



A riqueza lexical em produções escritas no 1.º ciclo: um estudo de caso¹

Lexical richness in elementary school written productions: a case study

Ana Catarina Costa

Universidade de Aveiro
anacatarina.pcosta@ua.pt

Rosa Lídia Coimbra

CLLC/DLC, Universidade de Aveiro
rlcoimbra@ua.pt
<https://orcid.org/0000-0002-8578-673X>

Luísa Álvares Pereira

CIDTFF/DEP, Universidade de Aveiro
lpereira@ua.pt
<https://orcid.org/0000-0002-9742-2351>

Resumo:

A pesquisa que aqui se apresenta enquadra-se no Projeto Díade, e dá conta de resultados da comparação da riqueza lexical de produções escritas com dois anos de intervalo, realizadas pelas mesmas díades de alunos. A recolha de *corpus* foi efetuada segundo o sistema RAMOS. O *corpus* recolhido permite o estudo longitudinal da escrita criativa de histórias como um processo e da interação entre escreventes, durante a redação colaborativa. O presente estudo incide em histórias inventadas por duas díades de alunos do ensino básico e incluiu as vertentes da diversidade e da densidade lexicais, mostrando a relevância destas pesquisas na compreensão da evolução da escrita, ao longo do percurso escolar.

Palavras-chave: Riqueza lexical; escrita colaborativa; histórias; ensino básico.

Abstract:

The present research is in the scope of the Díade Project, and shows the results of a comparison between the lexical richness of texts written within a two-year period by the same dyads of students. The corpus was collected according to the RAMOS system. The corpus collected allows the longitudinal study of creative story writing as a process and of the interaction between

1 O presente estudo foi financiado por Fundos Nacionais através da FCT – Fundação para a Ciência e a Tecnologia, I.P., no âmbito dos projetos UID/CED/00194/2019 e SFRH/BSAB/142988/2018. Enquadra-se no Projeto Interwriting/DIADE, em curso no grupo ProTextos, Ensino e Aprendizagem da Escrita de Textos (<http://protextos.web.ua.pt/>).



writers during collaborative writing. The present study focuses on stories invented by two dyads of elementary school students, and includes the aspects of lexical diversity and lexical density, showing the relevance of these researches in understanding the evolution of writing along the school years.

Keywords: Lexical richness; collaborative writing; stories; basic education.

Résumé:

La recherche présentée ici s'inscrit dans le projet Díade et montre les résultats de la comparaison de la richesse lexicale de productions écrites avec deux ans de distance par les mêmes dyades d'étudiants. Le corpus a été collecté selon le système RAMOS. Le corpus collecté permet l'étude longitudinale de l'écriture de récits créatifs en tant que processus et aussi l'interaction entre les écrivains lors de l'écriture collaborative. La présente étude porte sur les histoires inventées par deux dyades d'élèves du primaire et comprend les aspects de la diversité et de la densité lexicale, montrant la pertinence de ces recherches pour comprendre l'évolution de l'écriture tout au long du cursus scolaire.

Mots-clés: Richesse lexicale; écriture collaborative; histoires; éducation primaire.

Introdução

Sendo as palavras a matéria-prima da construção textual, a análise do léxico e da sua riqueza constitui um elemento fundamental no estudo da progressão dos alunos no que respeita ao desenvolvimento das suas competências de escrita.

Na pesquisa que aqui se apresenta, e que vem na senda de outras do grupo de trabalho em que se insere (ver, por exemplo, Santos *et al.*, 2018), dar-se-á conta dos resultados da riqueza lexical, nas vertentes da densidade e da diversidade, em produções escritas nas mesmas condições e com dois anos de intervalo.

Este tipo de estudos permite aferir, de uma forma objetiva e quantificável, a evolução dos alunos em termos do léxico utilizado na construção dos seus textos, elemento fundamental na evolução da competência escrita.

