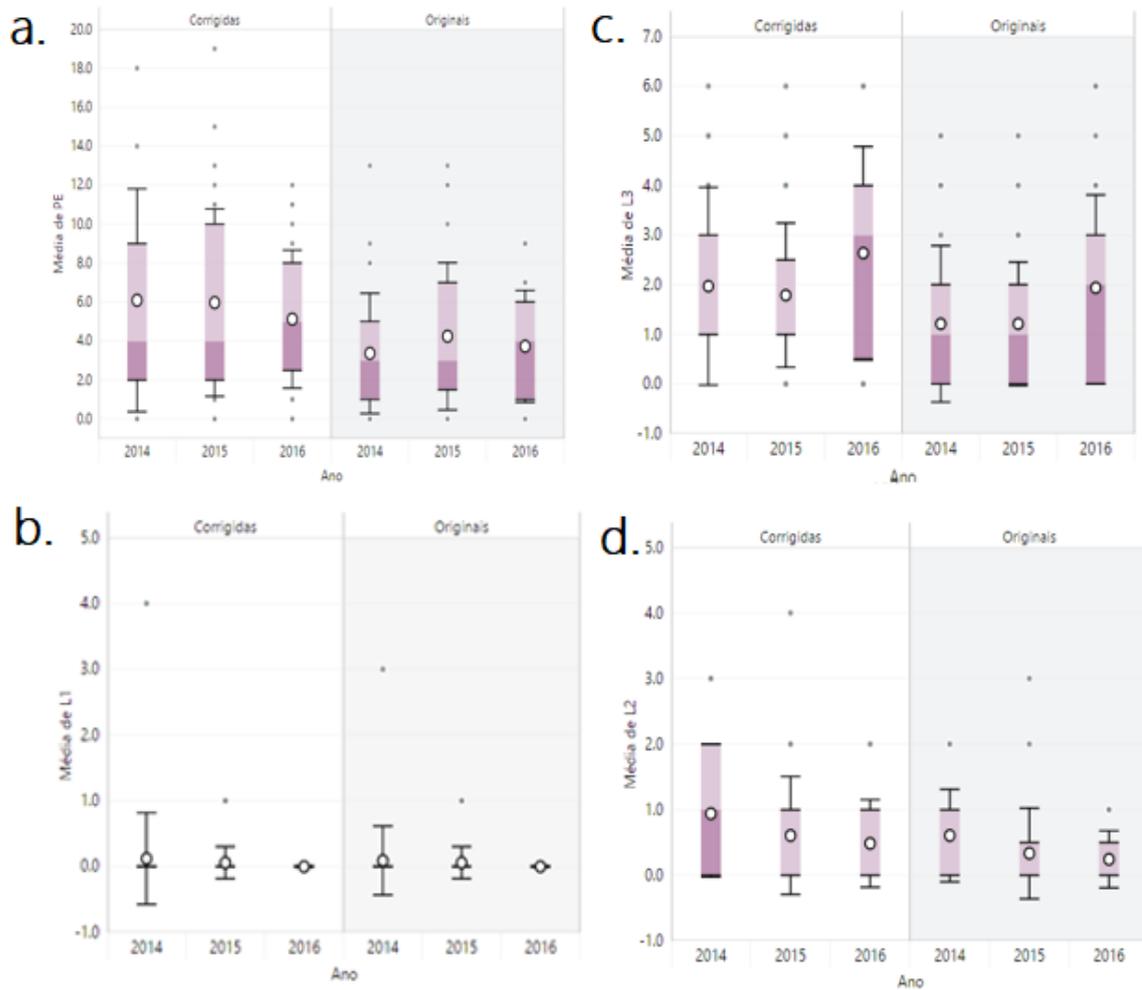
Fonte: A autora

Constatou-se também diferença significativa entre as versões corrigidas ($p = 0.02193$) e originais ($p = 0.01825$) para a produção de arestas (Figura 30, c.), novamente entre o primeiro e o último ano de coleta ($p \text{ adj.} = 0.01777975$ e $p \text{ adj.} = 0.01400102$ respectivamente). Tal resultado reflete o aumento significativo nas relações estabelecidas entre os itens lexicais de um texto – alicerce da face quantitativa da coesão textual.

Atributos de recorrência (Figura 31) indicaram significância apenas para L2, em versões originais das produções textuais, mas sem grande força ($p = 0.04139$), uma vez que nem mesmo para o primeiro e último ano houve distinção (2014 - 2015 $p \text{ adj.} = 0.07299646$, 2014 - 2016 $p \text{ adj.} = 0.06978742$, 2015 – $p \text{ adj.} = 0.85890887$). Na literatura, PE e L3 apresentaram correlação negativa com performance cognitiva em Mota et al. (2016), estudo no qual foi constatado que maiores quocientes de inteligência e melhores níveis de desenvolvimento da Teoria da Mente³⁴ foram acompanhados por menores incidências de PE e L3 em relatos autobiográficos de crianças entre 6 e 8 anos. Sabe-se que quocientes de inteligência podem sofrer flutuações ao longo do tempo, associadas ao desenvolvimento cerebral na adolescência (RAMSDEN et al., 2011). Como os participantes das escolas não foram avaliados quanto aos seus quocientes de inteligência, fortes inferências acerca desta provável correlação não são possíveis. Admite-se apenas a chance de seus níveis de QI não terem sofrido flutuações significativas durante esta janela temporal, o que justificaria parcialmente a pouca expressividade dos atributos de recorrência nesta análise – tendo em vista que os mesmos alunos foram acompanhados ao longo dos três anos.

³⁴ Teoria da Mente, especificamente, apresentou correlação negativa apenas com PE (MOTA et al., 2016, p. 8)

Figura 31: Atributos de recorrência (PE, L1, L2 e L3) (Longitudinal)

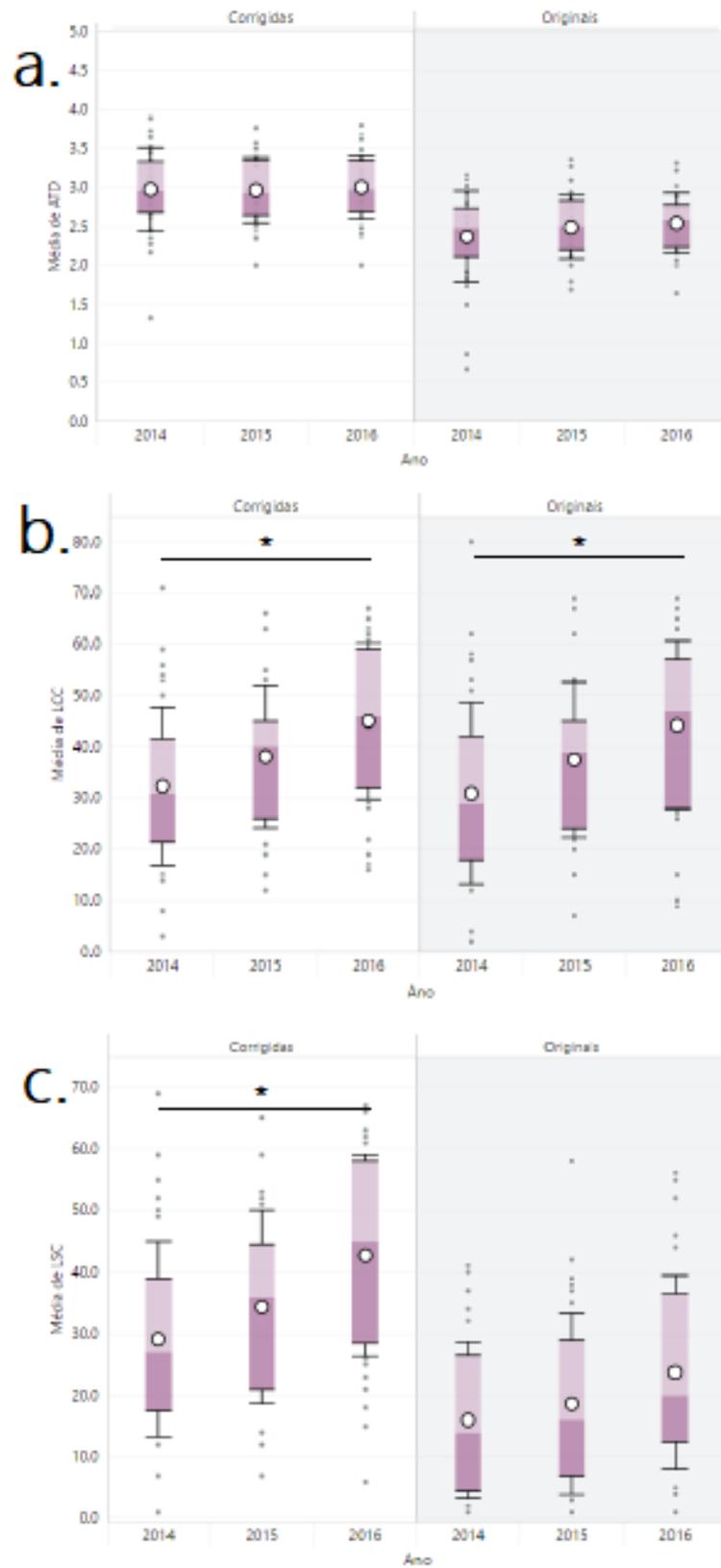


Fonte: A autora

As medidas de conectividade apresentaram diferenças significativas para LCC (Figura 32, b.; $p = 0.004135$ nas corrigidas e $p = 0.003697$ nas originais) e LSC (somente nas corrigidas, com $p = 0.003748$, Figura 32, c.). Já ATD (Figura 32, a.) mostrou-se irrelevante para a distinção longitudinal dos sujeitos de pesquisa a partir de suas produções textuais corrigidas e originais³⁵ – ou, em outras palavras, pouco expressiva da evolução das habilidades de escrita (e consequentemente de leitura) destas crianças.

³⁵ Houve diferença significativa para as análises normalizadas (Tabela 10), mas elas não serão discutidas neste trabalho.

Figura 32: Atributos de conectividade (ATD, LCC e LSC) (Longitudinal)



Fonte: A autora

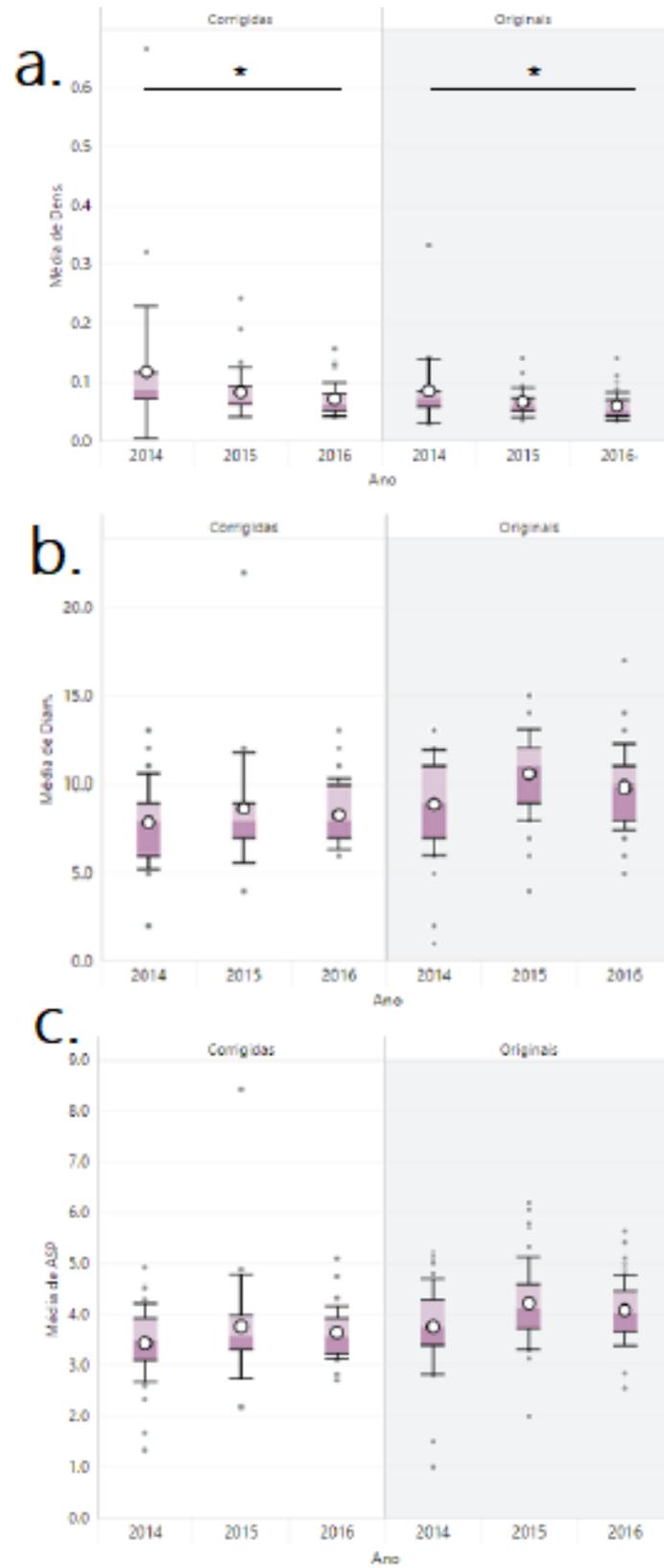
Mota et al. (2016) encontraram correlação positiva com performance de leitura para ambos os atributos – LCC e LSC. Logo, se forem considerados os resultados do teste Dunn aplicado, que, mais uma vez, indicaram distinção somente entre as produções de 2014 e 2016 - tanto nas corrigidas (p adj. = 0.002826689 para LCC e p adj. = 0.002761068 para LSC) quanto nas originais (p adj. = 0.002459695 para LCC), é possível inferir que as evidências encontradas por Mota et al. (ibid.) são reforçadas pelas evidências obtidas nesta análise longitudinal. Isto porque as médias de LCC e LSC – medidas quantitativas de qualidade de coesão textual – tanto nas versões originais quanto nas corrigidas, progrediram de 2014 para 2016³⁶, tal como se espera que aconteça com o avanço escolar (GÂNDARA; BEFI-LOPES, 2010).

Finalmente, a análise estatística das medidas de grafo globais revelou que apenas a densidade (Figura 33, a.) foi significativamente distintiva da escrita das crianças para cada um dos três anos de coleta (p = 0.006302 para as versões corrigidas e p = 0.005701), sendo 2014 e 2016 os anos com maiores diferenças entre suas densidades (p adj. = 0.004452921 para as corrigidas e p adj. = 0.003953354 para as originais). A densidade parece retratar a maleabilidade, a flexibilidade dos textos não no sentido de quantificar a imprecisão do texto, mas sim de mensurar quão fáceis de manipular, interpretar e compreender eles são; quanto maior a densidade, mais rígidos estão os fios que compõem o tecido textual e maior a sua espessura, o que prejudica sua maleabilidade. Bertola et al. (2014) identificaram correlação negativa entre densidade e status cognitivo geral e performance funcional, em idosos; aqueles com funções cognitivas mais preservadas produziram mais nós e arestas em redes com maiores diâmetro e ASP, mas menor densidade. As médias da presente análise sugerem que a densidade textual regride com o aumento da escolaridade³⁷; especificamente, que a densidade regride quase que exponencialmente com o aumento do número de palavras de um texto. Logo, a progressão escolar demonstra aprimorar a flexibilidade destes fios, diminuindo a espessura da tessitura textual sem torná-la inconsistente, mas sim mais manipulável, compreensível a quem lê o texto.

³⁶ Especificamente, 32,36 (2014), 38,12 (2015) e 45,06 (2016 para LCC nas corrigidas; 30,91 (2014), 37,52 (2015) e 44,21 (2016) nas para LCC originais; 29,15 (2014), 34,33 (2015) e 42,7 (2016) para LSC nas corrigidas.

³⁷ Especificamente, para as versões corrigidas, 0,11802 (2014), 0,08409 (2015) e 0,07227 (2016). Para as originais, 0,08663 (2014), 0,06721 (2015) e 0,06012 (2016).

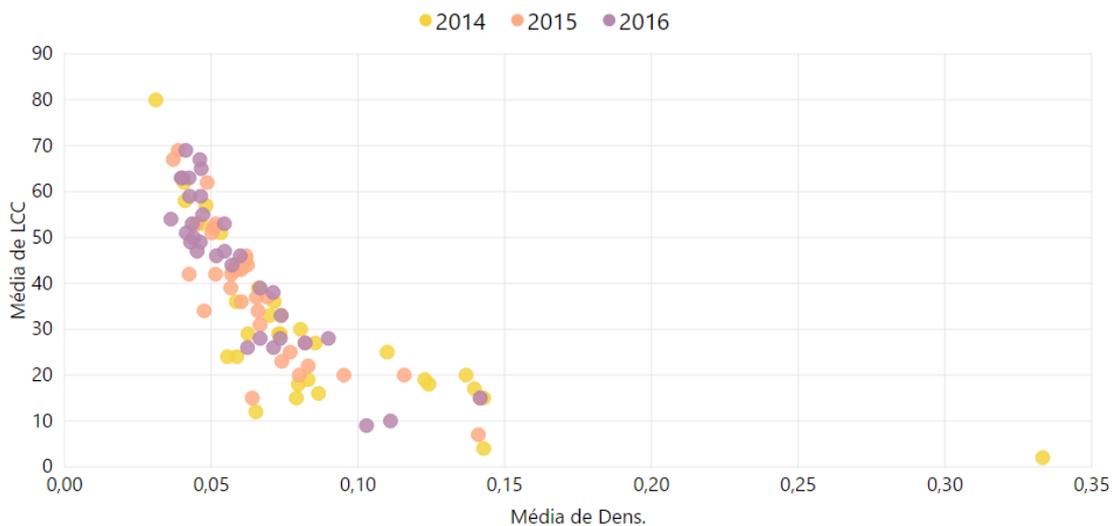
Figura 33: Atributos globais (Dens. Diam. e ASP) (Longitudinal)



Fonte: A autora

Por fim, ao serem comparadas as variações das médias de LCC e Dens., dois dos atributos significativamente diferentes para as produções originais dos três anos, percebe-se que as medidas parecem variar em sentidos opostos (Figura 34). Esta relação, portanto, indica que o aumento de nós é acompanhado por maiores agrupamentos de nós e menores densidades. E este padrão de conectividade, de acordo com as evidências discutidas, tende a acompanhar a progressão da escolaridade das crianças.

Figura 34: Relação entre as médias de Dens. e LCC. (Longitudinal)



Fonte: A autora

3.1.4 Análise transversal geral

Uma vez discutida a relação entre os atributos de grafo de textos escritos e a progressão na vida escolar no início do Ensino Fundamental, examinam-se agora as diferenças entre o grupo Ambulatório (N=52, 16 meninas e 36 meninos) e os três subgrupos experimentais do grande grupo Escolas - 2014 (N=80, 50 meninas e 30 meninos), 2015 (N=116, 69 meninas e 47 meninos) e 2016 (N=104, 61 meninas e 43 meninos). O propósito desta análise, essencialmente, consiste em verificar se há uma maior aproximação entre os padrões de conectividade do subgrupo 2014 e do grupo Ambulatório e, conseqüentemente, uma menor aproximação entre o subgrupo 2016 e o grupo Ambulatório.

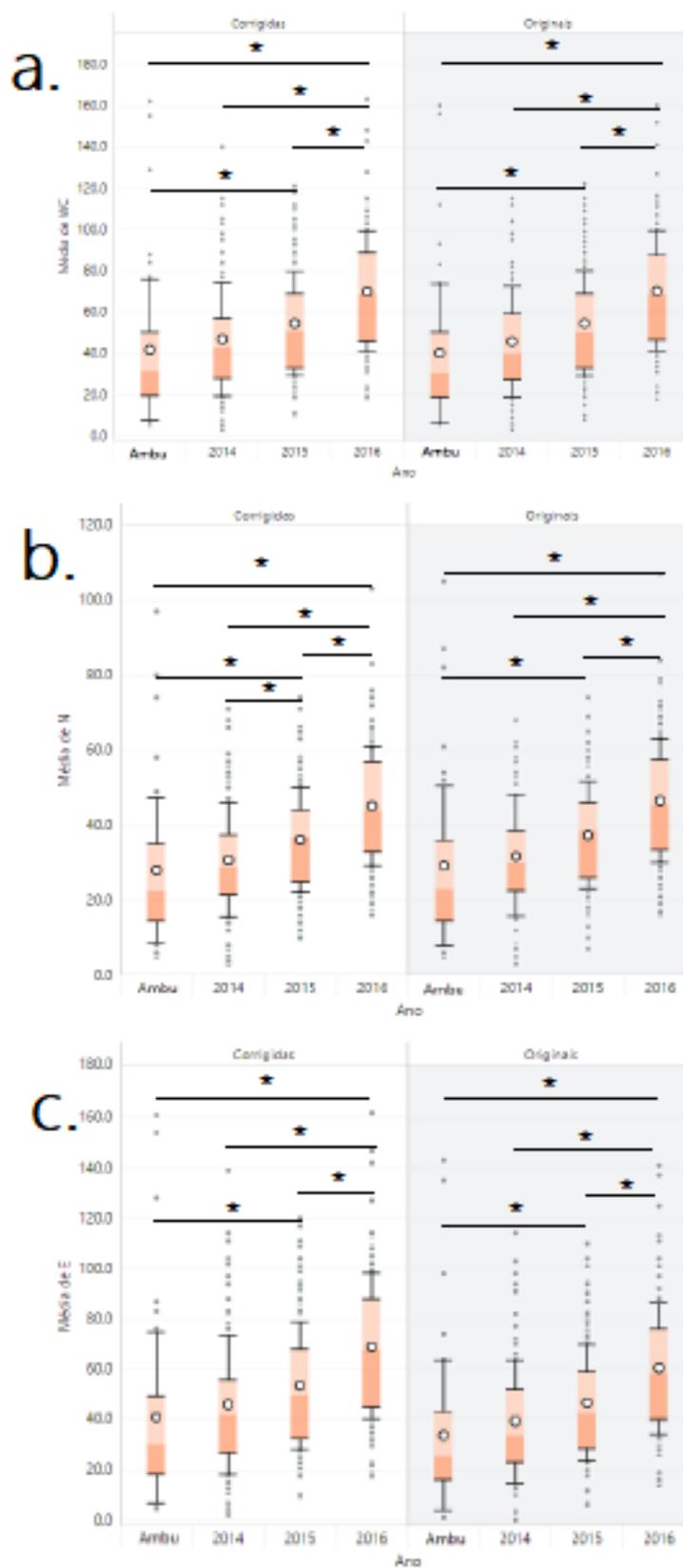
Mais uma vez, todos os atributos gerais – N e E –, somados a WC, apresentaram diferença significativa entre os grupos, tanto para as versões corrigidas ($WC_p = 5,054 \times 10^{-4}$, $N_p = 6,595 \times 10^{-12}$, $E_p = 5,054 \times 10^{-11}$) dos textos quanto para as originais ($WC_p = 3,859 \times 10^{-12}$, $N_p = 1,445 \times 10^{-11}$, $E_p = 6,951 \times 10^{-12}$), tal como ilustra a Figura 35. A aplicação do teste Dunn revelou que os grupos se mostraram substancialmente diferentes entre si na produção destes atributos, exceto 2014 e Ambulatório, que sempre demonstraram semelhança entre si para todos os atributos, em todas as análises³⁸. Assim, as médias de WC, N e E, para as duas versões de transcrição (Figura 35, a., b. e c. respectivamente), indicam ser coerente a expectativa de que a progressão escolar em fase de alfabetização e, conseqüentemente, o aprendizado da leitura e da escrita sejam acompanhados por um incremento gradativo na profusão lexical (WC), diversidade vocabular (N) e articulação (E) dos itens lexicais de um texto. Isto explica porque o padrão de conectividade de atributos gerais de 2014 é mais similar ao do Ambulatório, grupo composto por crianças cuja manipulação da língua escrita é deficitária.

