

Corpus de um livro didático: estudo exploratório voltado a dicionários para crianças em fase de alfabetização

Corpus of a children workbook: an exploratory study focused on dictionaries for children in the literacy phase

Submetido em: 14/08/2023

Aceito em: 09/11/2023

Janina Antonioli¹

Resumo: O uso de corpus linguístico é uma ferramenta consagrada para a compilação de dicionários. No entanto, no Brasil essa ferramenta ainda não foi utilizada para dicionários escolares de alfabetizandos. Para o presente artigo, buscou-se revisar a literatura que embasa metodologicamente o uso de corpus em dicionários pedagógicos e, com isso, avaliar quais informações um corpus compilado a partir de um livro didático poderia oferecer para a composição futura de um dicionário. Nessa perspectiva, o trabalho de Estopà (2019) contribuiu para estabelecer critérios pertinentes para o usuário mencionado, embora seus estudos sejam focados, especialmente, em dicionários terminológicos. Para este experimento, o uso de apenas um livro didático para a compilação do corpus teve como objetivo avaliar a produtividade dessa metodologia. Dentre substantivos, verbos e adjetivos, avaliou-se que tipos de lemas eram mais recorrentes e se eram frequentes no Corpus Brasileiro. Por fim, esta pesquisa exploratória é um trabalho piloto para o estabelecimento de procedimentos e critérios para a compilação de um corpus mais significativo que apoie a construção de um dicionário para alfabetizandos.

Palavras-chave: Corpus linguístico; Lexicografia Pedagógica; Alfabetização e Letramento.

Abstract: The use of corpus linguistics is an established tool for compiling dictionaries. However, in Brazil, this tool has not yet been used for school dictionaries targeted to literacy students. For the present article, an attempt was made to review the literature that methodologically supports the use of corpus in pedagogical dictionaries and, with that, to evaluate what information a corpus compiled from a children workbook could offer for the future composition of a dictionary. In this perspective, the work of Estopà (2019) contributed to establishing relevant criteria for the mentioned user, although her studies are focused, especially, on terminological dictionaries. For this experiment, the use of only one workbook to compile the corpus aimed to evaluate the productivity of this methodology. Among nouns, verbs and adjectives, it was evaluated which types of lemmas were more recurrent and if they were frequent in the Brazilian Corpus. Finally, this exploratory research is a pilot work for the establishment of procedures and criteria for the compilation of a more significant corpus that supports the construction of a dictionary for literacy students.

Keywords: Corpus linguistics; Pedagogical Lexicography; Literacy.

Introdução

A interface entre a Lexicografia e o uso de corpora é tão antiga quanto natural (Duran, 2008). A perspectiva da Linguística de Corpus, como área do conhecimento, contribui sobremaneira para o desenvolvimento da Lexicografia, pois aponta caminhos

¹ Doutoranda em Estudos da Linguagem pela Universidade Federal do Rio Grande do Sul (UFRGS). Lattes: <http://lattes.cnpq.br/2627486750658671>. OrcID: <https://orcid.org/0000-0002-7430-0001>. E-mail: nina.antonioli@protonmail.com.

e metodologias que enriquecem dicionários, especialmente dicionários com finalidades específicas, para públicos distintos (Oliveira, 2009). As ferramentas computacionais pensadas para apontar evidências linguísticas podem ser muito produtivas, também, para se pensar o que se deve apresentar à criança que está em fase de alfabetização, como um primeiro dicionário escolar.

Sendo assim, o presente artigo tem por objetivo servir de estudo inicial para bases teórico-metodológicas de um dicionário destinado a turmas em fase de alfabetização. O dicionário escolar teve e tem grande importância para o contexto escolar, seja para apresentar a obra lexicográfica e sua lógica interna à criança, seja para orientar sobre o léxico da língua (Dini, 2022; Ribeiro, De Paula, 2014; 2012). Rangel indica que o dicionário escolar:

[...] pode ser um instrumento bastante valioso para a aquisição de vocabulário e para o ensino e a aprendizagem da leitura e da escrita; e isso, para todas as áreas e para todas as horas, já que ler e escrever, dentro e fora da escola, fazem parte de muitas outras atividades. (Rangel, 2012, p. 18).