Contextualização teórica

O desenvolvimento do léxico na criança é um dos parâmetros mais evidentes e observáveis na evolução das suas competências comunicativas, tanto a nível escrito como oral. De facto, já há muito que estudos mostram que, nos primeiros anos de vida, o aumento do vocabulário é notório em cada ano que passa, indo de médias de 200 a 300 palavras, aos dois anos, para mais de 1000, aos três anos, e mais de 2600 palavras, aos seis anos (Sim-Sim, 1998, pp. 129-130).

Os mesmos estudos demonstram que não se trata apenas de uma questão quantitativa. A evolução do léxico na criança também se processa de um ponto de vista qualitativo, sendo a



criança mais velha capaz de fornecer categorizações e definições de conceitos de uma forma muito mais precisa do que em anos mais precoces.

Na análise de qualquer vocabulário, uma distinção fundamental é a da diferenciação entre palavras de conteúdo e palavras de função. Na tabela infra, apresentam-se as principais características distintivas destas duas grandes classes.

Tabela 1 – Comparação entre palavras de conteúdo e de função
(quadro elaborado com base no texto de Juste e Andrade, 2006, p. 130)

Palavras de conteúdo (lexemas)	Palavras de função (gramemas)
- formam conjuntos abertos, aos quais podem ser acrescentadas novas criações	- formam conjuntos fechados (não é fácil criar-se, por exemplo, uma nova preposição)
- apresentam significado lexical, são usadas para falar sobre o mundo extralinguístico	- apresentam um significado gramatical, exercem principalmente funções sintáticas, servindo de elementos de ligação frásica com baixa carga semântica própria
- podem ser nomes, adjetivos, verbos, advérbios (derivados de adjetivos)	- podem ser determinantes, preposições, conjunções, pronomes, interjeições, advérbios
- constituem o léxico	- apresentam propriedades gramaticais que fazem a diferença entre as línguas

Partindo da distinção entre estas duas grandes classes de palavras, a análise do vocabulário da criança incide, portanto, essencialmente, na primeira delas. O estudo do léxico no primeiro ciclo reveste-se de particular importância:

[C]onsideramos imprescindível que o 1.º ciclo contemple também atividades que visem aumentar o capital lexical das crianças e consolidar a sua consciência lexical, dado que a relação dialética que se estabelece entre léxico e sintaxe tem implicações ao nível da compreensão dos textos lidos e, conseqüentemente, no desempenho curricular dos alunos. (Pires, 2015, p. 431)

Ao longo da vida, o falante vai aperfeiçoando a riqueza lexical das suas produções escritas e orais. A riqueza lexical demonstrada na produção linguística é uma característica multidimensional e inclui diversos fatores. John Read (2000, p. 200-205) identifica quatro componentes da riqueza lexical:

- variação lexical (o mesmo que diversidade lexical), dependente da extensão do vocabulário, o que permite a não repetitividade, e é medida com base na taxa de palavras diferentes no total de palavras de um texto;
- densidade lexical, definida, e tal como introduzido por Ure (1971), como a proporção de palavras de conteúdo em relação a palavras de função, espelhando a concentração da informação, e a qual usualmente é mais acentuada na escrita do que na oralidade;



- sofisticação lexical, caracterizada pela escolha apropriada de vocábulos de baixa frequência, o que inclui o uso de termos técnicos e palavras pouco comuns, o que permite exprimir sentidos de forma precisa e sofisticada;
- correção lexical, parâmetro que Read designa por número de erros, que é uma medida que passa pela identificação de erros, baseando-se numa tipologia de Arnaud (1984, p. 19): ortográficos, de escolha lexical, morfológicos ou de interferência com outras línguas.

As medidas de todos estes parâmetros dão conta do grau de riqueza lexical do falante, mas aqui só nos ocuparemos das duas primeiras. A razão da escolha destes dois parâmetros prende-se com o facto de que “são, pela sua natureza quantitativa, bastante mais adequados a um estudo baseado em *corpus*, além de já estarem estabilizados na literatura enquanto indicadores fiáveis de verificação do desenvolvimento linguístico do repertório lexical” (Martins, 2016, p. 42).