É pertinente observar a diferença entre as médias dos atributos e sua relação com as considerações feitas a respeito do incremento da riqueza vocabular. Não por acaso, as médias de nós são inferiores às médias de número de palavras.

É compreensível que haja um aumento gradativo no número de nós entre as produções textuais do grupo Ambulatório e as produzidas em 2016, por crianças com diferentes níveis de proficiência leitora. Os subgrupos de 2014 e 2015, coerentemente, situam-se no intermédio desta progressão. Este resultado, se somado ao comportamento similar do número total de palavras e de arestas ao longo dos anos, demonstra que, mesmo havendo aumento no número de palavras, os nós permanecem menos numerosos que as palavras ao longo do tempo – o que faz sentido, visto que há palavras que se repetem e as relações que cada item lexical estabelece com os demais (E) acompanha este aumento.

³⁸ Resultados do Teste Dunn que apontaram semelhança entre 2014 e Ambulatório para atributos gerais: WC Corrigidas – p adj. = 0,1004, WC Originais – p adj. = 0,0982, N Corrigidas – p adj. = 0,1568, N Originais – p adj. = 0,1862, E Corrigidas – p adj. = 0,1004, E Originais – p adj. = 0,0913. Todos inferiores a 0,05 – logo, não significativos.

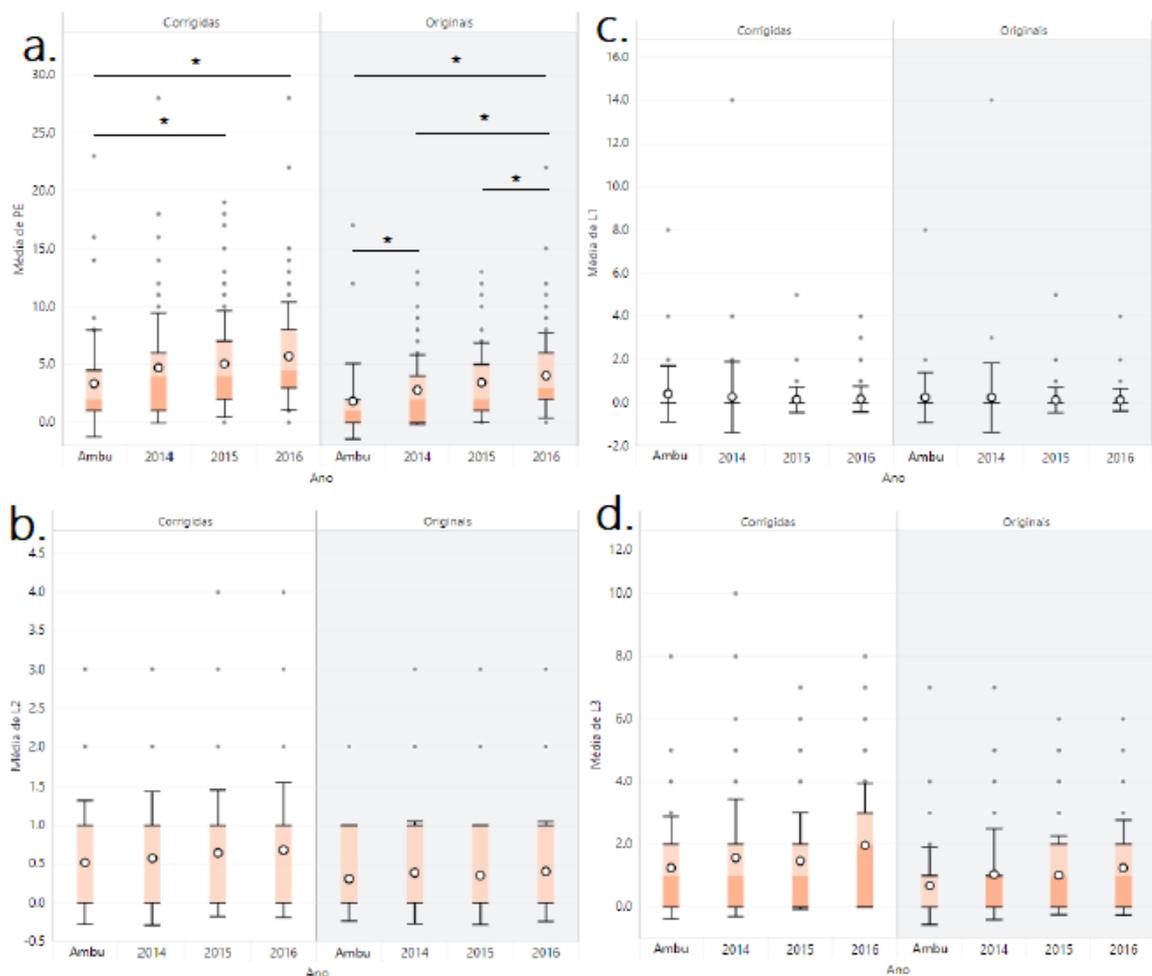
Figura 35: Atributos gerais (N e E) e número de palavras (WC) (Transversal Geral)



Fonte: A autora

Dos atributos de recorrência trabalhados, somente PE (Figura 36, a.) foi significativamente distintivo dos grupos (PE_p corrigidas = 0,00037 e PE_p originais = $1,986 \times 10^{-6}$), especialmente de 2015 e Ambulatório e ($P.adj$ corrigidas = 0,00952016 e $P.adj$ originais = 0,000359) e 2016 e Ambulatório ($P.adj$ corrigidas = 0,000133 e $P.adj$ originais = $9,46 \times 10^{-7}$); os demais ciclos de um, dois e três nós (Figura 36, c., b. e d. respectivamente) não foram expressivos de diferença entre os sujeitos de pesquisa, como mostra a Figura 36:

Figura 36: Atributos de recorrência (PE, L1, L2 e L3) (Transversal Geral)



Fonte: A autora

Complementando a discussão anterior a respeito do impacto de ciclos de nós neste estudo, reconhece-se a possibilidade de ciclos de um, dois ou três nós representarem a retomada intencional de itens lexicais, quer seja dando destaque a palavras específicas como um traço estilístico ou marcando ênfase. Admite-se, igualmente, a possibilidade de estes ciclos registrarem um simples encontro de itens lexicais fruto da exclusão da pontuação (em ambas as

versões) ou da falta de quebra de linha (nas corrigidas). Por fim, a hipótese de que estes ciclos sejam marcadores de uma desatenção no momento da escrita, que provoque a repetição de palavras, também é levantada. Problemas de atenção (DE SOUZA et al., 2007) e retomada intencional parecem possíveis e eventualmente comuns a disléxicos e a leitores típicos com diferentes anos de escolaridade, o que, entende-se, justifica a aparente homogeneidade entre os grupos para estes atributos – isto é, a irrelevância destes aspectos para a distinção entre os grupos.

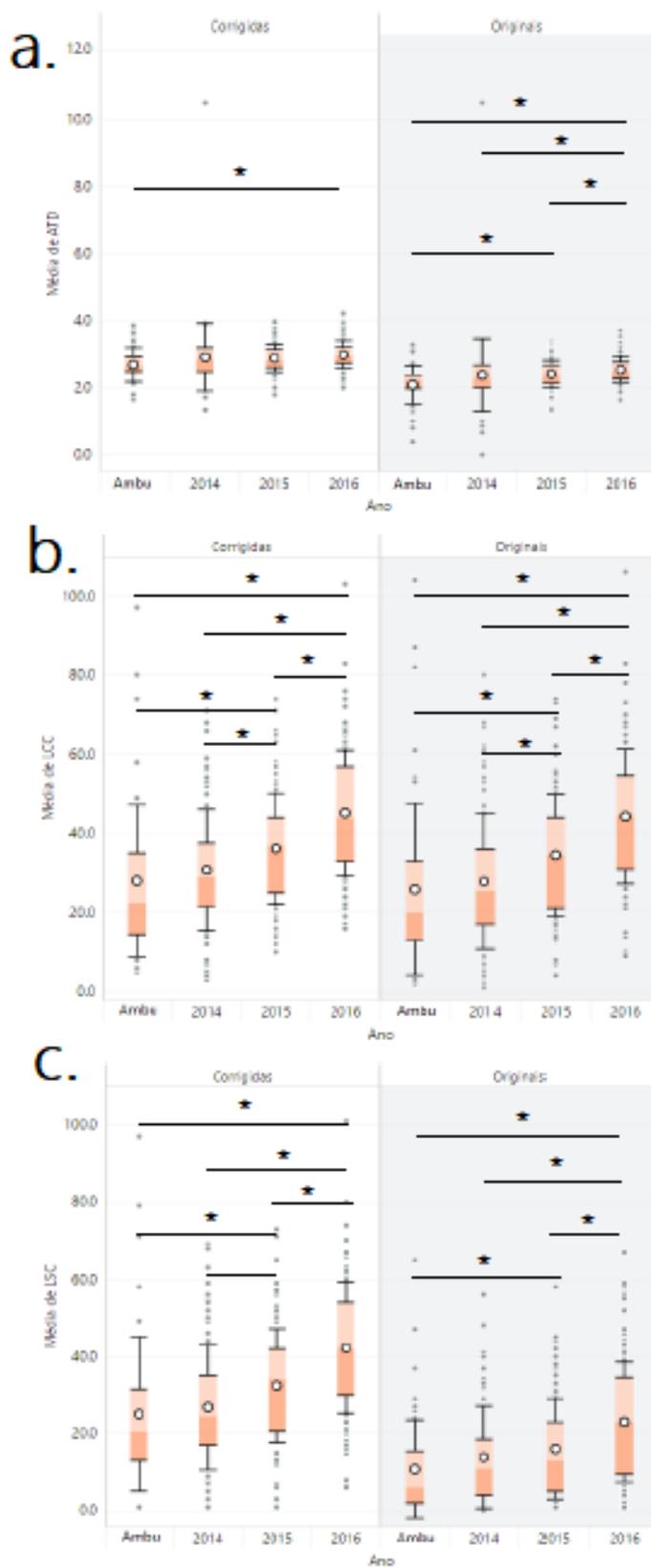
Os atributos de conectividade foram significativos também entre os textos corrigidos e os originais, mas com certas diferenças. ATD (Figura 37, a.), por exemplo, teve significância apenas entre Ambulatório e 2016 ($p = 0,00183$ e $p \text{ adj.} = 0,0008327$) nas corrigidas. Para as originais, no entanto, mostrou-se mais distintivo dos grupos ($p = 3,061 \times 10^{-6}$), não sendo eficaz apenas para a distinção de 2014 e 2015 ($p \text{ adj.} = 0,2602$) e 2014 e Ambulatório ($p \text{ adj.} = 0,06976$). LCC e LSC (Figura 37, b. e c.), entre as corrigidas, foram significativamente distintivos ($p = 6,595 \times 10^{-12}$ e $p = 1,038 \times 10^{-11}$), exceto para 2014 e Ambulatório ($p \text{ adj.} = 0,1568$ e $p \text{ adj.} = 0,35054$). Já entre as originais, LCC manteve o mesmo alcance de significância ($p = 5,59 \times 10^{-13}$ e $p \text{ adj.} = 0,2822$ para a diferença entre 2014 e Ambulatório). LSC, por sua vez, deixou de ser distintivo de 2014 e 2015 ($p \text{ adj.} = 0,1651$) e manteve-se pouco expressivo para 2014 e Ambulatório ($p \text{ adj.} = 0,2667$), sendo distintivo nos demais atributos ($p = 1,711 \times 10^{-7}$).

Estes resultados, se associados aos já reportados na literatura, indicam que a correlação positiva com performance de leitura encontrada por Mota et al. (2016) foi percebida³⁹, bem como a correlação negativa com déficits cognitivos⁴⁰, encontrada por Bertola et al. (2014). A expectativa atendida de que 2014 e Ambulatório tivessem padrões de conectividade similares não se deu ao acaso, uma vez que a fluência de leitura de crianças e adolescentes com dislexia (Ambulatório) tende a manter-se próxima da apresentada por crianças em fase de alfabetização (2014). Logo, a análise realizada, novamente, dialoga com as conclusões obtidas em estudos prévios e, assim, corrobora as expectativas do presente trabalho

³⁹ Ainda que os critérios de avaliação da fluência leitora sejam outros.

⁴⁰ Salvo as especificidades de ambos os estudos – do presente trabalho e do de Bertola et al. (2014), cujos problemas e escopo de pesquisa são distintos.

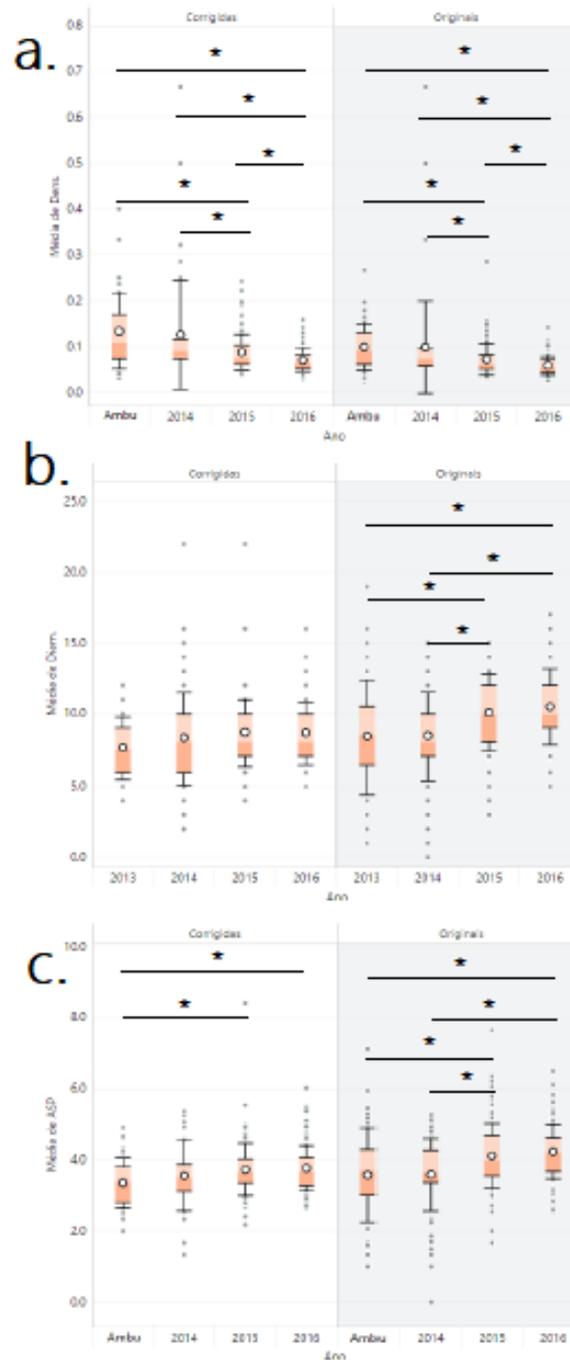
Figura 37: Atributos de conectividade (ATD, LCC e LSC) (Transversal Geral)



Fonte: A autora

Por fim, todos os atributos globais revelaram significância na análise transversal geral para os textos originais. Além disso, ASP (Figura 38, c., $p_{\text{corrigidas}} = 0,005489$ e $p_{\text{originais}} = 2,054 \times 10^{-5}$), Dens. (Figura 38, a., $p_{\text{corrigidas}} = 4,185 \times 10^{-11}$ e $p_{\text{originais}} = 2,102 \times 10^{-9}$) e Diam. (Figura 38, b., $p_{\text{corrigidas}} = 0,04962$ e $p_{\text{originais}} = 4,527 \times 10^{-6}$) não foram eficazes na distinção entre os textos de 2014 e os do Ambulatório, resultado em plena consonância com os anteriores e com as hipóteses de trabalho deste estudo.

Figura 38: Atributos globais (Dens., Diam. e ASP) (Transversal Geral)



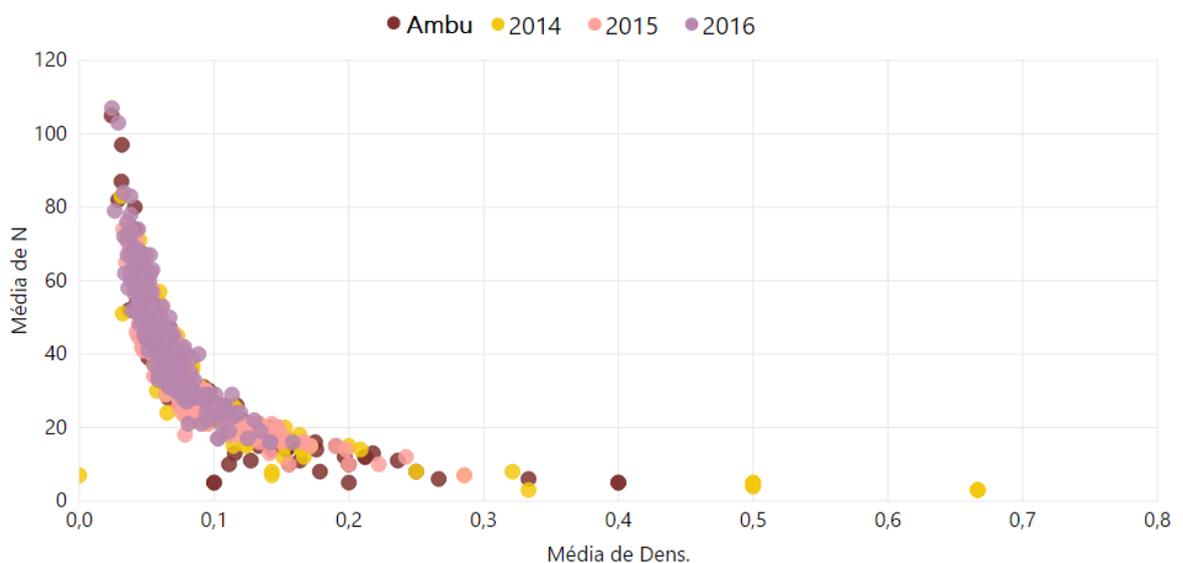
Fonte: A autora

Encontrar potencial de distinção, na análise transversal em questão, em atributos globais significa dizer que os textos de crianças com níveis diferentes de escolaridade apresentam diferentes dimensões, que os tornam mais ou menos inteligíveis. Quanto mais espessos (D), maior a probabilidade de serem condensados e compactados em pequenas dimensões (Diam. e ASP).

Em resumo, a análise transversal geral foi concluída com um resultado comum a todos os atributos: a ausência de diferença significativa entre as produções do subgrupo 2014 (Escolas) e as do grupo Ambulatório, o que corroborou a hipótese de trabalho que previa semelhança de padrões de conectividade entre os textos oriundos de ambos os grupos. Acerca dos atributos globais, já era sabido que diâmetro e ASP apresentaram correlação positiva com status cognitivo geral e performance funcional em Bertola et al. (2014). Constatar que grupos clínicos distintos (disléxicos e idosos com Alzheimer especificamente) parecem compartilhar um padrão de conectividade textual de atributos globais (ainda que seus grafos tenham diferentes estruturas) é interessante e pode contribuir para discussões acerca de convergências e divergências entre os comprometimentos cognitivos de caráter linguístico recorrentes na dislexia e na doença de Alzheimer.

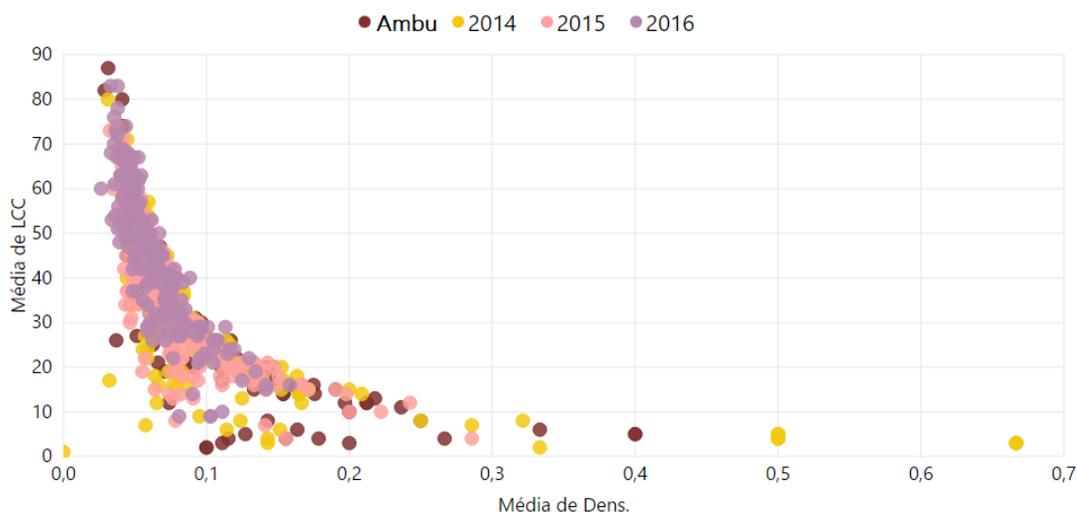
Assim, mais uma vez, constata-se um padrão de conectividade: quanto mais alta a média de diversidade vocabular (N), maior a escolaridade e menor a média de densidade (Figura 39). Igualmente, quanto maior o maior componente conectado, maior a escolaridade e menor a densidade média do texto (Figura 40).