Taljad, Prinsloo (2021, p. 199) reconhecem, por um lado, que “os dicionários infantis são fundamentais para estabelecer uma cultura de dicionário e são a porta de entrada para o uso sustentado e informado do dicionário”. Por outro lado, apontam, também, que “é surpreendente que muito pouca atenção seja dada a esses dicionários na pesquisa acadêmica” (Taljad, Prinsloo; 2021, p. 199). Os esforços embasados em estudos e pesquisas são, portanto, fundamentais para qualificar a obra lexicográfica destinada aos estudantes dos primeiros anos.

De acordo com a BNCC (Brasil, 2018, p. 109), Base Nacional Curricular Comum, a criança em fase de alfabetização deveria ser capaz de reconhecer, ler e compreender (ainda que com ajuda) um verbete de dicionário. O uso autônomo do dicionário é também previsto na BNCC, mas somente no 3º ano (Brasil, 2018, p. 115). Um dicionário pensado para alfabetizandos, portanto, precisa ser acessível e adequado a esse período de ensino.

Uma vez que a Educação Brasileira tem um robusto e consagrado programa de distribuição de livro didático chamado PNLD (Programa Nacional do Livro Didático), que fomenta as salas de aulas de todas as escolas públicas do país, entendeu-se que o

livro didático poderia fornecer pistas sobre como o dicionário poderia ser conformado para atender seu público, já que a BNCC não orienta sobre qual e como seria o vocabulário esperado para o primeiro e segundo anos, mas é a base que fundamenta a concepção dos próprios livros didáticos. Este é um estudo exploratório sobre apenas um livro, que visa observar as questões linguísticas presentes e quais dados poderiam ser potencialmente explorados para uma obra lexicográfica.

Revisão bibliográfica

De acordo com Sinclair (2005, tradução nossa), um corpus é “uma coleção de trechos de textos em formato eletrônico, selecionados de acordo com critérios externos para representar, na medida do possível, uma variedade linguística como fonte de dados para pesquisa linguística”². De acordo, ainda, com Faaß (2017, p. 124, tradução nossa), “o método de compilação de tais recursos é um fator crucial, e os metadados são altamente relevantes para decidir qual corpus existente é adequado para o propósito de encontrar determinados fenômenos”³.

O trabalho do lexicógrafo, partindo das evidências linguísticas apontadas através de corpora, não é recente e tem sido automatizado desde a década de 1960 (Faaß, 2017). Ao mesmo tempo, o advento das tecnologias da informação contribui para que tanto as ferramentas, como suas análises posteriores, sejam qualificadas e possam explicar e discutir os fenômenos linguísticos de forma mais precisa. Não seria diferente com a Lexicografia e com a Linguística de Corpus (Fuertes-Olivera, 2007; Mukherjee, 2006). Mukherjee (2006) aponta que o uso de corpora para fins pedagógicos, especialmente em interface com a Lexicografia, pode contribuir para a construção de ferramentas de ensino de língua materna e adicional. Diferentes corpora podem auxiliar para tarefas de produção e de compreensão, ou seja, as experiências e necessidades de seus usuários orientam os critérios de concepção e acesso ao corpus ou aos corpora. Almeida acrescenta que

² “a collection of pieces of language text in electronic form, selected according to external criteria to represent as far as possible, a language variety as source of data for linguistic research”.

³ “the method of compilation of such resources is a crucial factor, and metadata are highly relevant when deciding which existing corpus is adequate for the purpose of finding certain phenomena”.

O trabalho com corpus tem trazido para a Linguística em geral uma experiência mais empírica por ser baseado em coletâneas de textos autênticos [...] tanto em modalidade escrita quanto oral. [...] A essa característica empírica da língua, muito nos interessa quando queremos relacionar a Linguística de/com corpus com o ensino. (Almeida, 2015, p. 650).

Há, contudo, alguns limites e insuficiências nos dicionários escolares que Estopà (2019) aponta como recorrentes em várias obras:

- Insuficiente ou nula delimitação tanto da finalidade didática que a obra busca quanto do coletivo ao qual que se dirige.
- Seleção inadequada do corpus registrado.
- Falta de precisão ou excessiva generalidade na formulação de definições.
- Ausência de um critério uniforme na ordenação de acepções e na localização de elementos fraseológicos.
- Graves insuficiências na separação de informação gramatical.
- Escassa ou nula presença de exemplos que orientem o uso e favoreçam, portanto, a função codificadora do dicionário.
- Escassa atenção aos aspectos pragmáticos e sociolinguísticos do léxico (Estopà, 2019, p. 252, tradução nossa)⁴.