Partimos da hipótese de trabalho de que estes dois indicadores apresentarão melhores resultados em anos posteriores de escolaridade, já que o domínio do léxico é uma das dimensões trabalhadas em sala de aula.

Corpus e metodologia

Com o presente estudo pretende analisar-se a presença dos parâmetros da densidade e da diversidade lexical em histórias inventadas por quatro alunos portugueses em dois momentos distintos do seu percurso escolar: no 2.º e no 4.º anos do ensino básico. Analisou-se, para este estudo de caso, um total de oito produções escritas individuais: quatro recolhidas em 2015 e outras quatro em 2017, junto dos mesmos alunos de uma escola do concelho de Aveiro, em ambiente urbano. Em todas elas se verificaram as mesmas condições de produção. Na tabela 2, apresentam-se os dados dos referidos textos.

Tabela 2 – Identificação do corpus do presente estudo (código ProTextos EE_007)

Díade	2.º ano Data de recolha: 10/03/2015	4.º ano Data de recolha: 08/03/2017
D1	Informante D1A (sexo masculino) Título: <i>O Tiago e a bicicleta</i> Total de palavras: 96	Informante D1A Título: <i>A Catarina em Aporos</i> Total de palavras: 216
	Informante D1B (sexo masculino) Título: <i>O tubarão e o leão</i> Total de palavras: 122	Informante D1B Título: <i>Uma aventura no zoo</i> Total de palavras: 280
D2	Informante D2A (sexo masculino) Título: <i>Os dois gatinhos</i> Total de palavras: 137	Informante D2A Título: <i>O Cristal Mágico</i> Total de palavras: 274
	Informante D2B (sexo feminino) Título: <i>Computador “Frutónico”</i> Total de palavras: 191	Informante D2B Título: <i>Choo Yeng Hu na volta ao mundo</i> Total de palavras: 307



Na figura 1, mostram-se dois exemplos do *corpus* analisado, e em que são visíveis grandes diferenças ao nível da extensão do texto e da quantidade e tipo de rasuras escritas utilizadas (Coimbra, Pereira, & Cardoso, 2019).