Figura 39: Relação entre as médias de Dens. e N (Transversal Geral)



Fonte: A autora

Figura 40: Relação entre as médias de N e Dens. (Transversal Geral)



Fonte: A autora

3.1.5 Análise transversal geral entre categorias de fluência de leitura, por ano de coleta

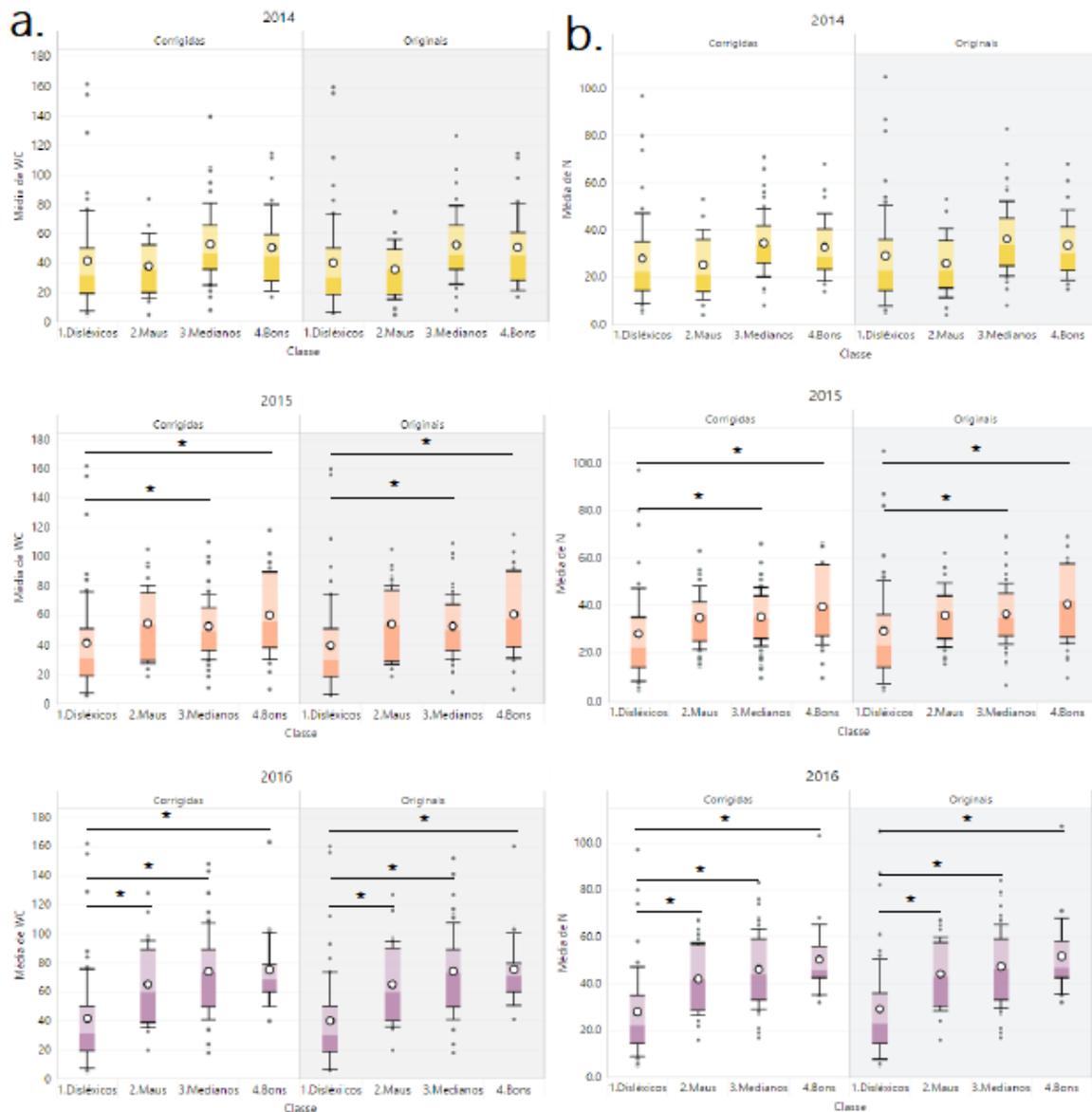
Compreender a flutuação das medidas de grafo entre os diferentes subgrupos das escolas - definidos com base nos *scores* dos alunos para um teste de fluência de leitura⁴¹ e no momento em que foram produzidos os textos - e o grupo Ambulatório/Disléxicos foi uma etapa crucial neste trabalho. Os resultados aqui discutidos dialogam com a segunda etapa da análise de dados, dedicada a avaliar e discutir a eficácia de classificadores SVM na distinção de tais grupos e, portanto, na predição de fluência de leitura, a partir de atributos de grafo. Em 2014, foram analisados 20 bons leitores (15 meninas e 6 meninos), 13 maus leitores (8 meninas e 5 meninos), 34 leitores medianos (20 meninas e 14 meninos) e 52 disléxicos (16 meninas e 19 meninos). Em 2015, foram analisados 21 bons leitores (9 meninas e 12 meninos), 24 maus leitores (14 meninas e 10 meninos), 51 leitores medianos (30 meninas e 21 meninos) e 52 disléxicos (16 meninas e 19 meninos). Por fim, em 2016, foram avaliados 21 bons leitores (12 meninas e 9 meninos), 21 maus leitores (14 meninas e 7 meninos), 34 leitores medianos (23 meninas e 11 meninos) e 52 disléxicos (16 meninas e 19 meninos).

⁴¹ Disponível em ANEXOS

O número de palavras já demonstrou ser um aspecto textual comum em análises de textos produzidos por adultos disléxicos e controles (TOPS et al., 2013), não sendo significativamente distintivo de textos do grupo clínico e do grupo com habilidades de leitura típicas. Este resultado se repetiu, no presente trabalho, entre as medidas de grafo das versões corrigidas de 2014 ($p = 0,07282$) e pode-se dizer que entre as versões originais também ($p = 0,03899$), uma vez que não houve distinção significativa entre os grupos⁴², de acordo com os resultados obtidos com a aplicação do teste Dunn. Foi encontrada significância entre os textos corrigidos e originais de 2015 ($p_{\text{corrigidas}} = 0,00633$ e $p_{\text{originais}} = 0,002725$) e, com maior força, de 2016 ($p_{\text{corrigidas}} = 9,453 \times 10^{-7}$ e $p_{\text{originais}} = 3,038 \times 10^{-7}$). Estes resultados indicam que a progressão escolar acentuou a diferença entre o número de palavras (Figura 41, a.) empregado pelos participantes do projeto com diferentes níveis de fluência leitora. As menores médias foram as apresentadas pelos disléxicos e as maiores, pelos bons leitores, configurando-se assim uma variação de mesmo sentido para fluência de leitura e número de palavras em uso em um texto, tal como ilustra a Figura 41:

⁴² Disléxicos e bons leitores – p adj. = 0,38882; Disléxicos e maus leitores – p adj. = 0,9212; bons leitores e maus leitores – p adj. = 0,57636; Ambulatório e medianos – p adj. = 0,05972; bons leitores e medianos – p adj. = 1,0; maus leitores e medianos – p adj. 0,34683.

Figura 41: Número de palavras (WC) e nós (E) (Transversal por ano e fluência de leitura)



Fonte: A autora

Para nós (Figura 41, b.) e arestas (Figura 41, a.), novamente, as produções de 2014 tiveram pouca distinção significativa. Nós não foram expressivos de fluência de leitura entre as corrigidas ($p_{\text{corrigidas}} = 0,05573$) e tampouco entre as originais ($p_{\text{originais}} = 0,06366$). Arestas apresentaram ligeira diferença, mas sem força para serem dadas como, de fato, significativas ($p_{\text{corrigidas}} = 0,04598$ e $p_{\text{originais}} = 0,03001$), uma vez que o teste Dunn não apontou diferenças entre os textos corrigidos e, para os originais, apenas entre ambulatório e medianos ($p_{\text{adj.}} = 0,04635$) - mas, ainda assim, fraca. Em 2015 já houve mais expressividade tanto para nós significativas ($p_{\text{corrigidas}} = 0,003281$ e $p_{\text{originais}} = 0,009398$) quanto para arestas significativas ($p_{\text{corrigidas}} = 0,003281$ e $p_{\text{originais}} = 0,001561$). Finalmente, 2016 confirmou as expectativas construídas com as análises

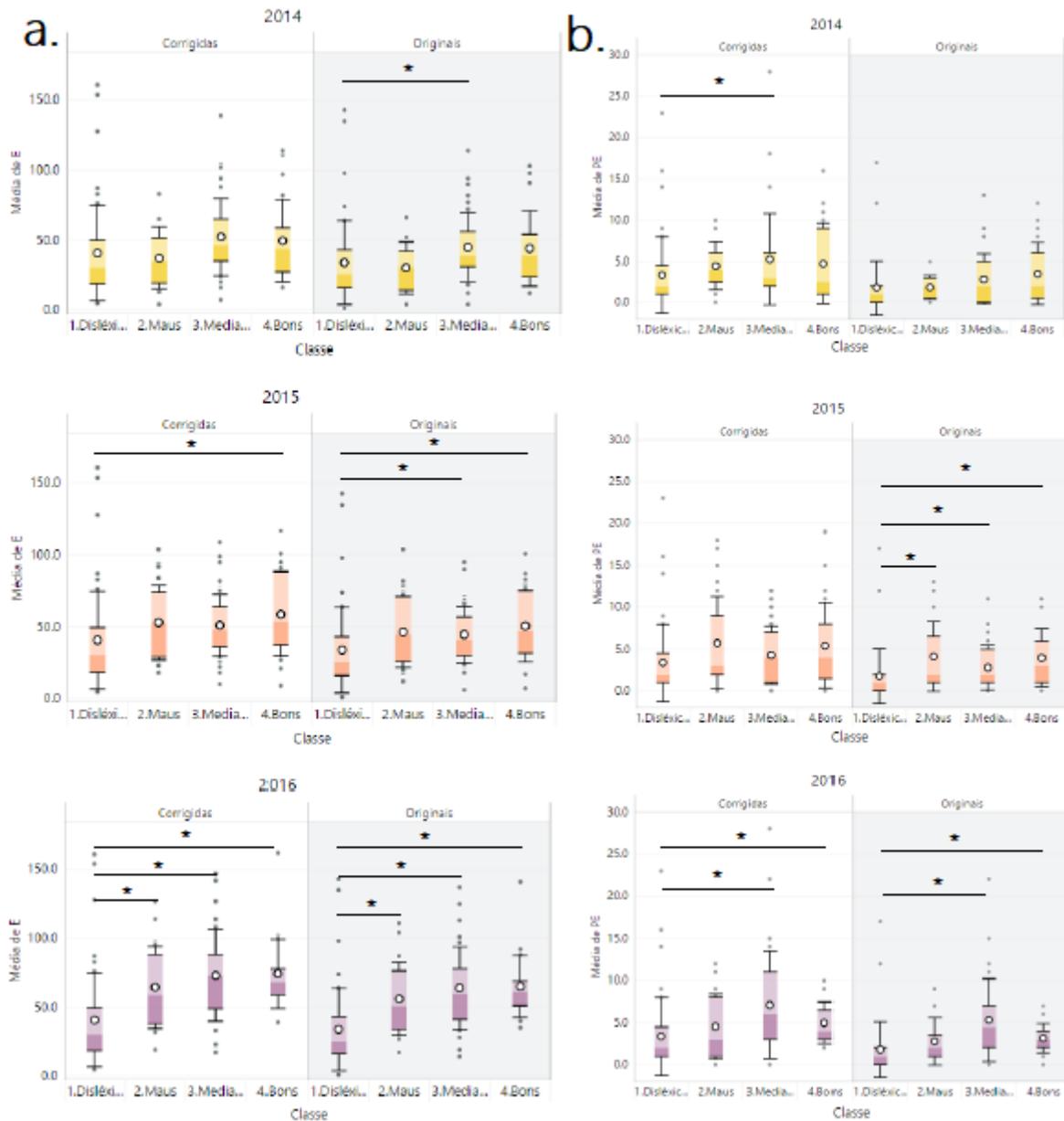
anteriores e revelou valores mais significativos de diferença entre os atributos ($p_{\text{corrigidas}} = 5,808 \times 10^{-7}$ e $p_{\text{originais}} = 2,098 \times 10^{-7}$). Tanto entre as corrigidas quanto entre as originais, quando comparados bons e maus, bons e medianos e maus e medianos, não foram encontradas diferenças significativas. Suspeita-se que o tamanho amostral enxuto tenha favorecido esta baixa distinção mesmo em 2016 - momento em que ela poderia ou deveria ser mais marcada. Além disso, é possível que os critérios de atribuição de fluência leitora não sejam suficientes para distribuir os participantes das escolas em diferentes subgrupos.

Os atributos gerais de grafo e o número de palavras, portanto, mostraram eficácia progressiva na categorização dos textos, acompanhando a progressão da escolaridade das crianças. Isto não só corrobora expectativas encontradas na literatura acerca do desenvolvimento das habilidades de leitura e escrita (GÂNDARA; BEFI-LOPES, 2010; ARAUJO, 2011) como também reitera a relevância já constatada em pesquisas prévias acerca destes atributos, ainda que fossem diferentes os seus escopos de estudo.

Dos atributos de recorrência, aquele com maior significância para cada rodada de análises por ano de coleta foi arestas paralelas, ou PE (Figura 42, b.). Enquanto que L1 e L2 não foram significantes em nenhuma análise, e L3 apresentou diferença significativa somente em 2016 - para corrigidas⁴³ e ($p_{\text{corrigidas}} = 0,02924$) e originais ($p_{\text{originais}} = 0,03558$) -, PE foi significativo em todos os anos. Em 2014, houve significância para corrigidas e originais ($p_{\text{corrigidas}} = 0,02285$ e $p_{\text{originais}} = 0,03759$). Porém, só houve distinção entre disléxicos e medianos para as corrigidas ($p \text{ adj.} = 0,03694$) e, para as originais, não foi obtida nenhuma diferenciação de pares. Em 2015, mais uma vez, a significância obtida para as corrigidas ($p_{\text{corrigidas}} = 0,01381$) não garantiu a diferenciação entre pares; para as originais, ($p_{\text{originais}} = 0,0003236$), foram diferenciados disléxicos e bons leitores ($p \text{ adj.} = 0,004284$), disléxicos e maus leitores ($p \text{ adj.} = 0,003776$) e disléxicos e medianos ($p \text{ adj.} = 0,01737$). Significâncias mais fortes foram obtidas em 2016, que, entre as corrigidas ($p_{\text{corrigidas}} = 0,0002271$), obteve resultados satisfatórios para disléxicos e bons leitores ($p \text{ adj.} = 0,01044$) e disléxicos e medianos ($p \text{ adj.} = 0,0004159$). Para as originais, essa diferenciação se manteve (disléxicos e bons leitores - $p \text{ adj.} = 0,001986$ e disléxicos e medianos - $p \text{ adj.} = 1,165 \times 10^{-5}$).

⁴³ Com distinção somente entre Ambulatório e medianos ($p \text{ adj.} = 0,4583327$), segundo os resultados do teste Dunn.

Figura 42: Arestas (E) e arestas paralelas (PE) (Transversal por ano e fluência de leitura)

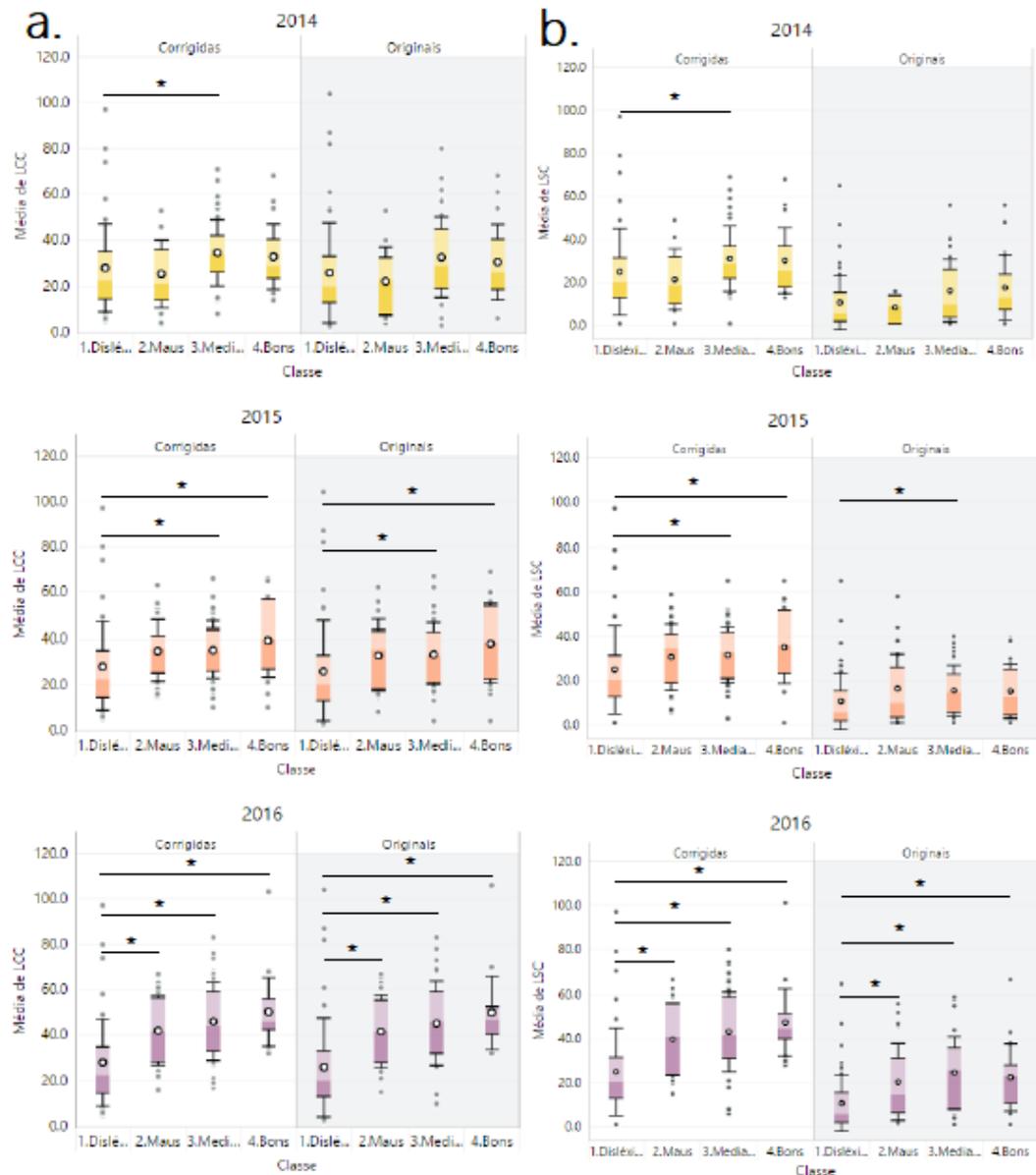


Fonte: A autora

Assim, arestas paralelas se afirmaram como o atributo de recorrência mais eloquente sobre grupos com diferentes categorias de fluência leitora. Evidências encontradas em estudos anteriores apontaram uma correlação negativa entre PE e performance cognitiva (QI e Teoria da Mente) (MOTA et al., 2016). Como os quocientes de inteligência dos participantes não foram controlados para as análises do presente estudo, não são possíveis inferências a este respeito. Pode ser reafirmada, apenas, a baixa expressividade de atributos de recorrência na caracterização de subgrupos experimentais com diferentes fluências de leitura.

Se comparadas as eficácias de atributos de conectividade LCC e LSC (Figura 43, a. e b. respectivamente) na distinção de fluências de leitura, pode-se dizer que os resultados foram semelhantes entre ambos, como ilustram os gráficos:

Figura 43: LCC e LSC (Transversal por ano de coleta e fluência de leitura)



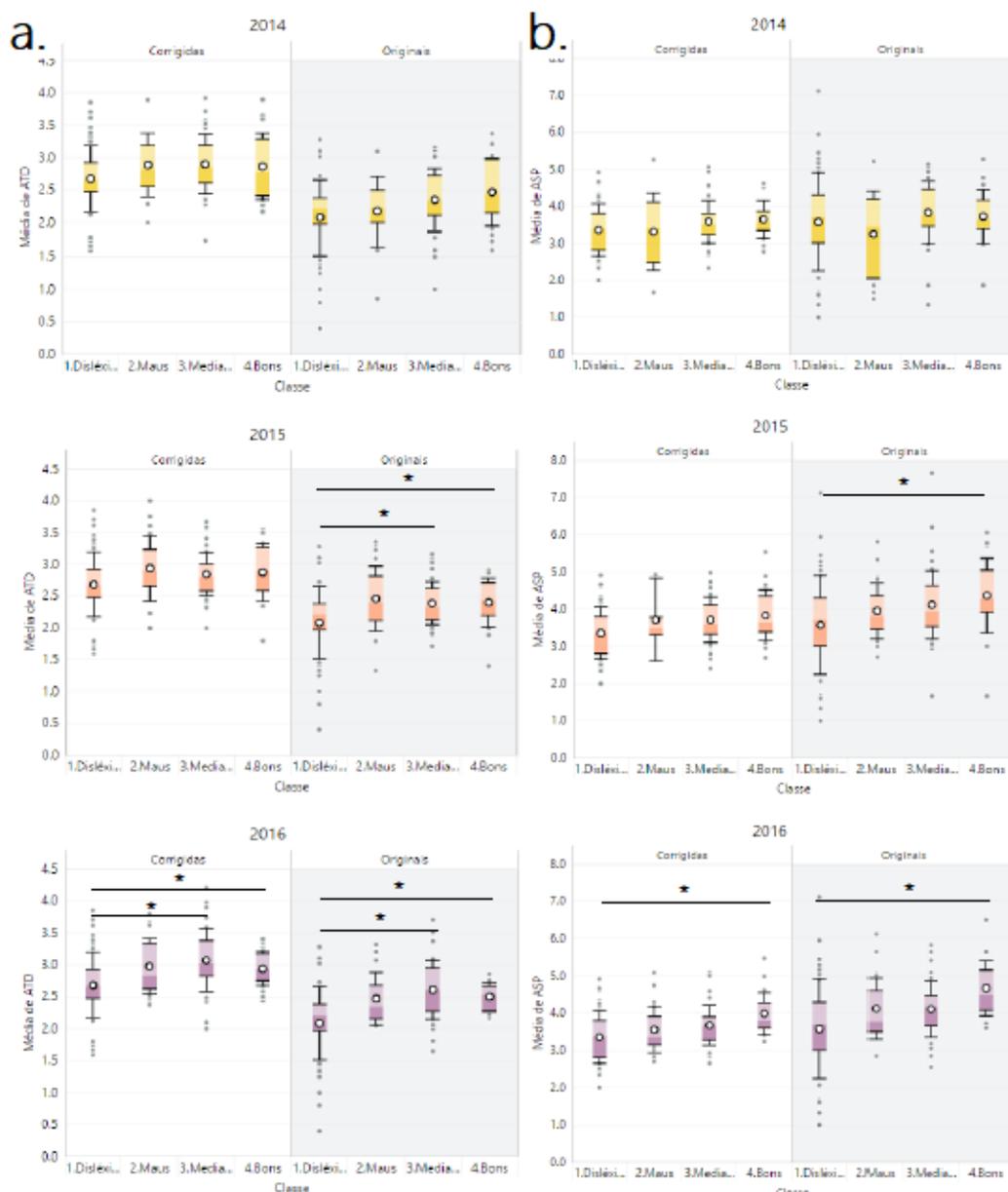
Fonte: A autora

Em 2015, LCC apresentou uma eficácia ligeiramente melhor que a LSC entre as originais, ao conseguir diferenciar disléxicos e bons leitores (LCC p adj. disléxicos – bons = 0,009411 ; LSC p adj. disléxicos – bons = 0,28068). Para os demais anos e análises, ambos apresentaram as mesmas eficácias, sendo a distinção entre disléxicos e medianos a única

recorrente entre todos os casos. Entende-se que este resultado decorra do tamanho amostral de medianos, superior ao tamanho dos demais grupos e, por isso, bastante expressivo nas análises).