Embora exista um debate sobre o suporte e o formato da obra lexicográfica, mesmo em ambiente escolar (Nomdedeu-Rull, Tarp, 2018; Tarp, Fuertes-Olivera; 2016), independentemente da forma como é apresentado, o dicionário continua sendo uma referência para o trabalho pedagógico de sala de aula (Brasil, 2018) e, por isso, a busca por qualificar a obra lexicográfica é fundamental para que não seja subutilizada. O currículo do ensino organizado para o período de alfabetização está detalhado na BNCC e esse documento programático pode fornecer informações sobre a concepção do corpus a ser compilado, como sugerem Marian (2014) e Pereira (2016). O corpus,

⁴ “Insuficiente o nula delimitación tanto de la finalidad didáctica que persigue la obra como del colectivo al que se dirige. Inadecuada selección del corpus registrado. Falta de precisión o excesiva generalidad en la formulación de las definiciones. Casos de circularidad o de pistas perdidas. Ausencia de un criterio uniforme en la ordenación de las acepciones y en la ubicación de los elementos fraseológicos. Graves insuficiencias en el apartado de la información gramatical. Escasa o nula presencia de ejemplos que orienten en el uso y favorezcan, por tanto, la función codificadora del diccionario. Escasa atención a los aspectos pragmáticos y sociolingüísticos del léxico”.

portanto, poderia contribuir para validar metodologicamente as escolhas lexicográficas, bem como orientá-las. Trabalhos como Estopà (2019), Tamayo Lozada, Ruiz Miyares (2021) sugerem caminhos de uso de corpora, ainda que suas questões metodológicas não avertam critérios que se relacionem diretamente à fase de alfabetização, que é tão específica e singular no processo de escolarização, mas também no currículo brasileiro.

Nesse sentido, embora exista produtividade reconhecida em dicionários baseados em estudos advindos da Linguística de Corpus e, além disso, constem experiências exitosas desses estudos em obras lexicográficas, é preciso encontrar, para o caso brasileiro, critérios que, além de estarem consonantes com as necessidades de uma criança em fase de alfabetização, devam estar alinhados ao currículo proposto.

Pressupostos para a elaboração da metodologia

As diferenças entre os registros da língua podem mudar de acordo com muitas variantes, sejam elas culturais, sociais e/ou etárias. A linguagem infantil, nesse sentido, apresenta características bastante típicas, assim muitas inovações nas línguas podem ser salientadas na fala e nas primeiras escritas das crianças (Wild, Kilgarriff, Tugwell, 2012). Dessa forma, a linguagem infantil costuma ser menos metafórica, utiliza frases com predomínio da ordem direta, confunde algumas marcações morfológicas menos usuais (como, por exemplo, plural e/ou gênero) (Chapman, 2000; Suppes, 1974; De Salles, Parente, 2007). Embora essas características citadas sejam mais presentes na primeira infância, a linguagem da criança vai se aperfeiçoando com o tempo e remanescências desses traços ocorrem nos primeiros anos da vida escolar (De Salles, Parente, 2007).

Assim sendo, por causa das peculiaridades da infância, é bastante natural a curiosidade sobre o quê e o quanto de palavras as crianças reconhecem – dúvidas que, *per se*, justificam projetos que analisam qualitativa e quantitativamente o conhecimento linguístico infantil (Segbers, Schroeder, 2017; Wild, Kilgarriff, Tugwell, 2012; Banerji, Gupta, Kilgarriff, Tugwell, 2013). Há, contudo, poucos corpora compilados com escrita das crianças. Estes raros trabalhos almejam oferecer subsídios para a prática docente em contextos bilíngues (Geiger-Jaillet, Ross; 2017); ou seu objetivo principal é registrar