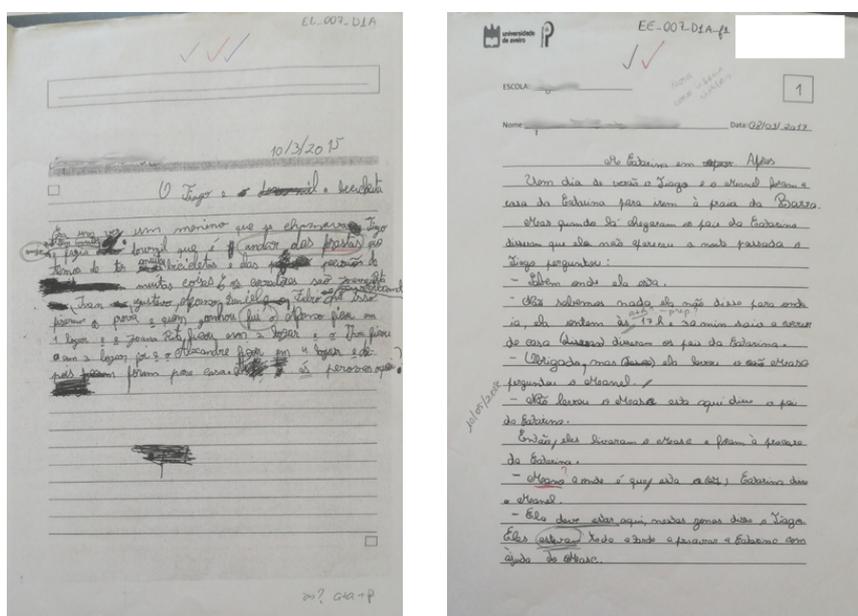


Figura 1 – Exemplos das produções escritas analisadas

Todos os textos foram redigidos em sala de aula, tendo a professora solicitado a redação de uma história inventada. Os textos seleccionados para o presente estudo foram, em cada ano, redigidos individualmente pelos alunos como produções finais de uma sequência total de oito produções, sendo a primeira individual, à qual se seguiram seis produções em díade; a oitava, e última, produção foi novamente individual e é essa que aqui se analisa. Foram seleccionados para este estudo de caso os textos da autoria de quatro alunos, pertencentes a duas das díades participantes. As recolhas do 2.º ano (primeira coluna da tabela 1) e do 4.º ano (segunda coluna da tabela 1) foram, portanto, feitas junto dos mesmos alunos, com um intervalo de dois anos e seguiram todos os procedimentos éticos, devidas autorizações e consentimentos informados. Todos os dados são tratados anonimamente.

A metodologia de recolha, neste projeto, segue o sistema RAMOS (Calil, no prelo), sendo gravado, com equipamento especial, todo o processo da construção do texto e, no caso das redações em díade, também a interação verbal entre os elementos do par.

Seguidamente, os textos foram analisados com o auxílio de uma folha de cálculo desenvolvida para o efeito por pesquisadores parceiros do projeto, no LAME, Laboratório do Manuscrito Escolar, da Universidade Federal de Alagoas (Santos *et al.*, 2018, p. 18). Nessa folha, foram transcritas todas as palavras do texto, distribuídas por colunas de acordo com as classes morfológicas a



que pertencem. As palavras repetidas são assinaladas com sombreado e contadas manualmente. Com estes totais e os totais do número de palavras em cada classe, a própria folha de cálculo apresenta as taxas de diversidade e densidade lexicais de cada texto.

A diversidade (também designada variedade lexical) depende da quantidade de palavras diferentes que o escrevente conhece e utiliza para redigir os seus textos. Se os lexemas escolhidos forem mais diversificados, verificar-se-ão menos repetições e a taxa de diversidade lexical aumenta. Esta taxa, chamada TTR (de *type token ratio*) é calculada dividindo o número total de palavras diferentes (*types*, tipos) pelo número total de palavras que ocorrem no texto (*tokens*, ocorrências). A proporção daqui resultante é quantitativamente apresentada entre 0 e 1 (ou em percentagem). Quanto mais elevado for este número, mais diverso é o vocabulário do texto analisado.

Por sua vez, a densidade lexical diz respeito à taxa de palavras de conteúdo (nomes, verbos, adjetivos, advérbios) em relação ao total de palavras utilizadas (ou seja, acrescentando, às anteriores, as palavras de função: determinantes, quantificadores, pronomes, conjunções, interjeições, verbos auxiliares). Este valor é interessante, pois um texto que contenha uma grande proporção de lexemas será mais rico em informação. Estudos indicam que as produções escritas tendem a apresentar uma riqueza lexical superior à das produções orais, já que a densidade lexical é “the kind of complexity that is typical of written language” (Halliday, 1985, p. 62). Em relação à língua inglesa, por exemplo, Ure (1971) estima que as produções orais tenham uma percentagem abaixo dos 40% e as produções escritas estejam acima deste valor. A densidade lexical de um texto é calculada dividindo o número total de lexemas pelo número total de palavras, daqui resultando uma determinada percentagem.

Estas duas medidas têm sido apontadas como fundamentais no estudo do vocabulário e os seus valores podem e têm sido utilizados para encontrar padrões característicos e efetuar estudos comparativos entre géneros textuais, falantes de diferentes sexos, idades, grau de desenvolvimento linguístico, entre outros estudos possíveis. Vejam-se, a título de exemplo, as pesquisas de Malvern *et al.* (2004), Johansson (2008), Alami *et al.* (2013), deBoer (2014), e Martins (2016).

Resultados

Aplicada a metodologia de recolha e tratamento dos dados acima exposta, obtiveram-se, para os valores de densidade e diversidade lexicais dos textos produzidos pelos mesmos quatro alunos com dois anos de intervalo, os gráficos 1 a 4, que a seguir se apresentam.