Os próximos gráficos exibem as medidas descritivas dos atributos ATD e ASP:

Figura 44: ATD e ASP (Transversal por ano de coleta e fluência de leitura)



Fonte: A autora

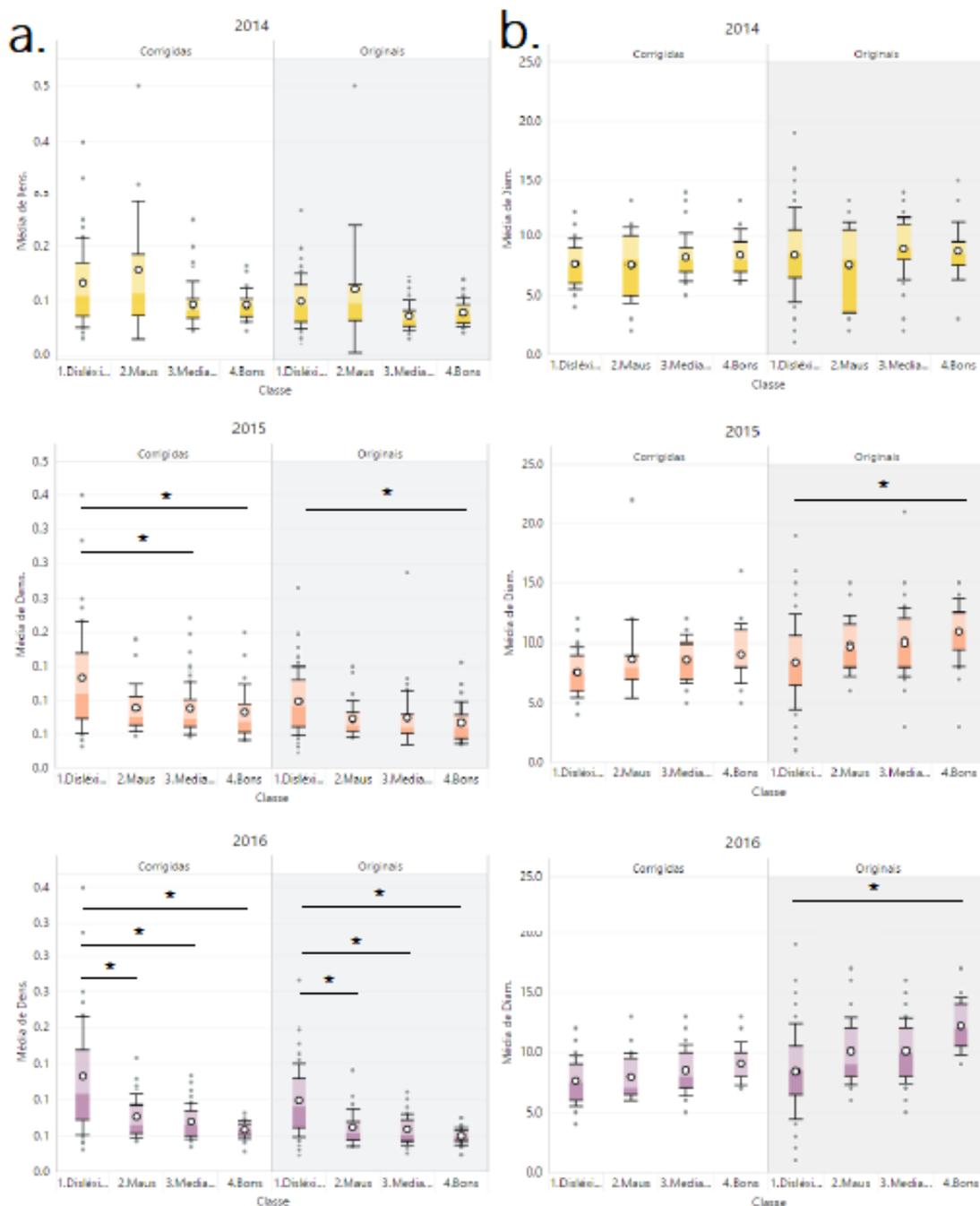
A Figura 44 retrata, mais uma vez, que o avanço da escolaridade garante mais distinção entre os grupos, uma vez que, em 2014, ATD e ASP não foram eficazes na diferenciação das categorias de fluência de leitura, para qualquer que tenha sido o tipo de transcrição – corrigidas ou originais (ATD $p_{\text{corrigidas}} = 0,149$ e ATD $p_{\text{originais}} = 0,0381$; ASP $p_{\text{corrigidas}} = 0,374$ e ASP

$p_{\text{originais}} = 0,462$). Em 2015 já houve melhora tímida neste cenário (ATD $p_{\text{corrigidas}} = 0,0594$ e ATD $p_{\text{originais}} = 0,0038$; ASP $p_{\text{corrigidas}} = 0,039$ e ASP $p_{\text{originais}} = 0,0267$), que adquiriu mais relevância em 2016 (ATD $p_{\text{corrigidas}} = 0,00064$; ATD $p_{\text{originais}} = 4,22 \times 10^{-6}$; ASP $p_{\text{corrigidas}} = 0,0033$ e ASP $p_{\text{originais}} = 0,0013$), quando mais pares de categorias de fluência de leitura foram diferenciados: disléxicos e bons, disléxicos e maus e disléxicos e medianos.

ASP, como já foi mencionado, apresentou correlação positiva com status cognitivo geral e performance funcional em BERTOLA et al. (2014). Os resultados desta rodada de análise (Figura 44, b.) indicam que pacientes com características de ordem cognitiva associadas ao prejuízo de habilidades de leitura foram os com menor médias de ASP em 2015 e 2016, apresentando consonância com o estudo citado, ainda que as populações e os escopos de estudo sejam diferentes. ATD, no mesmo estudo, foi inexpressivo de diferença de status cognitivo geral e performance funcional, entre os idosos. Mas, neste estudo, seus resultados de significância, bem como suas médias, indicaram um comportamento semelhante, uma vez que ele aumentou quando estavam em questão melhores habilidades de leitura (Figura 44, a.).

Finalmente, densidade (Figura 45, a.) se comportou de maneira similar à retratada para ATD e ASP. Em 2014, as análises estatísticas realizadas não revelaram diferenças significativas para dens. e diam. entre os grupos (Dens. $p_{\text{corrigidas}} = 0,0915$ e Dens. $p_{\text{originais}} = 0,1147$; Diam. $p_{\text{corrigidas}} = 0,0670$ e Diam. $p_{\text{originais}} = 0,0145$). Estes potenciais de distinção aumentaram com o avanço dos anos para dens., atingindo significância especialmente na distinção de disléxicos e bons leitores, em 2016, tanto para corrigidas (Dens. $p_{\text{adj.}} = 0,00029$) quanto para originais (Dens. $p_{\text{adj.}} = 0,000298$). Diâmetro, por sua vez, mostrou-se progressivamente significativo somente entre originais, sobretudo em 2016, quando diferenciou disléxicos e bons leitores (Diam. $p_{\text{adj.}} = 0,0001$) significativamente, segundo os resultados do teste Dunn aplicado. Estes resultados ilustram a prevalência constatada da densidade sobre o diâmetro, nas análises realizadas. Ambos são atributos globais, porém com diferentes relevâncias para a predição de fluência de leitura.

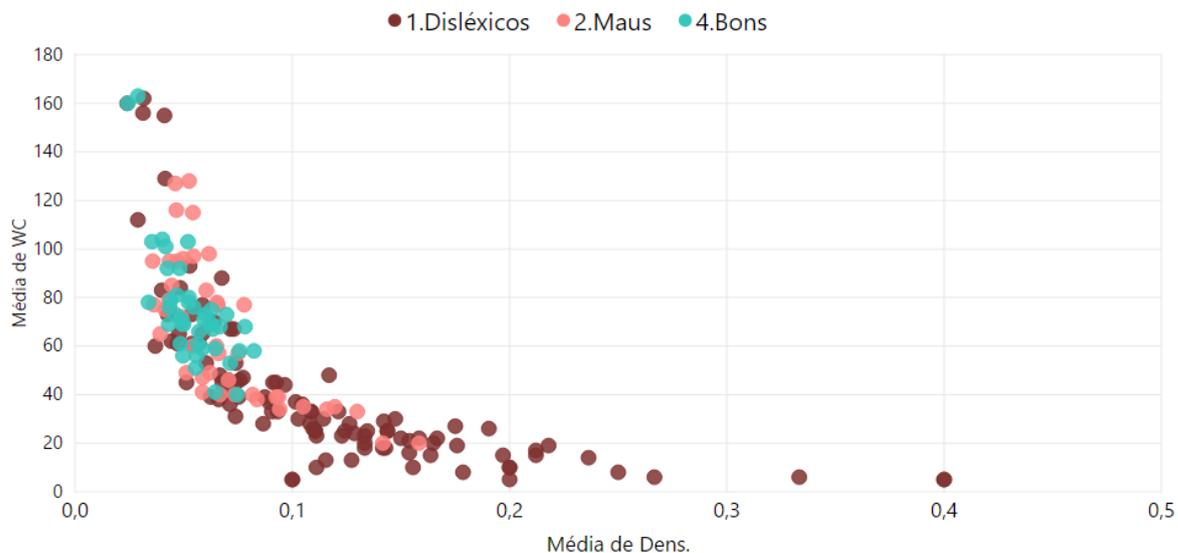
Figura 45: Dens. e Diam. (Transversal por ano de coleta e fluência de leitura)



Fonte: A autora

Em resumo, com base nos atributos de grafo extraídos dos textos originais de 2016, medidas de profusão lexical (WC) e espessura textual (Dens.) de uma mesma categoria de fluência de leitura, quando comparadas, revelaram que textos com maiores números de palavras tendem a apresentar baixa densidade – isto é, trata-se de dois atributos cujas variações ocorrem em sentido opostos (Figura 46).

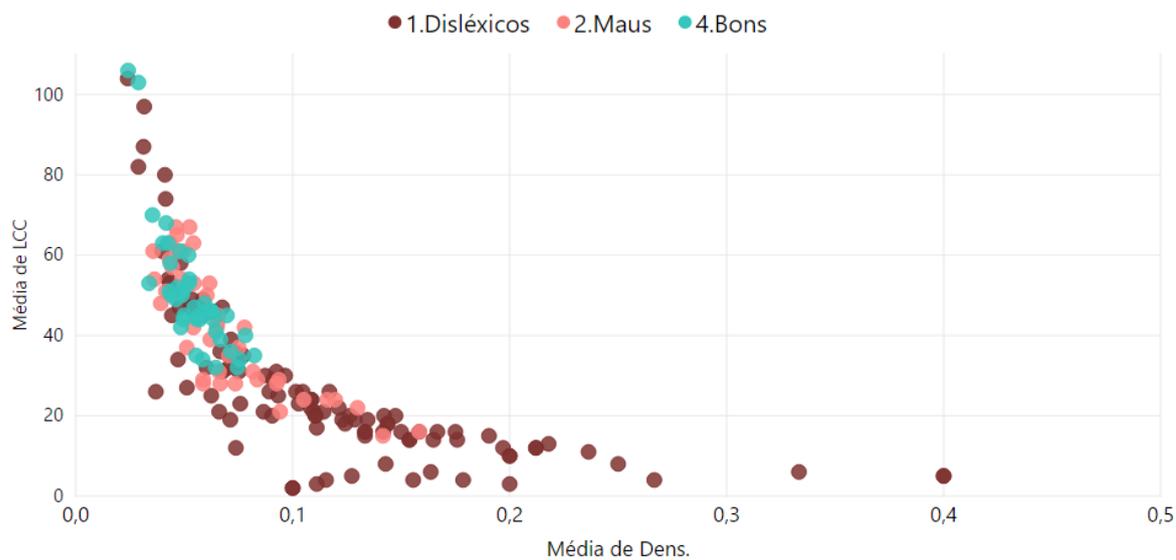
Figura 46: Relação entre médias de WC e Dens.



Fonte: A autora

Da mesma forma, textos com maiores agrupamentos de itens lexicais (LCCs), segundo as análises, são acompanhados por menores espessuras textuais (Dens.), como retrata a Figura 47:

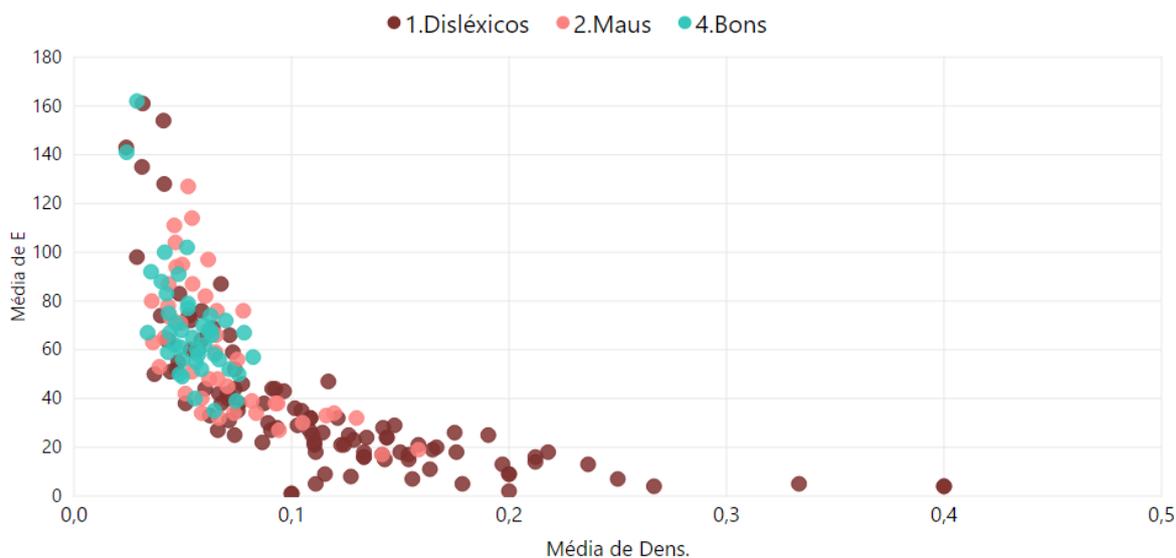
Figura 47: Relações entre médias de LCC e Dens.



Fonte: A autora

Igualmente relevante e diretamente ligada à discussão das relações acima, é a compreensão de que o aumento de relações entre itens lexicais de um texto (E) parece colaborar para que sua espessura diminua, tornando-o mais inteligível, como ilustra a Figura 48:

Figura 48: Relações entre as médias de E e Dens.



Fonte: A autora

3.1.6 Teias de médias de atributos

Como dito anteriormente, textos são como tecidos, uma vez que se comportam como produtos de diferentes fios cuja espessura e maleabilidade e dependem da qualidade e da quantidade destes fios, da maneira como estes estão entrelaçados e organizados a fim de, juntos, formar um todo maior. Comparam-se, neste estudo, os fios de um texto às relações de seus itens lexicais, que podem ser abordados de maneira qualitativa ou quantitativa. Neste trabalho, investiga-se quantitativamente a tessitura de produções textuais de crianças com diferentes níveis de fluência de leitura, a partir de atributos de grafo.

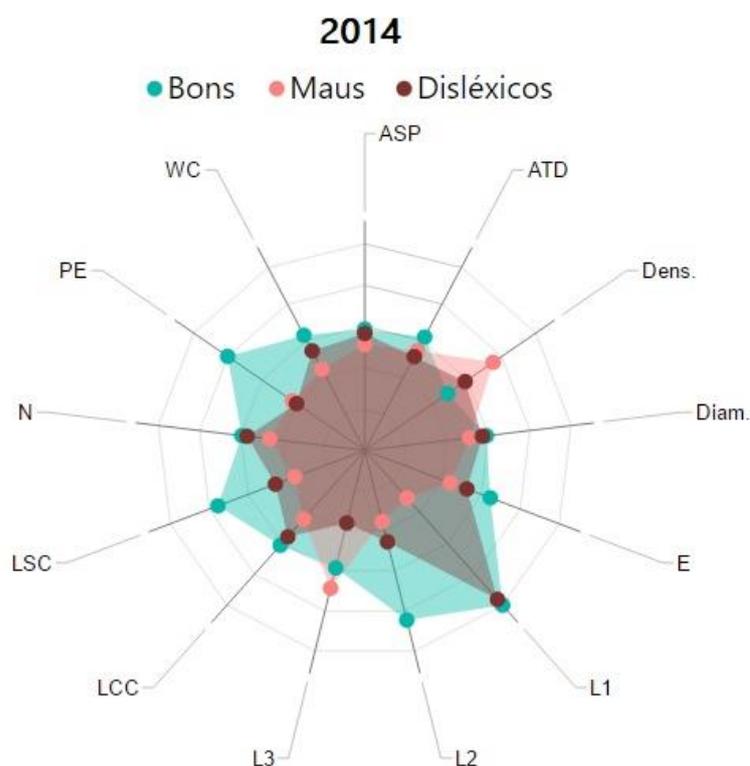
A fim de unir analogia e análise estatística, conclui-se esta discussão acerca das análises estatísticas descritivas e inferenciais realizadas com a apresentação das médias de atributos de

grafo de cada uma das três principais⁴⁴ categorias - bons leitores, maus leitores e disléxicos – em três teias – uma para cada ano de coleta. Entende-se que esta representação é pertinente tanto por oportunizar uma visão mais ampla da produção dos atributos quanto por dialogar com a própria atividade de “tecer textos”.

O cálculo que orientou a criação das teias foi a razão entre a média de cada atributo e a média geral⁴⁵ daquele atributo para cada ano, uma vez que houve a necessidade de normalizar os valores de médias para que todos tivessem uma base decimal comum e, assim, fosse possível uma melhor representação gráfica de todos juntos.

A inspeção das três teias oportuniza a visualização da distinção progressiva não só dos três subgrupos – bons leitores, maus leitores e disléxicos -, como também dos três anos de coleta, como mostra a teia de 2014:

Figura 49: Teia de médias de atributos de 2014



⁴⁴ A categoria dos leitores medianos foi descartada desta discussão final por ter pouca relevância teórico-prática dentro do escopo deste trabalho e, também, por conta de seu tamanho amostral, significativamente maior que os demais – o que pode ter enfraquecido os resultados de parte das análises inferenciais.

⁴⁵ As médias de bons leitores, maus leitores e disléxicos foram somadas e divididas por três. O resultado dessa operação serviu de base para a normalização dos dados.

A teia de 2014 revela contornos nebulosos que, em certa medida, sobrepõem-se. Neste momento, os alunos estavam no segundo ano do Ensino Fundamental, portanto haviam recentemente sido alfabetizados. Ainda que a amplitude da teia dos bons leitores seja maior que a dos maus leitores e disléxicos, nota-se que, para certos atributos, as médias foram muito similares. ATD, ASP, Diam., LCC e N são atributos que, como já fora comentado, mostraram-se distintivos de categorias de fluência de leitura ao longo dos três anos. No entanto, neste primeiro momento de coleta, os três grupos os produziram de maneira similar. Bons leitores destacaram-se na produção de PE, L2 e LSC – dois atributos de recorrência e um atributo de conectividade respectivamente. Mas, se forem considerados os valores de significância obtidos nas análises estatísticas das produções originais não normalizadas de 2014 (representadas por meio destas teias), será constatado que apenas PE foi significativamente diferente entre os três subgrupos. Logo, as teias refletem as médias descritivas dos atributos, não necessariamente as suas significâncias.

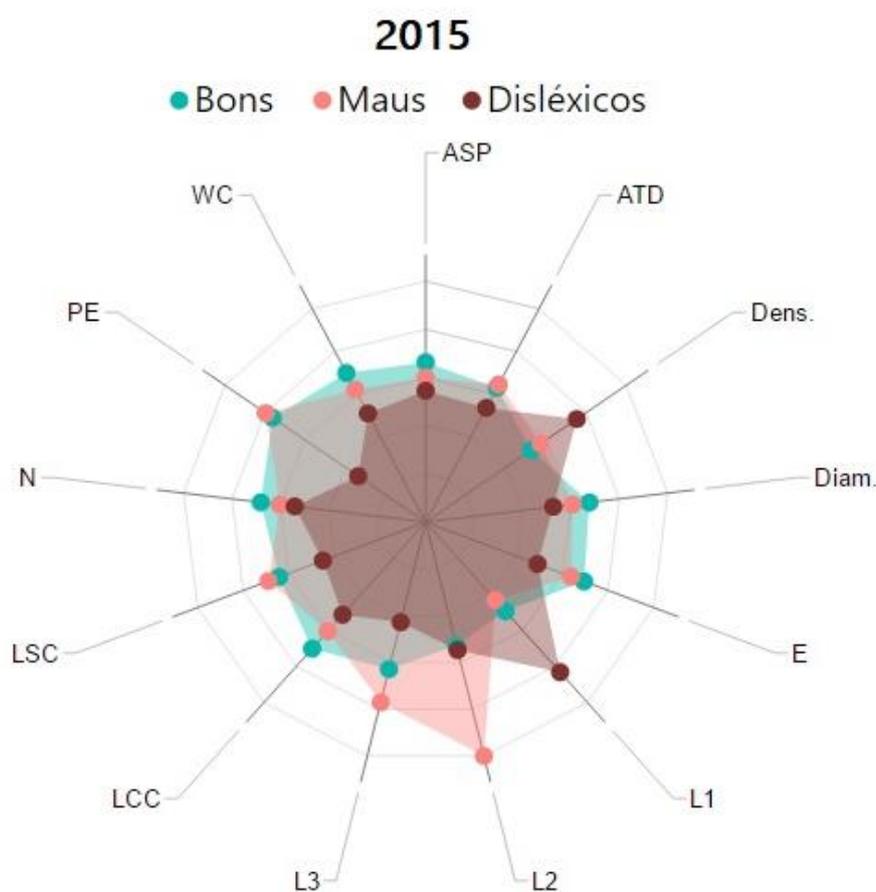
Os maus leitores apresentaram as menores produções médias para nove das 13 medidas analisadas: ASP, Diam., E, L1, L2, LCC, LSC, N e WC. Considerando o fato de que os maus leitores das escolas não passaram por um protocolo de diagnóstico de dislexia, é possível que, dentro deste grupo, haja crianças disléxicas. Esta hipótese ganha força se for destacado que a média mais alta de densidade foi a dos maus leitores, e não a dos disléxicos. A densidade, como já foi dito, mostrou-se um marcador do transtorno nas análises estatísticas. Foi observada maior média, entre os maus leitores, para L3 também. Contudo, ambos os atributos – L3 e Dens. - não foram significativamente diferentes entre os grupos nas análises de 2014. O que indica que, sim, houve maior recorrência destas medidas entre maus leitores, porém não a ponto de indicar um padrão gráfico a ser atribuído ao grupo, capaz de o distingui-lo dos demais.

As médias dos disléxicos para a maioria dos atributos apresentou uma variação muito parecida com a dos maus leitores, sendo ora ligeiramente mais baixa que a deste subgrupo das escolas – especificamente, para ATD, Dens. e PE -, ora ligeiramente mais alta – para ASP, Diam., E, L2, LCC, LSC, N e WC. Foi observada uma diferença maior entre ambos os grupos, para duas medidas de grafo: L1 e L3. Disléxicos produziram uma média de L1 bem superior à de maus leitores; estes apresentaram uma média de L3 maior que a média dos disléxicos para o mesmo atributo. Mas, como se sabe, ambas as medidas de recorrência não foram expressivas de diferenças entre as crianças.

Na teia de médias de atributos de 2015, é representada maior diferenciação entre os subgrupos maus leitores e disléxicos e menor distinção entre maus leitores e bons leitores. Em

uma primeira análise, esta diferença contraria a expectativa criada com a teia de 2014 de que a notável proximidade entre o grupo maus leitores e o grupo disléxicos refletiria a presença de crianças com dislexia não diagnosticada entre as más leitoras. Contudo, acredita-se que o ano de 2014, por ser o primeiro ano posterior à alfabetização das crianças, seja caracterizado por padrões de conectividades pouco delineados, em processo de consolidação. Assume-se que facilidades e dificuldades iniciais, particulares de cada criança e não necessariamente restritas ao universo dos transtornos de aprendizagem, configurem a teia de maneira caótica, assim como se constituem os subgrupos das escolas. A atribuição de fluência leitora às crianças, portanto, tem de ser relativizada para este ano. Eis a teia de 2015:

Figura 50: Teia de médias de atributos de 2015



Fonte: A autora

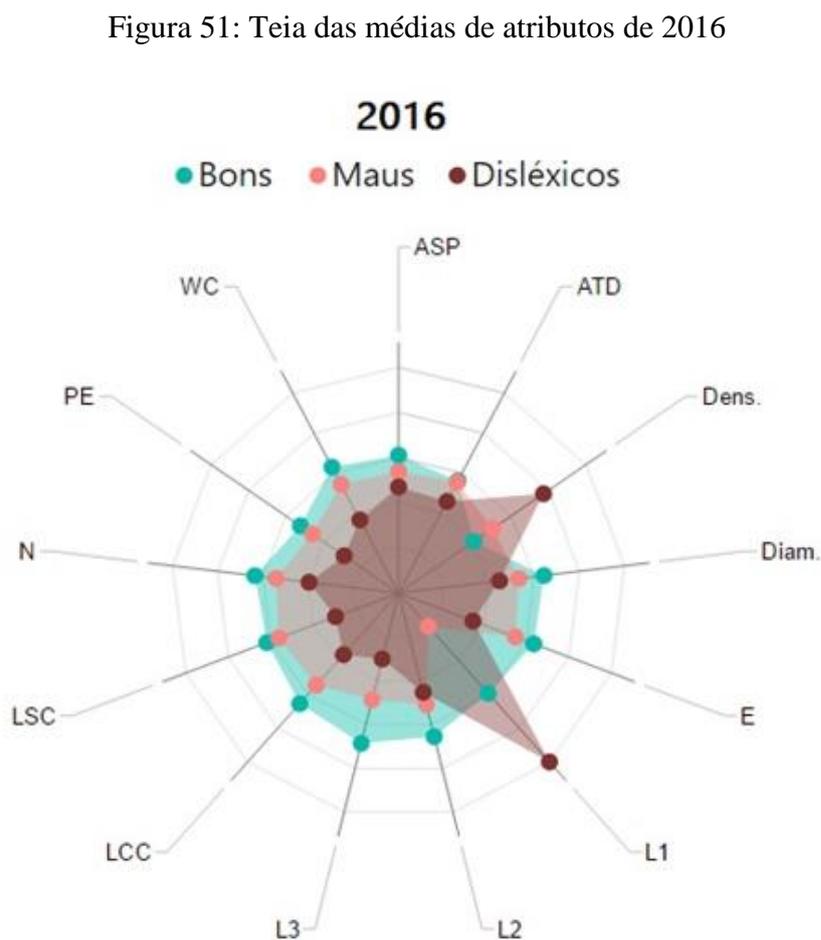
Observa-se que bons leitores começaram a construir textos mais balanceados em 2015 – ou teias textuais mais equilibradas entre seus fios de atributos. Neste aspecto, bons leitores

distinguiram-se dos demais, pois tanto maus leitores quanto disléxicos permaneceram apresentando picos de produções médias de alguns atributos.

Maus leitores, surpreendentemente, pareceram reverter as médias de 2014 para L2 e conservar as de L3. Acredita-se que esta diferença para L2 reflita idiosincrasias dos maus leitores, uma vez que eles não foram os mesmos em cada ano de coleta e, principalmente, que ciclos de dois nós não registram, necessariamente, problemas ortográficos, gramaticais ou pobreza vocabular, como já fora discutido.

A teia também ilustra o começo da acentuada densidade dos textos de disléxicos – esta, sim, com diferença significativa constatada. O grupo também apresentou médias altas para L1, mas sem significância. Ciclos de um nó, como já fora discutido, não se mostraram preditores de fluência leitora.

Eis a teia final, de 2016:



Fonte: A autora

A teia de 2016 atende às expectativas do presente estudo: apresenta diferenciação dos grupos e, especificamente, retrata complexidade e equilíbrio textual progressivos entre as três categorias de fluência leitora – crescentes dos menos fluentes para os mais fluentes. Na Figura 51, fica nítida a progressão entre os grupos, sendo o grupo dos disléxicos o com menor amplitude ao longo da teia, exceto para dois atributos: Dens. e L1.

Maus leitores apresentaram as menores médias de L1 nos três anos de coleta, ao mesmo tempo em que esta constância não foi percebida para L2 e L3. Este resultado reitera a hipótese de que ciclos dizem muito pouco a respeito da fluência leitora dos participantes, por serem motivados por diferentes questões – nenhuma delas restrita ao universo das dificuldades de produção textual vividas por disléxicos.

Outro resultado interessante é observado no eixo da densidade. O aumento de médias que acompanha o aumento da fluência leitora dos participantes não se aplica à densidade. Pelo contrário, as análises estatísticas revelaram que, quanto melhor o desempenho na leitura das crianças, menor é a espessura de seus textos. Em outras palavras, quanto mais fluente e desenvolta a criança no exercício da leitura, menos condensados são seus textos, no sentido de serem menos aglutinados e mais maleáveis.

Logo, a análise das três teias evidenciou as seguintes diferenças entre as categorias de fluência leitora:

- Os bons leitores, nos três anos, apresentaram melhor distribuição de atributos de grafo que os demais, uma vez que se estenderam entre todos os fios da teia de maneira mais equilibrada, sobretudo em 2016;
- Os maus leitores apresentaram diferença levemente acentuada para alguns atributos, em todos os anos. Em 2014, produziram médias mais altas de L3 e densidade; em 2015, produziram médias também mais altas para L2 e L3 e, em 2016, apresentaram queda aparente na produção de L1;
- Os disléxicos, se comparados com as demais categorias, sempre apresentaram as teias com pior distribuição, com produções mais acentuadas tanto para L1 quanto para densidade, em todos os anos.

Igualmente, a análise comparativa das conectividades observadas entre os anos sugere que, ao longo do tempo de coleta:

- Tanto bons leitores quanto maus leitores progressivamente equilibraram a conectividade de seus textos, garantindo uma distribuição similar dos atributos sobre suas teias – em especial bons leitores.
- Disléxicos, por outro lado, mostraram-se constantemente desequilibrados neste sentido, com picos de produção média de densidade e L1 em todos os anos.

Estes resultados sugerem que, quanto maior a fluência de leitura, mais equilibrada é a produção média dos componentes textuais que se refletem em atributos de grafo, como itens lexicais (ou nós) e ligações entre itens lexicais (ou arestas).

3.2 RESULTADOS DO CLASSIFICADOR SVM

Como mencionado previamente, optou-se por analisar as medidas de atributos de grafo extraídas dos textos pelo software *Speech Graphs*. Para tanto, tais medidas serviram como *input* para classificadores SVM, que se basearam em quatro métricas de Aprendizado de Máquina: acurácia (índice de frequência de precisão do classificador), revocação ou sensibilidade (fração de instâncias relevantes que são recuperadas), precisão ou valor preditivo positivo (fração de instâncias recuperadas que são relevantes) e F1 Score (índice da qualidade geral do modelo obtido pela medida harmônica entre precisão e revocação).

O classificador SVM foi treinado com um kernel RBF, organizando os dados de forma *leave one out* para o treinamento. Índices das métricas superiores a 70% foram considerados indicadores de predição, de modo que, quanto mais próximos de 100%, maior o potencial do classificador de prever a fluência de leitura das amostras; quanto mais próximos de 70% os índices, menores os seus potenciais de predição das diferenças analisadas.

Observa-se, a seguir, que as métricas utilizadas nas análises revelaram baixos índices de predição dos classificadores utilizados, sobretudo para as versões originais dos textos.

Tabela 16: Resultados dos classificadores SVM para originais

ORIGINAIS				
Bons x Disléxicos				
Ano	Acurácia	Revocação	Precisão	F1 Score
2014	0.61	0.57	0.61	0.55
2015	0.53	0.53	0.53	0.53
2016	0.72	0.73	0.72	0.72
Maus x Disléxicos				
Ano	Acurácia	Revocação	Precisão	F1 Score
2014	0.58	0.58	0.58	0.58
2015	0.56	0.56	0.56	0.56
2016	0.64	0.64	0.64	0.64
Bons x Maus				
Ano	Acurácia	Revocação	Precisão	F1 Score
2014	0.52	0.42	0.28	0.34
2015	0.27	0.25	0.18	0.21
2016	0.6	0.6	0.61	0.58

Já os resultados obtidos nas análises das versões corrigidas foram os seguintes:

Tabela 17: Resultados dos classificadores SVM para corrigidas

CORRIGIDAS				
Bons x Disléxicos				
Ano	Acurácia	Revocação	Precisão	F1 Score
2014	0.5	0.45	0.37	0.37
2015	0.64	0.63	0.63	0.63
2016	0.83	0.84	0.85	0.83
Maus x Disléxicos				
Ano	Acurácia	Revocação	Precisão	F1 Score
2014	0.62	0.62	0.63	0.61
2015	0.62	0.61	0.63	0.61
2016	0.7	0.7	0.7	0.7
Bons x Maus				
Ano	Acurácia	Revocação	Precisão	F1 Score
2014	0.58	0.47	0.3	0.37
2015	0.37	0.35	0.22	0.27
2016	0.26	0.26	0.23	0.24

As tabelas reportam valores assumidos neste estudo como indicadores de predição apenas para a distinção entre bons leitores e disléxicos de 2016 (tanto na análise das originais quanto na das corrigidas), dado este que apresenta consonância com as teias discutidas anteriormente. Contudo, a acurácia obtida para as corrigidas foi superior à encontrada para as originais – 83% e 72% respectivamente. Se consideradas a soma de ocorrências de significância dos atributos de grafo, para as quatro rodadas de análises realizadas⁴⁶, conclui-se que as versões originais, independentemente do ano de coleta, permitem a extração de atributos de grafo mais expressivos de diferenças entre as produções textuais dos participantes. Contudo, a “expressão das corrigidas” não ficou muito aquém neste sentido para o ano de 2016, se comparada com a das originais⁴⁷. Logo, entende-se que a diferença de acurácia obtida com os classificadores se deve ao fato de que o texto corrigido torna a medida de grafo mais confiável, porque

⁴⁶ Originais, corrigidas, originais normalizadas e corrigidas normalizadas.

⁴⁷ As corrigidas tiveram 10 atributos de grafo com significância e as originais, 11. O único atributo que não foi significativo para as corrigidas tendo o sido para as originais foi diâmetro. Esta informação está disponível na Tabela 14.

“normaliza” o léxico e garante, assim, a integridade de uma mesma representação semântica – que é perdida nas transcrições originais, nas quais *fósforo* e *foforo* são escritos em um mesmo texto e o classificador entende que são representações semânticas diferentes. Representações corrigidas, seguindo esta lógica, permitem uma melhor análise discursiva, não ortográfica.

Cabe salientar que a expectativa era a de que o classificador se mostrasse mais preciso para a distinção dos grupos nas demais análises de 2016. No entanto, isto não ocorreu, uma vez que os índices das quatro métricas se mantiveram próximos a 60% para a distinção entre maus leitores e disléxicos e próximos a 25% pra a distinção entre bons leitores e maus leitores, o que foi extremamente ruim.

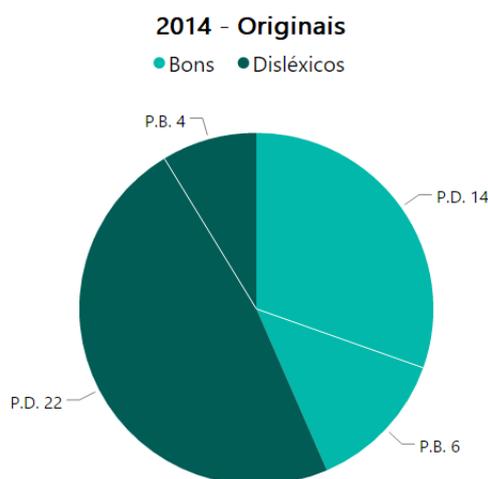
Possíveis justificativas são levantadas para este resultado:

- O grupo disléxicos foi balanceado aleatoriamente, o que abre margem para a possibilidade de terem se concentrado na amostra utilizada de 50% (N=26) do montante original a fração de crianças da amostra com crianças com nível baixo de dislexia; isto explicaria a razão de maus e disléxicos terem sido distinguidos pelo classificador com uma acurácia de apenas 60%;
- Por outro lado, é possível que a acurácia do classificador tenha sido baixa para a diferenciação de maus leitores e disléxicos porque é alta a probabilidade de crianças sem o diagnóstico de dislexia integrarem o subgrupo dos maus leitores, o que o aproximaria do grupo de disléxicos;
- Os critérios de categorização de fluência leitora adotados talvez não sejam eficientes na distinção de bons e maus leitores; esta limitação justificaria a baixíssima acurácia do classificador na distribuição de bons e maus leitores entre ambas as categorias – corrigidas e originais - e entre todos os anos;
- Não é descartada a possibilidade de que a alta variância nos dados, isto é, todas as métricas de grafos utilizadas, somadas ao WC, não sejam boas preditoras de acuidade de leitura para o tamanho de conjunto de dados trabalhados; logo, assume-se que o baixo número de exemplos analisados ou, fundamentalmente, a própria métrica de grafos talvez tenham impedido uma predição mais acurada;
- Finalmente, há grandes chances de que o modelo do classificador adotado não seja eficaz na generalização destes dados, neste contexto de pesquisa. Outros modelos talvez apresentassem melhor acurácia – ex.: árvore de decisão.

A fim de ilustrar a origem das inferências acerca da acurácia do classificador - métrica que melhor responde nossa principal pergunta de pesquisa⁴⁸ -, as matrizes de confusão das análises realizadas para cada ano e versão (original e corrigida) são, a seguir, apresentadas lado a lado, com o propósito de contrastá-las. Elegeu-se de cada rodada de análise as duas matrizes de confusão com maiores acurácias – uma das originais, outra das corrigidas -, cujos resultados foram reportados em gráficos de setores, que introduzem as discussões acerca de cada ano. Setores do gráfico identificados como “P.B.,” “P.M.” e “P.D.” referem-se, respectivamente, à “Predição para Bons Leitores”, “Predição para Maus Leitores” e “Predição para Disléxicos”. Cumpre lembrar que, nas matrizes, a diagonal principal (sempre em cinza e amarelo mais escuros) é o principal foco de interesse. Nela concentram-se os valores preditivos positivos (ou a chamada precisão).

Em 2014, foram analisadas 20 crianças com boa fluência leitora (15 meninas e 6 meninos), 13 crianças com má fluência leitora (8 meninas e 5 meninos) e 26 crianças com dislexia (7 meninas e 19 meninos). Eis os resultados deste ano:

Figura 52: Gráfico de predição de 2014 - Originais



Fonte: A autora

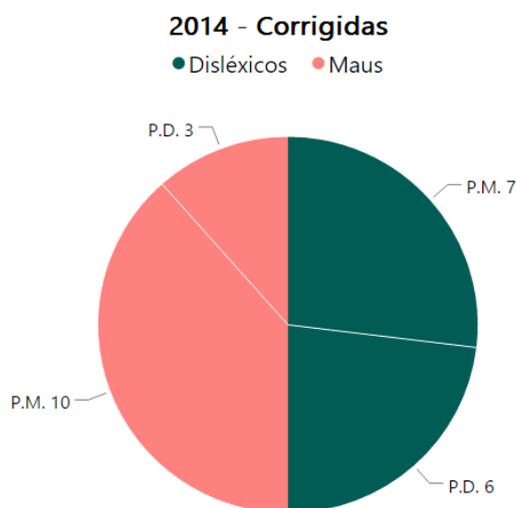
Verifica-se que, na distinção dos atributos extraídos das versões originais para bons leitores e disléxicos, as taxas de erro foram bem inferiores às ideais. Dos 20 bons leitores, apenas 30% (N=6)

⁴⁸ Por principal pergunta de pesquisa, entende-se: os atributos de grafo extraídos de produções textuais de crianças do Ensino Fundamental são preditivos da dislexia do desenvolvimento?

deles foram classificados corretamente. Dos disléxicos, 15,3% (N=4) deles foram categorizados como bons leitores.

Já a acurácia mais alta obtida em 2014, para as corrigidas, foi na distinção entre maus leitores e disléxicos – ainda assim ruim. Especificamente para esta análise, os dados do ambulatório foram balanceados de maneira a reunir uma amostra aleatória de 25% dos participantes (N=13; 3 meninas e 10 meninos), uma vez que o total de maus leitores de 2014 é menor, composto somente por 13 crianças. Sendo assim, dos 13 disléxicos, mais da metade das crianças (53,8%, isto é, 7) foram consideradas más leitoras pelo classificador. Como ilustra a Figura 53, entre os maus leitores de 2014, esta taxa de erro foi menor: 23,07% das 13 crianças com má fluência leitora (N=3) foram consideradas disléxicas:

Figura 53: Gráfico de predição de 2014 - Corrigidas



Fonte: A autora

A tabela com as matrizes de confusão a seguir inclui estes e os demais resultados:

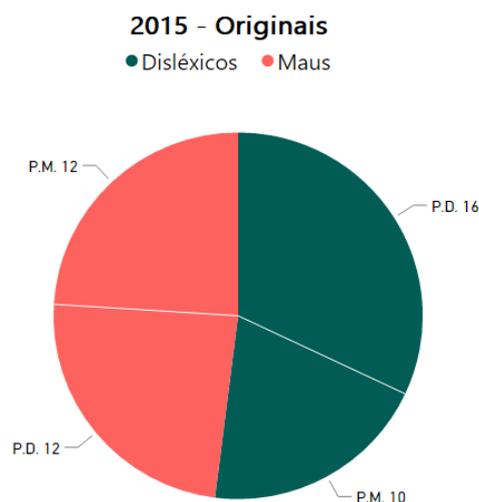
Tabela 18: Matrizes de confusão de 2014

BONS X DISLÉXICOS					
Originais Acurácia: 60,87%			Corrigidas Acurácia: 50%		
Pred./Categ.	Bons	Disléc.	Pred./Categ.	Bons	Disléc.
B	6	4	B	1	4
D	14	22	D	19	22
MAUS X DISLÉXICOS					
Originais Acurácia: 57,69%			Corrigidas Acurácia: 61,54%		
Pred./Categ.	Maus	Disléc.	Pred./Categ.	Maus	Disléc.
M	7	5	M	10	7
D	6	8	D	3	6
BONS X MAUS					
Originais Acurácia: 51,52%			Corrigidas Acurácia: 57,58%		
Pred./Categ.	Bons	Maus	Pred./Categ.	Bons	Maus
B	17	13	B	19	13
M	3	0	M	1	0

Os resultados do classificador dialogam com os resultados estatísticos encontrados para o ano de 2014. As produções textuais deste ano e seus respectivos atributos de grafo mostraram-se, mais uma vez, semelhantes – possivelmente em decorrência do fato das crianças terem sido alfabetizadas em um período próximo ao do momento da coleta, o que pode ter atenuado o abismo entre as performances de leitura e escrita de disléxicos e de leitores típicos e com dificuldades. As diagonais das matrizes de confusão corroboram esta ideia ao trazerem resultados interessantes, como os obtidos para a categorização de bons e maus leitores. Tanto na análise das originais quanto na análise das corrigidas, nenhum dos maus leitores foi classificado corretamente; todos os 13 foram preditos bons leitores, o que pode ser interpretado como um indicativo de homogeneidade entre o grupos. Entre maus leitores e disléxicos e bons leitores e disléxicos, aparentemente, esta homogeneidade diminuiu, ainda que a acurácia se mantenha bem baixa.