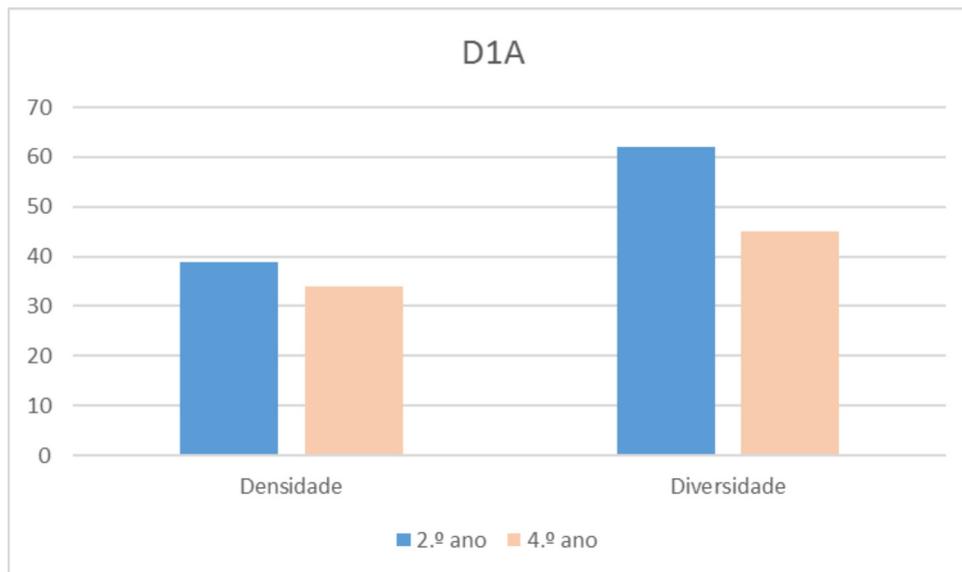


Gráfico 1 – Valores da densidade e diversidade lexicais nos textos do aluno D1A

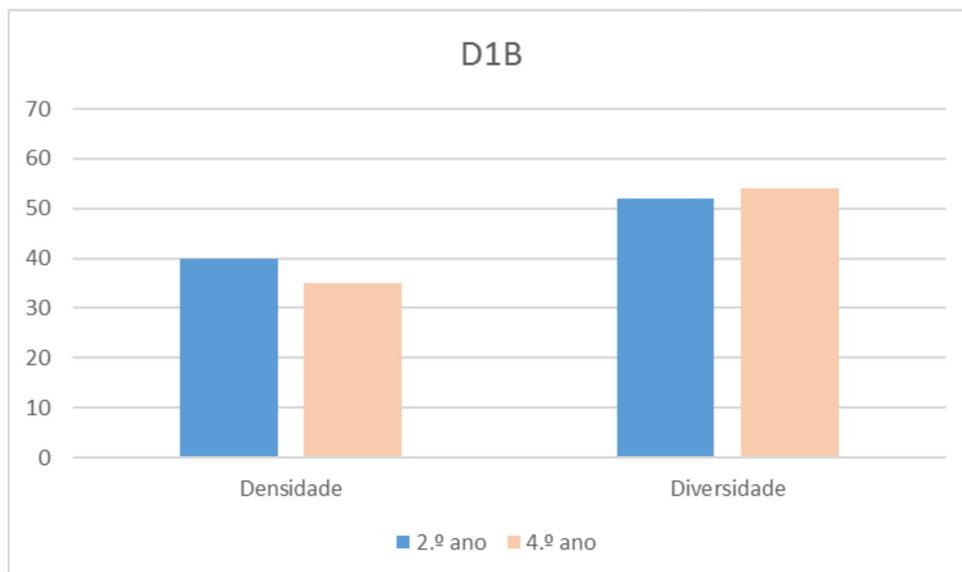


Gráfico 2 – Valores da densidade e diversidade lexicais nos textos do aluno D1B

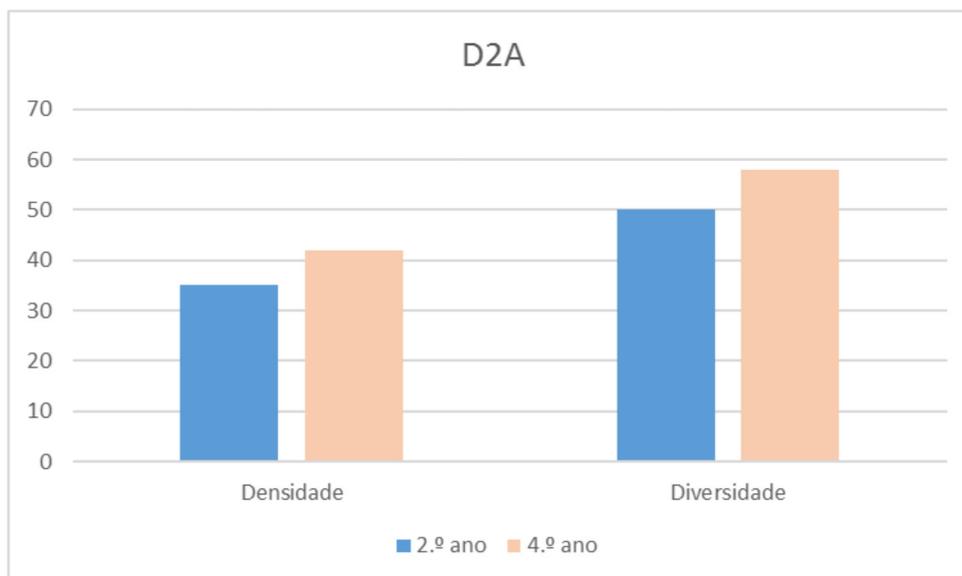


Gráfico 3 – Valores da densidade e diversidade lexicais nos textos do aluno D2A

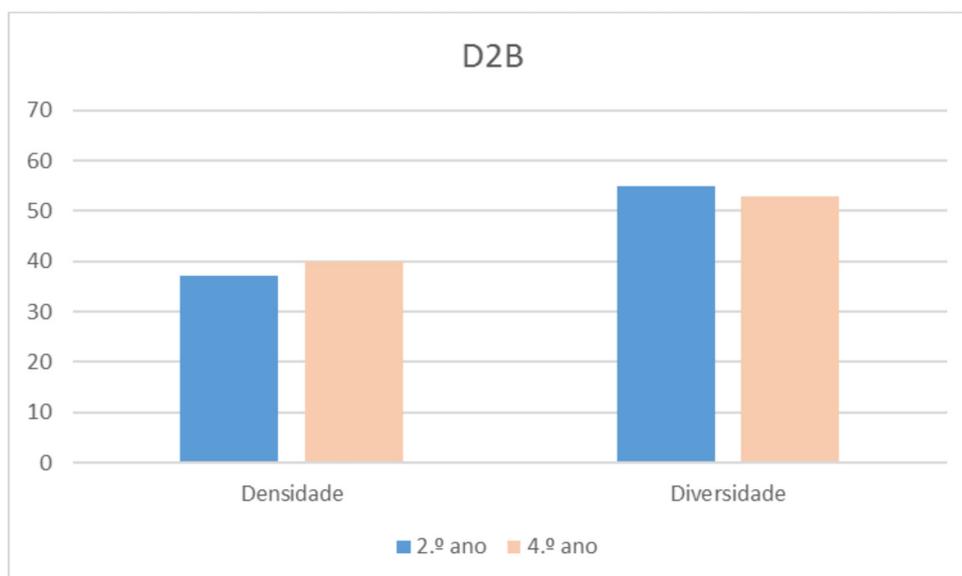


Gráfico 4 – Valores da densidade e diversidade lexicais nos textos do aluno D2B

A observação dos gráficos 1 a 4 permite-nos constatar que os informantes não apresentam taxas de diversidade e densidade lexicais muito diferentes entre o 2.º e o 4.º ano. Apenas o informante D2A viu as duas medidas aumentar. Os restantes ou apresentam pequenas subidas, ou



verificam-se mesmo pequenas descidas nos valores. Sendo estas medidas indicativas da evolução lexical do escrevente, seria, à partida, de esperar que todas as medidas fossem mais elevadas no 4.º ano em relação ao 2.º. Ora, isso só acontece em metade dos casos.