Seguem os resultados de 2015, momento em que foram analisados os atributos de grafo de 21 bons leitores (9 meninas e 12 meninos), 24 maus leitores (14 meninas e 10 meninos) e 26 disléxicos (7 meninas e 19 meninos):

Figura 54: Gráfico de predição de 2015 - Originalis

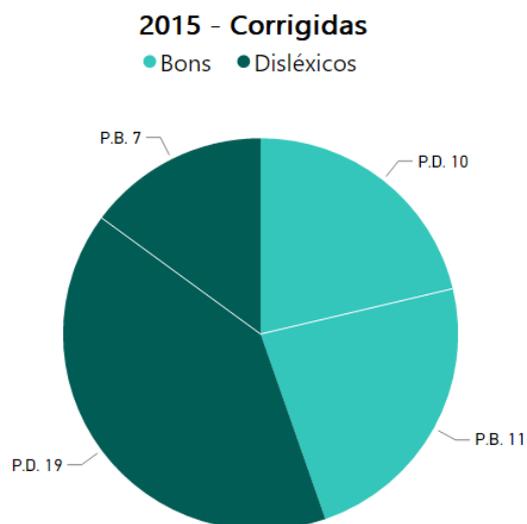


Fonte: A autora

Nota-se, novamente, uma baixíssima acurácia. Quando realizada a distinção entre maus leitores e disléxicos, 50% dos primeiros (12) foram classificados como disléxicos pelo classificador – o que consiste em uma alta taxa de erro -, enquanto que 38,4% (N=10) dos 26 participantes com diagnóstico de dislexia foram considerados maus leitores. Conforme já exposto anteriormente, esta baixa diferenciação entre maus leitores e disléxicos não contraria, de todo, as expectativas do presente estudo. Reafirma-se a possibilidade de parte dos integrantes do grupo maus leitores serem disléxicos, porém não diagnosticados.

A observação do gráfico de setores da distinção entre bons leitores e disléxicos de 2015, a partir dos atributos extraídos das transcrições corrigidas, indica que, uma vez mais, o potencial de prever a dislexia do desenvolvimento, por meio do aprendizado de máquina foi insatisfatório. A Figura 55 retrata uma taxa de erro de 47,6% (N=10) para a predição de crianças com boa fluência leitora e uma de 26,9% (N=7) para a predição de crianças disléxicas:

Figura 55: Gráfico de predição de 2015 - Corrigidas



Fonte: A autora

Resultados complementares, bem como os mencionados, são apresentados na tabela abaixo, com as matrizes de confusão de 2015:

Tabela 19: Matrizes de confusão de 2015

BONS x DISLÉXICOS					
Originais Acurácia: 53,19%			Corrigidas Acurácia: 63,83%		
Pred./Categ.	Bons	Disléxicos	Pred./Categ.	Bons	Disléxicos
B	10	11	B	11	7
D	11	15	D	10	19

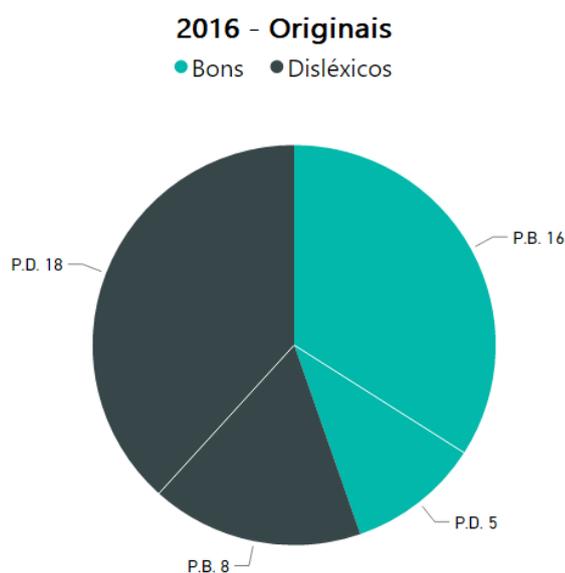
MAUS x DISLÉXICOS					
Original Acurácia: 56%			Corrigidas Acurácia: 62%		
Pred./Categ.	Maus	Disléx.	Pred./Categ.	Maus	Disléxicos
M	12	10	M	13	6
D	12	16	D	11	20

BONS x MAUS					
Originais Acurácia: 26,67%			Corrigidas Acurácia: 37,78%		
Pred./Categ.	Bons	Maus	Pred./Categ.	Bons	Maus
B	0	12	B	0	7
M	21	12	M	21	17

De todos os índices de baixa acurácia novamente obtidos, destacam-se os últimos, acerca da distinção entre bons e maus leitores. Em 2014, para esta distinção, o classificador não havia identificado nenhum mau leitor corretamente. Em 2015, o classificador não classificou nenhum bom leitor como tal, independentemente do tipo de transcrição sob análise; originais e corrigidas, novamente, revelaram similaridade entre os atributos de grafo de ambos os grupos. Este resultado reforça a ideia de se repensar sejam os critérios de atribuição de fluência leitora ou o próprio classificador, uma vez que, para 2014, esta homogeneidade era justificada, mas, para 2015, a expectativa era a de que houvesse uma melhora na acurácia do classificador, dado o avanço das crianças na formação escolar.

Em 2016, foram analisados 21 bons leitores (12 meninas e 9 meninos), 21 maus leitores (14 meninas e 7 meninos) e 26 disléxicos (7 meninas e 19 meninos). O classificador foi mais eficaz na distinção de bons leitores e disléxicos tanto na análise das medidas das transcrições originais (Figura 56) quanto das corrigidas (Figura 57). A acurácia para as corrigidas, contudo, foi mais alta, como já havia sido comentado: 23,8% (N=5) dos bons leitores foram classificados como disléxicos; 30,7% (N=8) dos disléxicos foram classificados como bons leitores. Ambas são taxas de erro bem altas se considerado o propósito deste estudo de verificar o potencial preditivo dos atributos de grafo para a dislexia do desenvolvimento.

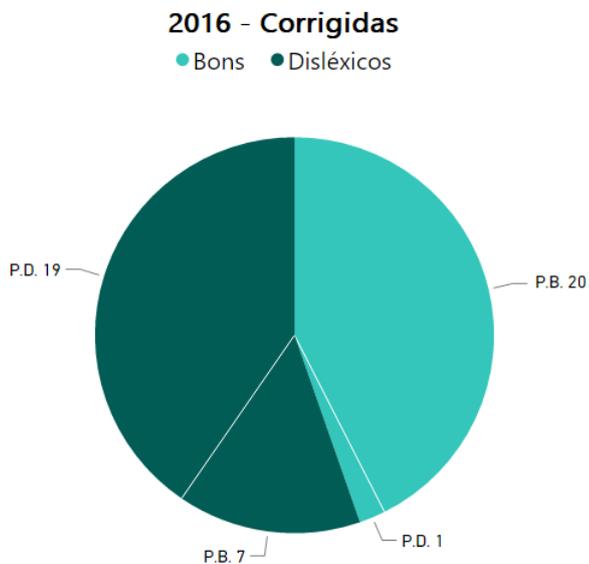
Figura 56: Gráfico de setores de 2016 - Originais



Fonte: A autora

A Figura 57 - que contempla o mesmo contraste de fluências leitoras, porém agora com os resultados obtidos com as análises das transcrições corrigidas de 2016 – retrata taxas de erro menores: 5% (N=1) dos bons leitores foram classificados como disléxicos e 26,9% (N=8) dos disléxicos foram classificados como bons leitores.

Figura 57: Gráfico de predição de 2016 - Corrigidas



Fonte: A autora

A tabela com as matrizes de confusão incluem estes e os demais resultados:

Tabela 20: Matrizes de confusão de 2016

BONS X DISLÉXICOS					
Originais Acurácia: 72,34%			Corrigidas Acurácia: 82,98%		
Pred./Categ.	Bons	Disléxicos	Pred./Categ.	Bons	Disléxicos
B	16	8	B	20	7
D	5	18	D	1	19

MAUS X DISLÉXICOS					
Originais Acurácia: 63,83%			Corrigidas Acurácia: 70,21%		
Pred./Categ.	Maus	Disléx.	Pred./Categ.	Maus	Disléxicos
M	13	9	M	15	8
D	8	17	D	6	18

BONS X MAUS					
-------------	--	--	--	--	--

Originals Acurácia: 59,52%			Corrigidas Acurácia: 26,19%		
Pred./Categ.	Bons	Maus	Pred./Categ.	Bons	Maus
B	16	12	B	9	19
M	5	9	M	12	2

Finalmente, em 2016, obteve-se índices de acurácia superiores ao ponto de corte estabelecido neste trabalho como indicativo de predição de fluência leitora. Infelizmente, estas acurácias se concentraram apenas nas distinções entre bons leitores e disléxicos, a partir das transcrições originais e corrigidas, com índices de 72,34% e 82,98% respectivamente. No entanto, aquela homogeneidade percebida nos dois anos anteriores entre bons e maus leitores, de acordo com a tabela, foi ligeiramente atenuada nas análises dos atributos extraídos das transcrições originais, sem deixar de surpreender. Desta vez, o classificador encontrou mais bons leitores entre os maus leitores do que o suposto. Dos 21 maus leitores, 57,14% (N=12) do grupo foi categorizado como bom leitor, resultado este que continuou a refletir inconsistências por parte do classificador ou dos critérios de definição de fluência leitora.

CONSIDERAÇÕES FINAIS

O presente trabalho se inseriu em um nicho investigativo até então pouco explorado por estudos em Psicolinguística. Abordar a coesão textual de participantes com diferentes níveis de fluência de leitura a partir da Teoria dos Grafos consistiu em uma escolha pouco ortodoxa, mas produtora e pertinente.

Em resumo, a dislexia é um transtorno de aprendizagem de natureza neurobiológica, caracterizado por uma acentuada e inesperada dificuldade em decodificar e compreender textos escritos (DEHAENE, 2012). Tal dificuldade advém de um aprendizado insatisfatório do princípio alfabético e de uma associação grafo-fonêmica deficitária, com significativo comprometimento do processamento fonológico (SHAYWITZ, 2008). A exploração científica acerca da expressão escrita de disléxicos tem se desdobrado, essencialmente, em investigações a respeito de aspectos linguísticos associados à ortografia e às classes de palavras empregadas por sujeitos com dislexia (TOPS et al., 2013; BASSO et al., 2017). Estudos de natureza quantitativa focados em mensurar a coesão dos textos deste grupo clínico eram, até então, desconhecidos. Inspirado no êxito de trabalhos anteriores que se utilizaram da Teoria dos Grafos para compreender a estrutura discursiva oral de psicóticos (MOTA et al., 2014), idosos com demência (BERTOLA et al., 2014) e escolares do Ensino Fundamental (MOTA et al., 2016), o presente trabalho inseriu-se neste nicho investigativo pouco explorado e obteve bons resultados.

Especificamente, este trabalho foi orientado pela expectativa de compreender:

- g) Se o tipo de transcrição textual impactaria as medidas de grafo extraídas dos textos, uma vez que o software utilizado para a extração dos atributos é sensível à quebra de linha e a correções na ortografia, componente textual expressivo de limitações impostas pela dislexia (BASSO et al., 2017);
- h) Se a normalização das medidas de grafo pelo número de palavras de cada texto seria igualmente distintiva de categorias de fluência de leitura e anos de escolaridade quanto as transcrições não normalizadas, uma vez que esta medida eliminaria o número de palavras de um texto, que consiste em um aspecto textual relevante em análises discursivas (MOTA et al., 2014; BERTOLA et al., 2014);
- i) Se leitores fluentes, leitores pouco fluentes e crianças diagnosticadas com dislexia apresentariam padrões específicos de conectividade textual, haja vista as correspondências, previstas na literatura, entre a compreensão e a produção escrita

(CAPOVILLA; CAPOVILLA, 2007; NAVAS; PINTO; DELLISA, 2009; AFFONSO et al., 2011; BASSO et al., 2017);

- j) Se uma análise longitudinal evidenciaria diferenças significativas entre as produções textuais dos participantes das escolas que produziram textos ao longo dos três anos do ACERTA, de modo a registrar mudanças no desempenho acadêmico decorrentes do avanço escolar (CAPOVILLA; CAPOVILLA, 2004; CAPOVILLA; CAPOVILLA, 2007; AFFONSO et al., 2011; GÂNDARA; BEFI-LOPES, 2010);
- k) Se as médias dos atributos de grafo dos textos produzidos em 2014, pelo grupo Escolas, apresentariam proximidade com as dos textos escritos por pacientes do Ambulatório, uma vez que a produção escrita de disléxicos tende a ser substancialmente inferior à esperada para suas idades e escolaridades, e o grupo de disléxicos é bem heterogêneo quanto à idade e escolaridade (DEHAENE, 2012; (SHAYWITZ, 2008);
- l) Se o potencial preditivo da ferramenta seria, assim, corroborado com o uso de técnicas de aprendizado de máquina, dentro deste contexto de pesquisa.

A análise estatística dos dados e os resultados de predição obtidos com a utilização dos classificadores SVM indicaram que:

- a) Transcrições originais são mais eficientes na diferenciação de grupos experimentais a partir de análises estatísticas descritivas e inferenciais, supostamente por se utilizarem de textos com diferentes números de palavras. Isto não significa, contudo, que sejam informativas aos classificadores, uma vez que este resultado não se repetiu nas predições realizadas pelos classificadores, que tiveram maior excelência classificando os grupos a partir de atributos extraídos de transcrições corrigidas dos textos;
- b) As análises de dados normalizados pelo número de palavras de cada texto não foram igualmente significativas de categorias de fluência de leitura quanto foram as não normalizadas. Supõe-se que a razão deste resultado resida na indesejada eliminação do número de palavras de cada texto, que demonstra ser uma característica relevante dos textos em análises discursivas (BERTOLA et al., 2014);
- c) Bons leitores, maus leitores e disléxicos apresentaram padrões específicos de conectividade textual com o avanço dos anos de coleta, que, presume-se, colaboraram para que diferentes estágios da consolidação de habilidades de leitura e escrita se manifestassem nas produções textuais e contrastassem, assim,

progressivamente mais com as limitações de disléxicos associadas a estas habilidades acadêmicas (CAPOVILLA; CAPOVILLA, 2007; NAVAS; PINTO; DELLISA, 2009; AFFONSO et al., 2011; BASSO et al., 2017);

- d) A análise longitudinal registrou avanço escolar e suas respectivas mudanças significativas nas habilidades de leitura e escrita somente entre os anos de 2014 e 2016, extremos do período analisado e, por isso, distintos entre si para algumas medidas quantitativas de coesão textual, em especial N, E, LCC e Dens. (CAPOVILLA; CAPOVILLA, 2004; CAPOVILLA; CAPOVILLA, 2007; AFFONSO et al., 2011; GÂNDARA; BEFI-LOPES, 2010);
- e) A análise transversal geral revelou haver significativa homogeneidade entre os textos escritos em 2014, pelos alunos das escolas envolvidas no ACERTA, e aqueles produzidos pelos pacientes com dislexia, do Ambulatório. Confirmou-se, assim, a expectativa de que a produção escrita de disléxicos tende a ser substancialmente inferior à esperada para suas idades e escolaridades – considerando que o grupo de disléxicos é bem heterogêneo quanto à idade e escolaridade, com sujeitos mais velhos e mais adiantados quanto à escolaridade que os do grupo Escolas (DEHAENE, 2012; (SHAYWITZ, 2008);
- f) O potencial preditivo da análise de grafos associada ao Aprendizado de Máquina foi constatado parcialmente. A acurácia dos classificadores se manteve bem aquém da expectativa, tendo relativo êxito somente entre as análises das transcrições originais e corrigidas de 2016, na diferenciação de bons leitores deste ano e disléxicos.

Cabe ressaltar que este estudo apresenta limitações que comprometem a generalização de seus resultados. Destacam-se o número reduzido de participantes recrutados (ou de textos analisados) e o desequilíbrio entre os tamanhos das amostras de cada grupo/subgrupo, que limitou a generalização dos resultados para toda a população de indivíduos acometidos pela dislexia. No entanto, os resultados encontrados indicam que o aprofundamento deste estudo com mais participantes e grupos experimentais mais equilibrados pode ser elucidativo de como se desdobram em amostras maiores e mais equânimes os indícios acerca do potencial de predição de fluência de leitura da análise de grafos.

São reconhecidas também limitações no recrutamento dos alunos das escolas, uma vez que questões de ordem cognitiva, socioeconômica e ambiental - tais como QI, idade por ano de coleta, renda familiar e hábitos de leitura próprios e familiares – não foram controlados.

Consideram-se todos possíveis fatores não determinantes, mas mediadores dos desempenhos de leitura e escrita dos participantes - por isso, relevantes.

Os modelos de classificadores SVM utilizados não demonstraram adequação e excelência na classificação dos textos para grupos de fluência leitora. Logo, os resultados obtidos na predição de fluência leitora não podem ser assumidos categoricamente, pois talvez reflitam limitações do modelo e não, necessariamente, da análise quantitativa diferencial de coesão textual.

Análises de significância de correlação linear não foram realizadas neste trabalho. Supõe-se que os resultados obtidos com este tipo de análise, em estudos futuros, complementariam a discussão empreendida e aprofundariam as constatações acerca dos padrões de conectividade textual dos grupos experimentais.

Quanto a iniciativas de estudo futuras que possam se servir das evidências aqui encontradas, destaca-se o aprimoramento de um protocolo de predição de fluência de leitura, baseado na análise de grafos, que poderia favorecer o diagnóstico dos pacientes do ACERTA. Para tanto, faz-se necessária uma melhor compreensão da ferramenta de análise de grafos, que pode vir a ser obtida com estudos futuros cujas amostras populacionais sejam mais robustas e equânimes. Além disso, como discutido anteriormente, as análises baseadas em transcrições sem normalizações foram mais eficientes para a diferenciação dos textos. As razões destes resultados não foram largamente discutidas, mas supõe-se que seja o reflexo dos diferentes números de palavras dos textos – que, automaticamente, alteram atributos como N, E, LCC, LSC, PE RE. Este parece ser um tópico merecedor de atenção, capaz de aprofundar a compreensão do potencial preditivo de fluência de leitura dos atributos de grafo.

Aliar análises de grafo a experimentos com RMf se mostra igualmente pertinente. Uma proposta de estudo futuro seria analisar, paralelamente, os dados comportamentais de produção textual utilizados neste estudo e os dados de RMf para tarefas de leitura de participantes das escolas e de pacientes do Ambulatório diagnosticados com dislexia. Seria uma oportunidade de comparar padrões de conectividade textual com padrões de atividade cerebral, a fim de encontrar correspondências entre manifestações distintas da dislexia do desenvolvimento.

Entende-se que a presente dissertação, como um subprojeto de pesquisa inserido dentro do ACERTA, demonstrou ter atendido ao propósito do Projeto de explorar ferramentas validadas cientificamente, capazes de beneficiar o diagnóstico da dislexia do desenvolvimento.

Mais do que isso, esta exploração do potencial preditivo da análise de grafos sobre a fluência de leitura serviu de base para investigações futuras interessadas em melhor compreender o reflexo das habilidades de leitura em medidas quantitativas de textos escritos.

Se consideradas as medidas de grafo com maior prevalência de significância em todas as análises realizadas, cujas produções médias por subgrupos experimentais – bons leitores, leitores medianos, maus leitores e disléxicos – acompanharam direta ou inversamente a progressão de suas fluências de leitura, é formado o seguinte padrão de conectividade associado à fluência leitora:

Figura 58: Padrão de conectividade associado à fluência leitora

↑ Fluência de leitura
 ↑WC* ↑N ↑E ↑LCC ↑ASP ↓Dens

Fonte: A autora

Número de palavras, número de itens lexicais, quantidade de relações entre os componentes de um texto, o maior agrupamento de componentes de um texto e a média de menores distâncias entre pares de itens lexicais foram medidas cujas variações acompanharam, em mesmo sentido, o aumento da fluência de leitura nas análises estatísticas realizadas. Contrariamente, mas igualmente constante, a densidade acompanhou inversamente a progressão da fluência leitora – quanto maior foi a fluência de leitura de uma criança, menor foi a densidade de seu texto.

Em resumo, as evidências encontradas reúnem indícios quantitativos e significativos de que a fluência de leitura dos participantes deste estudo se reflete na composição e na coesão de seus textos. Ambas parecem depender de três esferas de conectividade essencialmente: uma geral – que contempla a quantificação de itens lexicais e suas relações (N e E) -, uma associativa – que diz respeito às relações entre os componentes de um texto (ATD, LCC e LSC) – e uma global, que mensura o texto (ou o grafo) como um todo, enquanto unidade (ASP, Dens. e Diam). Dito de outra forma, a coesão de um texto é uma característica complexa, que se reflete no tecido textual quando ele apresenta mais ou menos pontos e fios – isto é, mais ou menos itens lexicais que se relacionam em grande ou em pequena escala -, maiores ou menores dimensões – obtidas pelo diâmetro e pelas medidas médias de extensão do tecido textual – e, sobretudo, pelo seu grau de maleabilidade – este resultante de quão densa ou espessa é a tessitura do texto.

Finalmente, entende-se que um texto escrito, enquanto tecido, pode ser mais rígido ou mais maleável, mais amplo ou mais aglutinado, mais robusto ou mais modesto. Tais

características parecem não se dar ao acaso. A análise de dados e a literatura revisada sugerem que há fatores de natureza cognitiva e socioambiental que tecem os textos, determinando os arranjos destes fios – distribuições estas passíveis de serem mensuradas quantitativamente e, sabe-se agora, potenciais preditoras de fluências de leitura.

REFERÊNCIAS

AFFONSO, M. J. C. O. et al. Avaliação de escrita na dislexia do desenvolvimento: tipos de erros ortográficos em prova de nomeação de figuras por escrita. **Revista CEFAC**, v. 13, n. 4, p. 628–635, 2011.

ALEXANDERSON, G. About the cover: Euler and Königsberg's Bridges: A historical view. **Bulletin of the american mathematical society**, v. 43, n. 4, p. 567–573, 2006.

AMERICAN PSYCHIATRIC ASSOCIATION. Diagnostic and Statistical Manual of Mental Disorders, 5th Edition (DSM-5). **Diagnostic and Statistical Manual of Mental Disorders 4th edition TR.**, p. 280, 2013.

ANGELELLI, P.; MARINELLI, C. V.; ZOCCOLOTTI, P. Single or dual orthographic representations for reading and spelling? A study of Italian dyslexic-dysgraphic and normal children. **Cognitive Neuropsychology**, v. 27, n. 4, p. 305–333, 2010.