Outros estudos para o português demonstram que não é linear encontrar uma correlação entre estas medidas e a progressão escolar dos alunos. Por exemplo, Martins (2016), analisando um conjunto de 244 textos produzidos por alunos portugueses do 5.º, 7.º e 10.º anos de escolaridade, constata que “os resultados obtidos indicam que, quanto à complexidade lexical, há uma correlação positiva apenas entre a diversidade e a progressão escolar, não sendo detetada qualquer correlação entre a densidade, tomada globalmente, e a progressão escolar”.

No caso do nosso estudo, acresce que há um fator que pode ter eventualmente pesado contra o aumento destas taxas. Trata-se da extensão do texto. Se compararmos os textos produzidos no 2.º ano com os que foram elaborados no 4.º ano, verificamos que praticamente a sua extensão foi duplicada, em termos do número total de palavras, como se observa no gráfico 5. Ora, se um texto tem uma extensão maior, é expectável que, a certa altura, se comecem a repetir palavras que já foram utilizadas antes, o que fará diminuir, pelo menos, a medida da diversidade lexical.

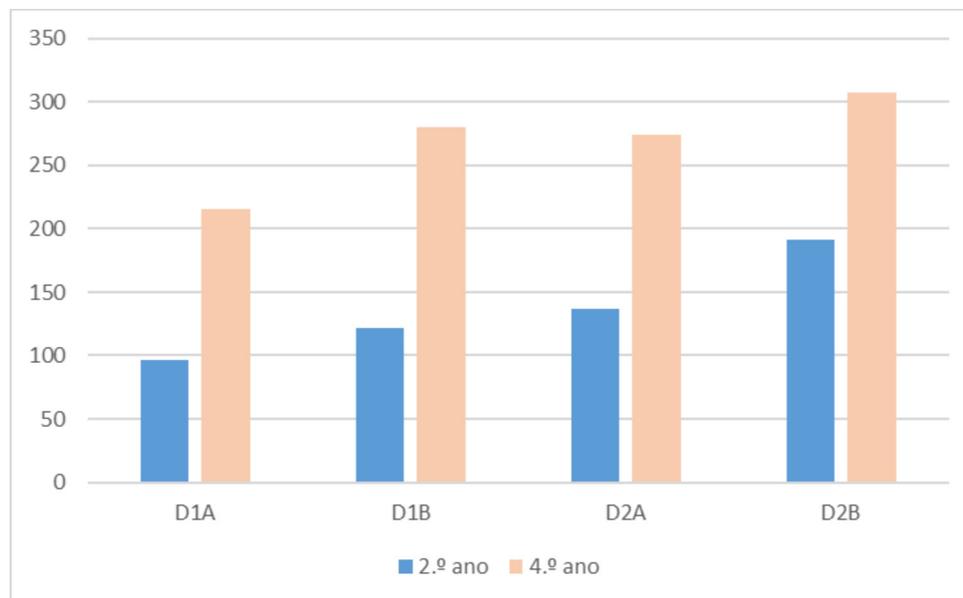


Gráfico 5 – A extensão dos textos do *corpus* em número total de palavras

Não tendo sido dados aos alunos limites quanto ao número de palavras a utilizar na construção das suas histórias, é um dado interessante verificar, pelo gráfico 5, que, no 4.º ano, todos os alunos redigiram textos muito mais extensos. Esta constatação pode indicar um maior à vontade sentido pelos informantes quanto à atividade da escrita e, como referido acima, explicar alguma repetitividade lexical responsável por não se ter verificado um aumento sistemático nos valores da riqueza lexical.



Conclusão

As medidas de diversidade e densidade lexicais, como fatores determinantes na quantificação da riqueza vocabular de um texto, foram utilizadas para comparar produções narrativas, junto dos mesmos informantes, em dois momentos da escolaridade básica. Embora, por si só, não sejam indicadoras da qualidade textual, podem dar um bom contributo para a pesquisa longitudinal sobre o desenvolvimento das competências lexicais, sendo, no entanto, fundamental o seu confronto com outros fatores, como é o caso da extensão do texto.

Estudos como o que aqui apresentamos podem trazer contributos interessantes, no que respeita à avaliação da progressão do aluno, em termos de riqueza lexical. Não existindo muitos estudos longitudinais para o português, importa alargar estas análises a *corpora* mais extensos. Esse é um objetivo que iremos continuar a ter em mente, no nosso grupo de pesquisa, tendo já sido efetuadas recolhas que permitirão a expansão destas análises a mais diádes e anos de escolaridade. A etiquetagem dos dados, processo minucioso e extremamente demorado, é o maior obstáculo à prossecução do trabalho. Daí que estudos de caso, como o que aqui apresentamos, se revelem fundamentais para apontar pistas de trabalho futuro.

Referências

- Alami, M., Sabbah, M. & Iranmanesh, M. (2013). Male-Female discourse difference in terms of lexical density. *Research Journal of Applied Sciences, Engineering and Technology*, vol. 5, n.º 23, 5365-5369.
- Arnaud, P. J. L. (1984). The lexical richness of L2 written productions and the validity of vocabulary tests. In T. Culhane, C. Klein-Braley & D. K. Stevenson (eds.). *Practice and problems in language testing* (pp. 14-28). *Occasional Papers* n.º 29. University of Essex.
- Calil, E. (no prelo). Sistema Ramos: Método para captura multimodal de processos de escritura a dois no tempo e no espaço real da sala de aula. *ALFA Revista de Linguística*, vol. 63.
- Coimbra, R. L., Pereira, L. Á. & Cardoso, I. (2019). A progressão na escrita de fábulas: estudo longitudinal no Ensino Básico. *Indagatio Didactica*, vol. 11, n.º 1, 49-68.
- deBoer, F. (2014). Evaluating the comparability of two measures of lexical diversity. *System*, n.º 47, 139-145.
- Halliday, M. A. K. (1985). *Spoken and written language*. Geelong: Deakin University Press.
- Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: a developmental perspective. *Lund University Working Papers*, n.º 53, 61-79.
- Juste, F., & Andrade, C. R. F. (2006). Tipologia das rupturas de fala e classes gramaticais em crianças gagas e fluentes. *Pró-Fono Revista de Atualização Científica*, vol. 18, n.º 2, 129-140.
- Malvern, D., Richards, B. J., Chipere, N. & Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Basingstoke: Palgrave Macmillan.
- Martins, M. G. C. (2016). *Complexidade textual e progressão escolar em dois registos: um estudo de correlação baseado em um corpus quasi-longitudinal* (tese de doutoramento). Lisboa: Universidade de Lisboa. Disponível em https://repositorio.ul.pt/bitstream/10451/23963/1/ulsd072796_td_Mario_Martins.pdf (acesso em 20-08-2019).



- Pires, N. A (2015). “Tu qué pô o péu na beça?” – Consciência lexical e aprendizagem da leitura: o papel dos prestadores de cuidados para o desenvolvimento lexical da criança. *Exedra*, vol. especial 1. *Didática do português: investigação e práticas*, 426-443.
- Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Santos, E. S., Calil, E., Pereira, L. Á. & Coimbra, R. L. (2018). Diversidade e densidade lexical em textos escritos por alunos recém-alfabetizados: um estudo descritivo de produções individuais e em díades. *Calidoscópio*, vol. 16, n.º 1, 25-32.
- Sim-Sim, I. (1998). *Desenvolvimento da Linguagem*. Lisboa: Universidade Aberta.
- Ure, J. (1971). Lexical density and register differentiation. In G. Perren, & J. L. M. Trim (eds.). *Applications of linguistics* (pp. 443-452), London: Cambridge University Press.