ARAUJO, A. **Aprendizagem infantil: uma abordagem da neurociência, economia e psicologia cognitiva**. 1. ed. Rio de Janeiro: Academia Brasileira de Ciências, 2011.

AZEVEDO, A. F. **CÉREBRO, LEITURA E DISLEXIA: UM ESTUDO EXPERIMENTAL SOBRE A LEITURA E AS BASES NEURAI DA DISLEXIA EM MONOLÍNGUES E APRENDIZES DE INGLÊS COMO L2, COM O USO DE RESSONÂNCIA MAGNÉTICA FUNCIONAL**. [s.l.] Pontifícia Universidade Católica do Rio Grande do Sul, 2016.

BASSO, P. et al. A produção textual escrita de adultos com dislexia de desenvolvimento. 2017.

BELLMAN, R. **An introduction to artificial intelligence: Can computers think?** [s.l.] Thomson Course Technology, 1978.

BENEDICT, H. Early lexical development: comprehension and production. **Journal of child language**, v. 6, n. 2, p. 183–200, 1979.

BERTOLA, L. et al. Graph analysis of verbal fluency test discriminate between patients with Alzheimer's disease, mild cognitive impairment and normal elderly controls. **Frontiers in aging neuroscience**, v. 6, n. July, p. 1–10, 2014.

BERWICK, R. C. et al. Evolution , brain , and the nature of language. **Trends in Cognitive Sciences**, v. 17, n. 2, p. 91–100, 2013.

BIBER, D. **Dimensions of register variation: a cross-linguistic comparison**. Cambridge: Cambridge University Press, 1995.

BICKERTON, D. **Adam’s tongue: how humans made language, how language made humans**. [s.l.] Macmillan, 2009.

BORDIN, I. A. S.; MARI, J.; CAEIRO, M. F. Validação da versão brasileira do “Child Behavior Checklist” (CBCL) (Inventário de Comportamentos da Infância e Adolescência): dados preliminares. **Revista ABP-APAL**, v. 17, n. 2, p. 55–66, 1995.

BUCHWEITZ, A. A aplicação de um modelo para a compreensão na leitura em primeira e segunda línguas. In: **Aspectos cognitivos e instrucionais da leitura**. 1. ed. Bauru: EDUSC, 2008. p. 73–98.

BUCHWEITZ, A. O desafio da dislexia. p. 1–2, fev. 2016.

CAPOVILLA, A. G. S.; CAPOVILLA, F. C. **Problemas de leitura e escrita: como identificar, prevenir e remediar numa abordagem fônica**. [s.l.] Memnon, 2007.

CAPOVILLA, F. C.; CAPOVILLA, A. G. S. Research on the role of phonology, orthography and cognitive skills upon reading, spelling and dyslexia in Brazilian Portuguese. International handbook of dyslexia: a cross-language comparison and practice guide. In: **International handbook of dyslexia: a cross-language comparison and practice guide**. [s.l.: s.n.]. p. 159–172.

CASELLA, E. B.; AMARO JR., E.; DA COSTA, J. C. As bases neurobiológicas da aprendizagem da leitura. In: **Aprendizagem infantil: uma abordagem da neurociência, economia e psicologia cognitiva**. Rio de Janeiro: Academia Brasileira de Ciências, 2011. p. 37–78.

CASTLES, A.; COLTHEART, M. Varieties of developmental dyslexia. **Cognition**, v. 47, n. 2, p. 149–180, 1993.

COCQUYT, E. ET AL. The role of the right hemisphere in the recovery of stroke-related aphasia: A systematic review. **Journal of Neurolinguistics**, v. 44, p. 68–90, 2017.

COLTHEART, M. Modeling Reading: The Dual-Route Approach. In: SNOWLING, M. J.;

- HULME, C. (Eds.). . **The science of reading: a handbook**. [s.l.] Oxford: Blackwell, 2013. p. 6–23.
- COSTA, A. C. et al. Ambulatório de Aprendizagem do Instituto do Projeto ACERTA (Avaliação de Crianças em Risco de Transtorno de Aprendizagem): métodos e resultados em dois anos. In: **Neuropsicologia do desenvolvimento: infância e adolescência**. 1. ed. Porto Alegre: [s.n.]. p. 151–158.
- COSTA, M. E. et al. Scale-invariant transition probabilities in free word association trajectories. **Frontiers in integrative neuroscience**, v. 3, n. September, p. 19, 2009.
- CURTIN, S.; MANIS, F. R.; SEIDENBERG, M. S. Parallels between the reading and spelling deficits of two subgroups of developmental dyslexia. **Reading and Writing** ·, v. 14, n. 5, p. 515–547, 2001.
- DE SOUZA, I. G. et al. Dificuldades no diagnóstico de TDAH em crianças. **J bras psiquiatr**, v. 56, n. 1, p. 14–18, 2007.
- DEDIU, D. et al. On the antiquity of language : the reinterpretation of Neandertal linguistic capacities and its consequences. **Frontiers in psychology**, v. 4, n. July, p. 397, 2013.
- DEHAENE, S. et al. How learning to read changes the cortical networks for vision and language. **Science**, v. 330, n. 6009, p. 1359–64, 2010a.
- DEHAENE, S. et al. How learning to read changes the cortical networks for vision and language. **Science (New York, N.Y.)**, v. 330, n. 6009, p. 1359–64, dez. 2010b.
- DEHAENE, S. **Os neurônios da leitura: como a ciência explica a nossa capacidade de ler**. [s.l.] Editora Penso, 2012.
- DEHAENE, S. Learning to read changes the cortical networks for vision and verbal language. v. 48, n. 1, p. 148–152, 2013.
- DEHAENE, S. ET AL. Why do children make mirror errors in reading? Neural correlates of mirror invariance in the visual word form area. **Neuroimage**, v. 49, n. 2, p. 1837–1848, 2010.
- DEUSCHLE, V. P.; CECHELLA, C. O déficit em consciência fonológica e sua relação com a dislexia: diagnóstico e intervenção. **Revista CEFAC**, v. 11, n. 2, p. 194–200, 2009.
- DIAS, R. S.; ÁVILA, C. R. F. Uso e conhecimento ortográfico no transtorno específico da leitura Orthographic knowledge and use in children with dyslexia. **Revista da Sociedade**

Brasileira de Fonoaudiologia, v. 13, n. 4, p. 381–390, 2008.

EHRI, L. C. Learning to read words: Theory, findings, and issues. **Scientific Studies of reading**, v. 9, n. 2, p. 167–188, 2005.

ELLIS, A. W. **Leitura, escrita e dislexia: Uma análise cognitiva**. Porto Alegre: Artes Médicas, 1995.

FERRARI, G. A State of the art in Computational Linguistics. In: **Linguistics Today. Facing a Greater Challenge**. Amsterdam: John Benjamins, 2004. p. 163–186.

FLETCHER, C. R.; BLOOM, C. P. Causal Reasoning in the Comprehension of Simple Narrative Texts. **Journal of Memory and language**, v. 27, n. 3, p. 235–244, 1988.

FLETCHER, J. M. et al. **Transtornos de aprendizagem: da identificação à intervenção**. [s.l.] Artmed, 2009.

FLETCHER, P. C. et al. Other minds in the brain: a functional imaging study of “theory of mind” in story comprehension. **Cognition**, v. 57, n. 2, p. 109–128, 1995.

FRITH, U. Beneath the surface of developmental dyslexia. In: PATTERSON, K.; MARSHALL, J.; COLTHEART, M. (Eds.). . **Surface dyslexia**. London: Erlbaum, 1985. v. 120–121p. 301–330.

FROST, S. J. et al. Phonological awareness predicts activation patterns for print and speech. **Annals of Dyslexia**, v. 59, n. 1, p. 78–97, 2009.

FUCHS, L. S. et al. Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. **Scientific studies of reading**, v. 5, n. 3, p. 239–256, 2001.

GÂNDARA, J. P.; BEFI-LOPES, D. M. Tendências da aquisição lexical em crianças em desenvolvimento normal e crianças com alterações específicas no desenvolvimento da linguagem. **Revista da Sociedade Brasileira de Fonoaudiologia**, v. 15, n. 2, p. 297–304, 2010.

GANSCHOW, L.; SPARKS, R. L. Learning disabilities and foreign language difficulties: Deficit in listening skills? **Journal of Reading, Writing and Learning Disabilities International**, n. 2, p. 305–319, 1986.

GARNHAM, A. **On-line construction of representations of the content of texts**. [s.l.] Indiana University Linguistics Club, 1982.

GEORGIEWA, P. ET AL. Phonological processing in dyslexic children: a study combining functional imaging and event related potentials. **Neuroscience letters**, v. 318, n. 1, p. 5–8, 2002.

GLENBERG, A. M.; MEYER, M.; LINDEM, A. Mental Models Contribute to Foregrounding Text Comprehension during. **Journal of Memory and language**, v. 26, n. 1, p. 69–83, 1987.

GOLDMAN, S. R. et al. Comprehending and Learning From Internet Sources : Processing Patterns of Better and Poorer Learners. **Reading Research Quarterly**, v. 47, n. 4, p. 356–381, 2012.

GOLDMAN, S. R.; VARMA, S. CAPPING the construction-integration model of discourse comprehension. In: **Discourse comprehension: Essays in honor of Walter Kintsch**. [s.l: s.n.]. p. 337–358.

GRAESSER, A. S.; SINGER, M.; TRABASSO, T. Minimal or global inference during reading. **Journal of Memory and Language**, v. 33, n. 4, p. 421, 1994.

GUARESI, R.; ALMEIDA, O. L. Dislexia: uma proposta de protocolo para rastreio em pré-escolares e escolares das séries iniciais. In: **Aprendizado da leitura e da escrita: a ciência em interfaces**. [s.l: s.n.]. p. 408.

HAN, J.; PEI, J.; AMBER, M. **Data mining: concepts and techniques**. Elsevier, 2011. [s.l.] Elsevier, 2011.

HANLEY, J. R.; HASTIE, K.; KAY, J. Developmental surface dyslexia and dysgraphia: An orthographic processing impairment. **The Quarterly Journal of Experimental Psychology**, v. 44, n. 2, p. 285–319, 1992.

HIGOUNET, C. **História Concisa da Escrita**. São Paulo: Parábola Editorial, 2014.

HILLIS, A. E.; RAPP, B. C. Cognitive and neural substrates of written language: comprehension and production. In: **The cognitive neurosciences**. [s.l.] MIT Press, 2004. p. 775–787.

HILLS, T., MAOUENE, J., RIORDAN, B., SMITH, L. B. **Contextual diversity and the associative structure of adult language in early word learning**. Proceedings of the 31st Annual Conference of the Cognitive Science Society. **Anais...2009**

- HOEFT, F. et al. Neural systems predicting long-term outcome in dyslexia. **Proceedings of the National Academy of Sciences of the United States of America**, v. 108, n. 1, p. 361–6, jan. 2011.
- HOVERS, E. Invention, reinvention and innovation: the makings of Oldowan lithic technology. In: **Origins of Human Innovation and Creativity**. 16. ed. [s.l: s.n.]. p. 51–68.
- I CANCHO, R. F.; SOLÉ, R. V. The small world of human language. **Proceedings of the Royal Society of London B: Biological Sciences**, v. 268, n. 1482, p. 2261–2265, 2001.
- JONES, C. M. Construction of a mental model. **Sources of coherence in reading**, p. 335–352, 1995.
- KINTSCH, W. The role of knowledge in discourse comprehension: A construction-integration model. **Psychological review**, v. 95, n. 2, p. 163, 1988.
- KINTSCH, W.; VAN DIJK, T. A. Toward a model of text comprehension and production. **Psychological Review**, v. 85, n. 5, p. 363–394, 1978.
- LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **Nature**, v. 521, n. 7553, p. 436–444, 2015.
- LENT, R. **Cem bilhões de neurônios?** 2ª ed. São Paulo: Atheneu, 2010.
- LISHMAN, W. A. Developmental dyslexia. **Journal of Neurology, Neurosurgery & Psychiatry**, v. 74, n. 12, p. 1603–1605, 2003.
- LUZ, J. P.; BUCHWEITZ, A. O método GraphoGame como catalisador da proficiência leitora. **Anais do IX Colóquio de Linguística, Literatura e Escrita Criativa [DES]LIMIARES DA LINGUAGEM**, p. 318–330, 2016.
- MANDLER, J. M.; JOHNSON, N. S. Remembrance of things parsed: Story structure and recall. **Cognitive psychology**, v. 9, n. 1, p. 111–151, 1977.
- MATTOS, P. et al. Apresentação de uma versão em português para uso no Brasil do instrumento MTA-SNAP-IV de avaliação de sintomas de transtorno do déficit de atenção/hiperatividade e sintomas de transtorno desafiador e de oposição. **Revista de Psiquiatria do Rio Grande do Sul**, v. 28, n. 3, p. 290–297, dez. 2006.
- MCCARTHY, J. **What is artificial intelligence?**
- MEYLER, A. et al. Brain activation during sentence comprehension among good and poor

- readers. **Cerebral cortex (New York, N.Y. : 1991)**, v. 17, n. 12, p. 2780–7, dez. 2007.
- MOLL, K. et al. Cognitive mechanisms underlying reading and spelling development in five European orthographies. **Learning and Instruction**, v. 29, p. 65–77, 2014.
- MONARD, M. C.; BARANAUSKAS, J. A. **Conceitos sobre Aprendizado de Máquina**. [s.l.] Manole Ltda, 2003.
- MOOJEN, S. **A escrita ortográfica na escola e na clínica: Teoria, avaliação e tratamento**. São Paulo: Casa do Psicólogo, 2009.
- MORGAN, T. J. H. et al. Experimental evidence for the co-evolution of hominin tool-making teaching and language. **Nature Communications**, v. 6, p. 1–8, 2015.
- MORKEN, F. et al. Children with dyslexia show cortical hyperactivation in response to increasing literacy processing demands. **Frontiers in Psychology**, v. 5, n. DEC, p. 1–14, 2014.
- MOTA, N. **Análise de grafos aplicada a relatos de sonhos: ferramenta diagnóstica objetiva e diferencial para psicose esquizofrênica e bipolar**. [s.l.] UFRN, 2013.
- MOTA, N. B. et al. Speech graphs provide a quantitative measure of thought disorder in psychosis. PloS one,. **PLoS ONE**, v. 7, n. 4, p. e34928, 2012.
- MOTA, N. B. et al. Graph analysis of dream reports is especially informative about psychosis. **Scientific reports**, v. 4, p. 1–7, 2014.
- MOTA, N. B. et al. A Naturalistic Assessment of the Organization of Children’s Memories Predicts Cognitive Functioning and Reading Ability. **Mind, Brain, and Education**, v. 10, n. 3, p. 184–195, 2016.
- NAVAS, A. L. G. P.; PINTO, J. C. B. R.; DELLISA, P. R. R. Avanços no conhecimento do processamento da fluência em leitura: da palavra ao texto. **Revista da Sociedade Brasileira de Fonoaudiologia**, 2009.
- NOLAN, K. A.; CARAMAZZA, A. Modality-independent impairments in word processing in a deep dyslexic patient. **Brain and Language**, v. 16, p. 232–264, 1982.
- NOLAN, K. A.; CARAMAZZA, A. An analysis of writing in a case of deep dyslexia. **Brain and Language**, v. 20, n. 2, p. 305–328, 1983.
- OECD. **PISA 2012 Results**. Disponível em: <<http://www.oecd.org/pisa/keyfindings/pisa-2012->

results.htm>. Acesso em: 10 dez. 2017.

OJANEN, E. et al. GraphoGame-A catalyst for multi-level promotion of literacy in diverse contexts. **Frontiers in Psychology**, v. 6, n. MAY, p. 1–13, 2015.

OLSON, R. K. **Genes, environment, and dyslexia the 2005 norman geschwind memorial lecture** *Annals of Dyslexia*, 2006.

ORE, O. **Graphs and their uses**. Cambridge: Cambridge University Press, 1990.

OTHERO, G. D. Á. Lingüística Computacional : uma breve introdução. **Letras de hoje**, v. 41, n. 2, p. 341–351, 2006.

OTTOSSON, T. **The World in Mind: Mental Models of Spatial Relations**. [s.l.] Gothenburg University, 1987.

PARDELLI, G.; BIAGIONI, S. Quando la linguistica incontra l'informatica: una riflessione terminologica. **SCIRES-IT-SCIENTIFIC RESEARCH AND INFORMATION TECHNOLOGY**, v. 3, n. 1, p. 67–78, 2013.

PEGADO, F. Aspectos cognitivos e bases cerebrais da alfabetização: um resumo para o professor. In: EDUFRRN (Ed.). . **Aprendizado da leitura**. Natal: [s.n.]. p. 376.

PENNINGTON, B. F. From single to multiple deficit models of developmental disorders. **Cognition**, v. 101, n. 2, p. 385–413, 2006.

PINKER, S. **The language instinct: The new science of language and mind**. [s.l.] Penguin UK, 1995.

PUNT, M. ET AL. Minor neurological dysfunction in children with dyslexia. **Developmental medicine and child neurology**, v. 52, n. 12, p. 1127–1132, 2010.

RAMSDEN, S. et al. Verbal and nonverbal intelligence changes in the teenage brain. **Nature**, v. 479, n. 7371, p. 113–116, 2011.

RAMUS, F.; PIDGEON, E.; FRITH, U. The relationship between motor control and phonology in dyslexic children. **Journal of Child Psychology and Psychiatry**, v. 44, n. 5, p. 712–722, 2003.

REIS, A; CASTRO-CALDAS, A. Illiteracy: a cause for biased cognitive development. **Journal of the International Neuropsychological Society : JINS**, v. 3, n. 5, p. 444–450, 1997.

RICHARDSON, U.; LYYTINEN, H. The GraphoGame Method: The Theoretical and Methodological Background of the Technology-Enhanced Learning Environment for Learning to Read. **Human Technology: An Interdisciplinary Journal on Humans in ICT Environments**, v. 10, n. 1, p. 39–60, 2014.

ROCHA, F. T. Conhecendo e respeitando a fisiologia cerebral da leitura. In: **Aprendizado da leitura e da escrita**. Natal: Edufrn, 2015. p. 408.

RUECKL, J. G. et al. Universal brain signature of proficient reading: Evidence from four contrasting languages. **Proceedings of the National Academy of Sciences of the United States of America**, v. 112, n. 50, p. 15510–15515, nov. 2015.

RUSSELL, S.; NORVIG, P.; INTELLIGENCE, A. A modern approach. In: **Artificial Intelligence**. Englewood Cliffs: Prentice-Hall, 1995. v. 25p. 27.

SALLES, J. F. **Habilidades e dificuldades de leitura e escrita em crianças de 2ª série: abordagem neuropsicológica cognitiva**. [s.l.] Universidade Federal do Rio Grande do Sul, 2005.

SALLES, J. F. DE; PARENTE, M. A. DE M. P.; MACHADO, S. DA S. As dislexias de desenvolvimento: aspectos neuropsicológicos e cognitivos. **Interações**, v. 9, n. 17, p. 109–132, 2004.

SARAIVA, R., MOOJEN, S., MUNARSKI, R. **Avaliação da compreensão leitora: Textos expositivos**. Porto Alegre: Casa do Psicólogo, 2007.

SARDINHA, T. B. Lingüística de corpus: histórico e problemática. **Delta**, v. 16, n. 2, p. 323–367, 2000.

SCHERER, L. C. Como os hemisférios cerebrais processam o discurso: evidências de estudos comportamentais e de neuroimagem. **Linguagem e Cognição**, p. 77, 2009.

SCLIAR-CABRAL, L. Evidências a favor da reciclagem neuronal para a alfabetização. **Letras de Hoje**, v. 45, n. 3, p. 43–47, 2010.

SHAYWITZ, S. E.; SHAYWITZ, B. A. Dyslexia (specific reading disability). **Biological psychiatry**, v. 57, n. 11, p. 1301–9, 2005.

SHAYWITZ, S. E. Dyslexia. **The New England journal of medicine**, v. 338, n. 5, p. 307–312, 1998.

- SHAYWITZ, S. E. et al. Functional disruption in the organization of the brain for reading in dyslexia. **Proceedings of the National Academy of Sciences**, v. 95, n. 5, p. 2636–2641, mar. 1998.
- SHAYWITZ, S. E. **Overcoming dyslexia: A new and complete science-based program for reading problems at any level.** [s.l.] Knopf, 2003.
- SHAYWITZ, S. E. **Vencer a dislexia.** 05/2008 ed. [s.l.] Porto Editora, 2008.
- SIGMAN, M. ET AL. Neuroscience and education: prime time to build the bridge. **Nature neuroscience**, v. 17, n. 4, p. 497–502, 2014.
- STEIN, N. L.; GLENN, C. G. STEIN, Nancy L.; GLENN, Christine G. An Analysis of Story Comprehension in Elementary School Children: A Test of a Schema. 1975.
- TOPS, W. et al. Beyond spelling: the writing skills of students with dyslexia in higher education. **Reading and Writing**, v. 26, n. 5, p. 705–720, 2013.
- TRABASSO, TOM. LANGSTON, M. Modeling causal integration and availability of information during comprehension of narrative texts. In: **The construction of mental representations during reading.** [s.l.: s.n.]. p. 25.
- TRABASSO, T.; VAN DEN BROEK, P. Causal thinking and the representation of narrative events. **Journal of Memory and Language**, v. 24, n. 5, p. 612–630, 1985.
- TRUDEAU, R. J. **Introduction to Graph Theory.** Nova Iorque: Dover Publications, 1976.
- VAN DEN BROEK, P. Causal inferences and the comprehension of narrative texts. **Psychology of learning and motivation**, v. 25, p. 175–196, 1990.
- VAN DEN BROEK, P.; YOUNG, M.; LINDERHOLM, Y. T. T. The landscape model of reading: Inferences and the online construction of a memory representation. In: **The construction of mental representations during reading.** [s.l.: s.n.]. p. 71–98.
- VAN DIJK, T. A.; KINTSCH, W. **Strategies of discourse comprehension.** Nova Iorque: Academic Press, 1983.
- VIEIRA, R.; LIMA, V. Linguística computacional: princípios e aplicações. **SBC - Jornadas de Atualização em Inteligência Artificial (JAIA)**, v. 3, p. 47–86, 2001.
- VITEVITCH, M. S. What Can Graph Theory Tell Us About. v. 51, n. April, p. 408–422, 2008.

VUKOVIĆ, M.; VUKOVIĆ, I.; MILLER, N. Acquired dyslexia in Serbian speakers with Broca's and Wernicke's aphasia. **Journal of Communication Disorders**, v. 61, p. 106–118, 2016.

WILSON, R. J. **Introduction to Graph Theory**. 4. ed. Harlow: Longman, 1996.

WIMMER, H.; MAYRINGER, H. Dysfluent reading in the absence of spelling difficulties: A specific disability in regular orthographies. **Journal of Educational Psychology**, v. 94, n. 2, p. 272–277, 2002.

ANEXOS

Anexo A – Questionário de anamnese

ANAMNESE AMBULATORIO DE APRENDIZAGEM

1. DADOS DE IDENTIFICAÇÃO

ENTREVISTADOR:	DATA:
NOME DO PACIENTE:	D. NASC.:
NOME DO RESPONSÁVEL:	IDADE (MESES):
TEL 1:	TEL 2:
É CANHOTO: (S) (N)	

2. QUEIXA PRINCIPAL

3. DESENVOLVIMENTO

PESO NASC. (GRAMAS):

PREMATURIDADE: (S) (N)

INTERNAÇÃO PERINATAL: (S) (N)

IDADE QUE COMEÇOU A CAMINHAR (MESES):

IDADE QUE COMEÇOU A FALAR (MESES):

4. SAÚDE

PROBLEMAS DE SAÚDE:

AUDIOMETRIA ALTERADA: (S) (N) (N SABE)	OTITES DE REPETIÇÃO: (S) (N) (N SABE)
AV. OFTALMOLÓGICA ALTERADA: (S) (N) (N SABE)	INTERNAÇÃO HOSPITALAR EM UTI: (S) (N) (N SABE)
USO ESPORÁDICO DE ÁLCOOL: (S) (N)	USO ESPORÁDICO DE DROGAS: (S) (N)

USO CR. DE ÁLCOOL: (S) (N)	USO CR. DE DROGAS: (S) (N)
USO CR. DE MEDICAMENTOS: (S) (N)	DIAG. NEUROLÓGICO OU PSIQ.: (S) (N)
USO DE MEDICAMENTOS NEUROPSIQ.: (S) (N)	
ESTÁ EM TRAT. NEUROLÓGICO: (S) (N)	JÁ FEZ TRAT. NEUROLÓGICO: (S) (N)
ESTÁ EM TRAT. PSIQUIÁTRICO: (S) (N)	JÁ FEZ TRAT. PSIQUIÁTRICO: (S) (N)
ESTÁ EM TRAT. PSICOLÓGICO: (S) (N)	JÁ FEZ TRAT. PSICOLÓGICO: (S) (N)
ESTÁ EM TRAT. FONOAUDIOLÓGICO: (S) (N)	HÁ QUANTO TEMPO (MESES)
JÁ FEZ TRAT. FONOAUDIOLÓGICO: (S) (N)	QUANTO TEMPO (MESES)
ESTÁ EM TRAT. PSICOPEDAGÓGICO: (S) (N)	HÁ QUANTOS TEMPO (MESES)
JÁ FEZ TRAT. PSICOPEDAGÓGICO: (S) (N)	QUANTO TEMPO (MESES)
HOUVE MELHORAS SIGNIFICATIVAS NA APRENDIZAGEM: (S) (N)	

5. ESCOLA

ESCOLA PÚBLICA: (S) (N)	ANO ESCOLAR ATUAL:
N. DE REPETÊNCIAS:	IDADE QUE ENTROU NA ESCOLA (MESES):
IDADE DE INÍCIO DAS DIFICULDADES ESCOL. (MESES):	N. DE SUSPENSÕES:
N. DE EXPULSÕES:	BULLYING: (S) (N)

6. APRENDIZAGEM

DIFICULDADE PARA LER: (S) (N)	LEITURA LENTA: (S) (N)
-------------------------------	------------------------

ENTENDE MELHOR QUANDO ALGUÉM LÊ: (S) (N)	DIFICULDADE PARA ESCREVER: (S) (N)	
TROCA OU OMITTE LETRAS: (S) (N)	DIFICULDADE PARA INTERPRETAR TEXTOS: (S) (N)	
DIFICULDADE EM MATEMÁTICA: (S) (N)	DIFICULDADE EM OUTRAS MATÉRIAS: (S) (N)	
NECESSITA AJUDA PARA ESTUDAR: (S) (N)	REFORÇO ESCOLAR: (S) (N)	
QUEIXA ESCOLAR DE DESATENÇÃO (S) (N) (N SABE)	QUEIXA ESCOLAR DE INQUIETUDE (S) (N) (N SABE)	QUEIXA ESCOLAR DE IMPULSIVIDADE (S) (N) (N SABE)

7. FAMÍLIA

H. FAMILIAR DE DIFICULDADE ESCOLAR: (S) (N) (N SABE)	H. FAMILIAR DE REPETÊNCIA: (S) (N) (N SABE)	H. FAMILIAR DE EXPULSÕES/ SUSPENSÕES: (S) (N) (N SABE)
QUEIXA FAMILIAR DE DESATENÇÃO (S) (N) (N SABE)	QUEIXA FAMILIAR DE INQUIETUDE (S) (N) (N SABE)	QUEIXA FAMILIAR DE IMPULSIVIDADE (S) (N) (N SABE)
MORA COM MÃE (S)(N)	MORA COM PAI (S)(N)	MORA COM PADRASTO (S)(N)
MORA COM MADRASTA (S)(N)	MORA COM ALGUM AVÔ (S)(N)	N. DE IRMÃOS
N. DE MORADORES NA CASA	N. DE QUARTOS	ALGUM FAMILIAR EM USO CR. DE ÁLCOOL: (S)(N)
ALGUM FAMILIAR EM USO CR. DE DROGAS: (S)(N)	ALGUM FAMILIAR COM DIAGNÓSTICO NEUROL. OU PSIQ.: (S) (N)	ALGUM FAMILIAR EM USO DE MEDICAM. NEUROPSIQ.: (S) (N)

RENDA MÉDIA MENSAL FAMILIAR:

COMO É O RELACIONAMENTO COM FAMILIARES:

Anexo B – Questionário ABIPEME

O critério Abipeme sucede a um outro preconizado pela Associação Brasileira de Anunciantes - o critério ABA - e foi desenvolvido pela Associação Brasileira de Institutos de Pesquisa de Mercado com as mesmas finalidades do anterior, ou seja, dividir a população em categorias segundo padrões ou potenciais de consumo. Como o anterior, esse critério cria uma escala ou classificação socioeconômica por intermédio da atribuição de pesos a um conjunto de itens de conforto doméstico, além do nível de escolaridade do chefe de família.

O critério resultante, conhecido por ABA/Abipeme, passou por revisão atualizada desenvolvida pela Marplan Brasil e LPM/Burke, chegando-se a um novo critério de pontuação, atualmente em vigor, que passou a denominar-se simplesmente “critério Abipeme”. A classificação socioeconômica da população é apresentada por meio de cinco classes, denominadas A, B, C, D e E correspondendo, respectivamente, a uma pontuação determinada.

Alguns dos itens de conforto no lar (aparelho de videocassete, máquina de lavar roupa, geladeira com ou sem freezer acoplado e aspirador de pó) recebem uma pontuação independentemente da quantidade possuída; outros (automóvel, TV em cores, banheiro, empregada mensalista e rádio) recebem pontuações crescentes dependendo do número de unidades possuídas. Da mesma forma, a instrução do chefe da família recebe uma pontuação segundo o grau de escolaridade.

Assim, temos os seguintes valores de pontuação:

INSTRUÇÃO	Abipeme
Analfabeto / Primário incompleto	0
Primário Completo / Ginásial Incompleto	5
Ginásial Completo / Colegial Incompleto	10
Colegial Completo / Superior Incompleto	15
Superior Completo	21

ITENS DE CONFORTO FAMILIAR - CRITÉRIO ABIPEME

Os pontos estão no corpo da tabela abaixo:

ITENS DE POSSE	Não Tem	1	2	3	4	5	Mais de 6
Automóvel	0	4	9	13	18	22	26
Televisor em cores	0	4	7	11	14	18	22
Banheiro	0	2	5	7	10	12	15

Empregada mensalista	0	5	11	16	21	26	32
Rádio (excluindo do carro)	0	2	3	5	6	8	9
Máquinas de lavar roupa	0	8	8	8	8	8	8
Videocassete	0	10	10	10	10	10	10
Aspirador de pó	0	6	6	6	6	6	6
Geladeira comum ou com freezer	0	7	7	7	7	7	7

Os limites de classificação são definidos por:

CLASSES	CRITÉRIO Abipeme
A	89 ou mais
B	59/88
C	35/58
D	20/34
E	0/19

Anexo C - Tarefa de Compreensão Leitora – Textos de 2014, 2015 e 2016 respectivamente.

O BEBÊ ELEFANTE

O BEBÊ ELEFANTE VIVE EM GRUPOS COM SUA MÃE E OUTRAS FÊMEAS E FILHOTES.

ELE NASCE COM QUASE UM METRO DE ALTURA E 100 KILO DE PESO.

COM TRÊS DIAS DE VIDA, O FILHOTE JÁ COMEÇA A CAMINHAR, AGARRANDO-SE COM SUA TROMBA NA CAUDA DA MÃE.

O ELEFANTINHO GOSTA DE BRINCAR E SE DIVERTE JOGANDO BOLA DE BARRO SECO.

AS ELEFANTAS DO BANDO PROTEGEM E VIGIAM O BEBÊ DESDE QUE ELE NASCE. SE ELE SE MACHUCA OU É AMEAÇADO POR OUTRO ANIMAL SELVAGEM, AS OUTRAS ELEFANTAS AJUDAM A MÃE A SOCORRÊ-LO.

A LONTRA

A LONTRA É UM MAMÍFERO CARNÍVORO COM HÁBITOS AQUÁTICOS. ELA VIVE NAS MARGENS DOS RIOS E LAGOS ONDE CONSTRÓI SEUS ABRIGOS. PARA SUA SOBREVIVÊNCIA E REPRODUÇÃO, NECESSITA QUE ESTES AMBIENTES SEJAM DESPOLUÍDOS E COM BASTANTE VEGETAÇÃO.

COMO UM ANIMAL CARNÍVORO, A LONTRA ALIMENTA-SE, PRINCIPALMENTE, DE PEIXES, ALÉM DE ANFÍBIOS, RÉPTEIS, AVES, PEQUENOS MAMÍFEROS, INSETOS E CRUSTÁCEOS.

SEU CORPO É ALONGADO E SUA CAUDA, COMPRIDA E FLEXÍVEL SERVE DE IMPULSO E DE LEME AO NADAR. AS PATAS CURTAS, COM CINCO DEDOS UNIDOS POR UMA MEMBRANA, FUNCIONAM COMO NADADEIRAS, PERMITINDO QUE ELA NADE DE FORMA RÁPIDA E VIGOROSA. O PELO CASTANHO, DENSO E SEDOSO DA LONTRA É MUITO VALIOSO NO MERCADO DA MODA PARA A CONFECÇÃO DE CASACOS.

A GESTAÇÃO DA LONTRA, DURA NOVE SEMANAS, AO FINAL DA QUAL NASCEM DOIS OU TRÊS FILHOTES.

A DESTRUIÇÃO DE SEU HÁBITAT PELO HOMEM, ALIADA À CAÇA PREDATÓRIA EM FUNÇÃO DA PELE, TORNOU A LONTRA UMA ESPÉCIE, AMEAÇADA DE EXTINÇÃO.

AS LHAMAS

AS LHAMAS SÃO ANIMAIS TÍPICOS DAS MONTANHAS DOS ANDES, NA AMÉRICA DO SUL.

APESAR DE SEREM MAMÍFEROS DA FAMÍLIA DOS CAMELOS, NÃO TÊM CORCOVA E SÃO BEM MAIS BAIXOS DO QUE ELES. AS LHAMAS SÃO FORTES E SE MOVIMENTAM COM AGILIDADE PELAS TRILHAS ACIDENTADAS DAS MONTANHAS. ELAS FORAM FACILMENTE DOMESTICADAS PELO HOMEM, POIS SÃO ANIMAIS DÓCEIS E CALMOS, NÃO MORDEM E NEM DÃO COICES.

ASSIM, HÁ MUITO TEMPO, AS LHAMAS SÃO UTILIZADAS PELA POPULAÇÃO DOS ANDES COMO ANIMAL DE CARGA, PODENDO CARREGAR ATÉ 60 QUILOS DE PESO E PASSAR 2 OU 3 DIAS SEM COMER, DURANTE AS VIAGENS. ELAS TAMBÉM FORNECEM LÃ, COURO E CARNE PARA O HOMEM, SENDO QUE ESTA, POR SER SABOROSA, É MUITO APRECIADA PELOS HABITANTES DA REGIÃO.

A SUA FORMA DE DEFESA É MUITO CURIOSA: QUANDO MALTRATADAS, DÃO UMA ENORME CUSPIDA EM SEUS AGRESSORES.

Anexo D - Tarefa de Compreensão Leitora – Questionário de perguntas dirigidas

Nome: _____ Escola: _____

Turma: _____ Avaliador: _____

Tempo: _____ Velocidade (9420/tempo): _____

1 -A lontra - reconto

Tipos de animal	Mamífero	Carnívoro
------------------------	----------	-----------

Pontuação: 0 - nenhuma das 2; 1 - um das 2; 2 - as duas.

Tipos de animal	Hábitos aquáticos
------------------------	-------------------

Pontuação: 0 - ausência; 1 - presença;

Habitat	Margens de rios e/ou lagos	Águas despoluídas	Com muita Vegetação
----------------	----------------------------	-------------------	---------------------

Pontuação: 0 - nenhuma das 3; 1 - uma das 3; 2 - duas ou três.

Alimentação	Pequenos animais ou Animais	Anfíbios e/ou Répteis e/ou Aves e/ou (Pequenos) Mamíferos e/ou Insetos e/ou Crustáceos	Pequenos animais terrestres e/ou aquáticos
--------------------	---------------------------------------	--	--

Pontuação: 0 - nenhuma das 3; 1 - uma das 3; 2 - duas ou três.

1 - Características	Corpo alongado ou Corpo comprido ou São compridas	Cauda comprida e/ou Cauda longa e/ou Cauda flexível e/ou A cauda serve como leme
----------------------------	---	--

Pontuação: 0 - nenhuma das 2; 1 - um das 2; 2 - as duas.

2- Características	Patas curtas e/ou Pequenas patas	Dedos unidos e/ou Membrana (pele) entre dedos	As patas funcionam como nadadeiras e/ou
---------------------------	--	---	--

Pontuação: 0 - nenhuma das 3; 1 - uma das 3; 2 - duas ou três.

Dizer que funcionam como nadadeiras é apenas consequência dos dedos serem unidos, assim, só deve ser pontuada se a informação anterior estiver presente no relato.

3 - Características	Nadam de forma rápida e/ou São rápidas ao nadar
----------------------------	---

Pontuação: 0 - nenhuma das 3; 1 - uma das 3; 2 - duas ou três.

4 - Características	Pelo Castanho	Pelo sedoso e/ou Pelo macio
----------------------------	---------------	---------------------------------------

Pontuação: 0 - nenhuma das 2; 1 - um das 2; 2 - as duas.

Gestação	9 semanas	2 a 3 filhotes
-----------------	-----------	----------------

Pontuação: 0 - nenhuma das 2; 1 - um das 2; 2 - as duas.

Utilidade para o homem	Confecção de casacos
-------------------------------	----------------------

Pontuação: 0 - ausência; 1 - presença;

Extinção	Caça pretória	Destruição do habitat
-----------------	---------------	-----------------------

Pontuação: 0 - nenhuma das 2; 1 - um das 2; 2 - as duas.

Total Geral (20): _____ Percentagem de evocação (total x 100/20): _____

2 -Questionário

Que tipo de animal é a lontra?

Tipos de animal	Mamífero	Carnívoro	De hábitos aquáticos
------------------------	----------	-----------	----------------------

Pontuação: 0 - nenhuma das 3; 1 - uma das 3; 2 - duas ou três

Onde ela vive?

Habitat	Margens de rios e/ou lagos	Águas despoluídas	Com muita Vegetação
----------------	-----------------------------------	-------------------	---------------------

Pontuação: 0 - nenhuma das 3; 1 - uma das 3; 2 - duas ou três.

De que se alimenta a lontra?

Alimentação	Pequenos animais ou Animais	Anfíbios e/ou Répteis e/ou Aves e/ou (Pequenos) Mamíferos e/ou Insetos e/ou Crustáceos	Pequenos animais terrestres e/ou aquáticos
--------------------	---------------------------------------	--	--

Pontuação: 0 - nenhuma das 3; 1 - uma das 3; 2 - duas ou três.

Quais as características da lontra?

Pontuação: 0 - nenhuma característica ou característica não presente no texto; 1- uma característica; 2- duas ou três; 3 - mais de três.

Qual é a utilidade da pele da lontra para as pessoas?

Utilidade para o homem	Confecção de casacos
-------------------------------	----------------------

Pontuação: 0 - ausência; 1 - presença;

Porque a lontra está ameaçada de extinção?

Extinção	Caça pretória	Destruição do habitat
-----------------	---------------	-----------------------

Pontuação: 0 - nenhuma das 2; 1 - um das 2; 2 - as duas.

***Você acha correto usar pele de animais para fazer roupas?**

***Como as lontras podem ser salvas do perigo de extinção?**

***não são pontuadas**

Total Geral (12) : _____	Porcentagem de evocação (total x 100/12): _____
---------------------------------	--

Nome: _____ **Escola:** _____

Turma: _____ **Avaliador:** _____

Tempo: _____ **Velocidade (3600/tempo):** _____

Anexo D – Tarefa de Leitura de Palavras e Pseudopalavras

Avaliação da leitura de palavras e pseudopalavras (Salles, 2005)

Palavra-alvo	leitura	Palavra-alvo	leitura	Palavra-alvo	leitura
1. sapo		21. casa		41. garganta	
2. bola *		22. alanare	alanare alanari	42. atmosfera *	
3. zure	Zure Zuri	23. fixo *		43. crafissoca	crafissoca crafissoca
4. alimento		24. truga	truga	44. tiro	
5. cratilo	cratilo cratilu	25. cemitério		45. fetre	Fetre fetri
6. conversa *		26. resposta *		46. conjunto	
7. noite		27. zero *		47. droga *	
8. táxi *		28. varpa	varpa	48. nefoxosa	nefoksoza nefochoza nefossoza
9. cifo	cifo cifu	29. amarela *		49. crime	
10. fantoche		30. sanverca	sanverca sanvérca	50. croquete *	
11. parágrafo *		31. dapel	dapel	51. brasa	
12. mosferata	mosferata	32. mago		52. escorpião	
13. erva *		33. crucifixo *		53. toxo	tokse - toksi toche - tochi tasse - tossi
14. vaca		34. arquistro	arquistro arquistru	54. letra *	
15. bano	bano banu	35. isca		55. resultado	
16. trabalho		36. febre *		56. azercico	azercico azercicu
17. orquestra *		37. tonchafe	tonchafe tonchafi	57. porta *	
18. arfa	arfa	38. papel		58. saxofone *	
19. elefante		39. tapi	tapi	59. indivíduo	
20. boxe *		40. exercício *		60. paresta	parésta parêsta

Impressões gerais:

ESCORE FINAL (palavras erradas: 60): _____

Palavras (40) lidas erradas: _____

Regulares (20):

* Irregulares (20):

Pseudopalavras (20) lidas erradas: _____

APÊNDICES

Apêndice A – Termo de Consentimento Informado

Termo de Assentimento e Consentimento Informado

TERMO DE ASSENTIMENTO E CONSENTIMENTO LIVRE E ESCLARECIDO REDES NEURAIAS DE LEITURA

PESQUISADOR RESPONSÁVEL: Prof. Dr. Augusto Buchweitz – Instituto do Cérebro,
PUCRS

Seu filho (a) está sendo convidado a participar de uma pesquisa sobre leitura e o funcionamento do cérebro. O objetivo da pesquisa é entender como o cérebro funciona enquanto uma pessoa está lendo. Para descobrir isso, pessoas que não tenham doenças e já saibam ler, como seu filho (a), estão sendo convidadas a participar da pesquisa. Caso o (a) Sr.(a) dê sua autorização, seu filho fará alguns testes para avaliar como está a leitura dele(a) e, depois, fará um exame de ressonância magnética do cérebro. Durante o exame, seu filho ficará deitado na máquina de ressonância, olhando um televisor no qual aparecerão algumas palavras para ele ler e algumas cores para ele dizer o nome. O único incômodo do exame é um ruído, mas ele (a) estará usando fones de ouvido para diminuir o barulho. A ressonância magnética não faz mal à saúde, porque não utiliza radiação, como o raio-x e a tomografia. Às vezes, porém, o exame pode gerar um pouco de angústia ou preocupação, pois algumas crianças sentem-se mal por ficar paradas dentro da máquina durante alguns minutos. Se isso acontecer, nenhum problema; o exame pode ser parado a qualquer momento, caso a criança não queira continuar. Benefícios: não há benefício direto para as pessoas que participarem deste estudo, mas elas estarão ajudando outras pessoas a entender melhor o funcionamento do cérebro. Sigilo e privacidade: Todas as informações da pesquisa serão guardadas pelos pesquisadores e só eles terão acesso a essas informações. O nome do seu filho (a) não será utilizado; apenas códigos, como letras e números, serão usados para identificar os dados. Quando esses dados forem usados em textos, aulas e cursos ninguém poderá identificá-los. Despesas e compensações: se houver algum gasto de dinheiro para vocês virem ao centro de pesquisa, esse gasto será pago pelos pesquisadores (até um limite de R\$40 reais por participante).

Acredito ter sido suficientemente informado a respeito das informações que li ou que foram lidas para mim, descrevendo o estudo. Ficaram claros para mim quais são os objetivos do estudo, os procedimentos a serem realizados, seus desconfortos e riscos, as garantias de proteção dos meus dados e de esclarecimentos permanentes. Ficou claro também que se eu tiver gastos com transporte para ir até o centro de pesquisa, estes gastos serão compensados pelos pesquisadores, desde que não ultrapassem R\$ 40 reais. Concordo voluntariamente na participação de meu filho e poderei retirar o meu consentimento a qualquer momento, antes ou durante o mesmo, sem penalidades ou prejuízo ou perda de qualquer benefício que eu possa ter adquirido, ou no meu atendimento neste Serviço.

Se houver perguntas sobre esse estudo, favor entrar em contato com o Prof. Dr. Augusto Buchweitz no seguinte endereço: Instituto do Cérebro, Av. Ipiranga, 6690; Partenon; Porto Alegre/RS. Fone: 3320-3485 ramal 2693 ou o Comitê de Ética em Pesquisa da PUCRS, fone 3320-3345.

Dou meu consentimento de espontânea vontade e sem reservas para participar deste estudo.

Assinatura do paciente/representante legal

Assinatura do pesquisador responsável

Data //