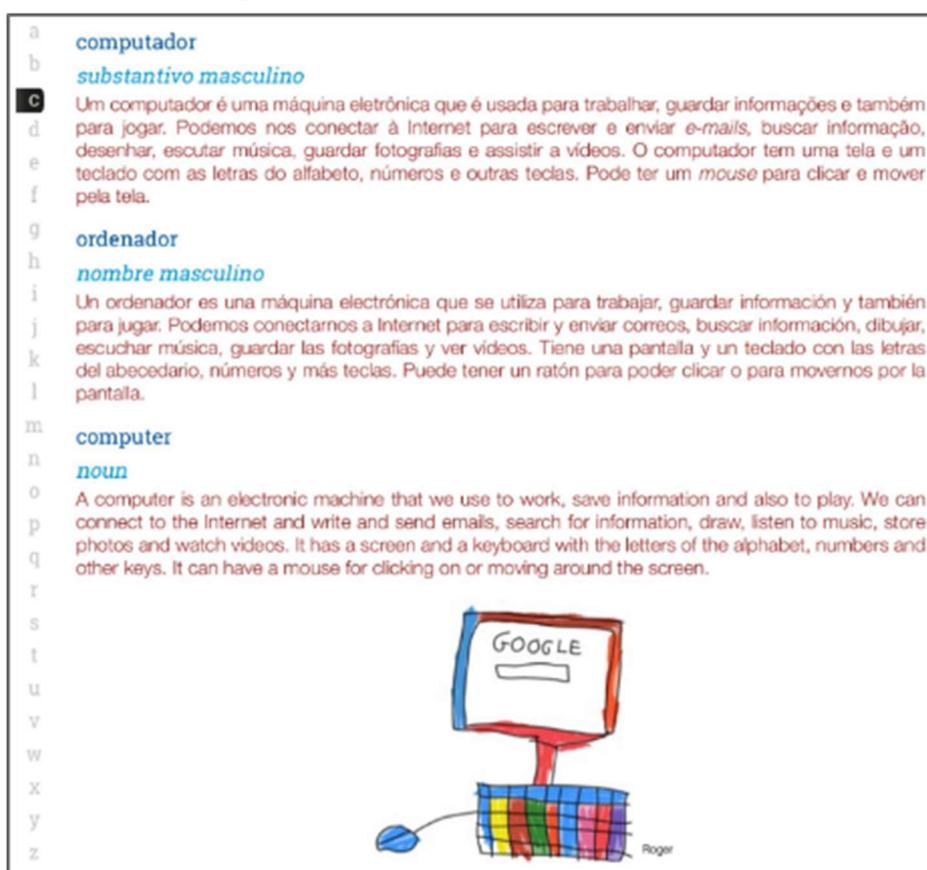
a fala (Batliner, Blomberg, D'arcy, Elenius, Giuliani, Gerosa, Hacker, Russell, Steidl, Wong; 2005; Pascual, Guevara, 2012); ou, ainda, seu foco está em crianças muito pequenas, abarcando o período da aquisição da linguagem, como o Corpus Infantil Longitudinal, que “desenvolve pesquisas relacionadas à aquisição da língua materna a partir da coleta, transcrição e marcação de dados orais e gestuais espontâneos de crianças brasileiras” (CIL, 2021).

Em Língua Portuguesa, não há projetos robustos de compilação de corpus da linguagem de crianças, seja na fase pré-escolar, seja na fase dos primeiros anos escolares, à exceção do corpus do LEAL (Laboratório de Estudos em Aquisição de Linguagem), da Universidade de São Paulo, USP. Esse corpus visa oferecer “à comunidade de linguistas, professores e pesquisadores três corpora sobre a fala, a escrita e o ambiente linguístico da criança” e um desses corpora é composto “por um conjunto de cadernos digitalizados, pertencentes a crianças do 1º ao 8º ano do Ensino Fundamental”, que se constituiu “de um conjunto digitalizado de cadernos e livros didáticos que foram utilizados durante o processo de escolarização de 5 crianças paulistas [...] cujos dados de desenvolvimento de fala foram também coletados” (CORPUS Escrita, 2021).

Mesmo que pouco se saiba sobre dicionários para escolares, abundam trabalhos que indicam a necessidade de aprofundamento e de estudos nessa seara (Sousa, 2008; Avila Martín, 2000; Lima, 2015; Brasil, 2012). Trabalhos robustos como Estopà (2019) e Estopà; Bevilacqua (2022) sobre dicionários de ciências para crianças das séries iniciais sugerem a necessidade de critérios específicos e bem delineados, que nem sempre podem ser replicados a outros públicos, desde a conformação e compilação do corpus, até os componentes canônicos do dicionário. Ou seja, a obra “para iniciantes” também é entendida como uma obra de “acesso” da criança ao universo dos dicionários. O corpus compilado por Estopà (2019), baseado nas explicações orais e desenhos que 500 crianças davam para termos das ciências naturais, é a “base da construção das definições elaboradas com uma metodologia em etapas que implica na construção – desconstrução – reconstrução” (Estopà, 2019, p. 255). A autora apresenta o passo a passo metodológico que conformou as decisões lexicográficas: a) selecionar as palavras mais adequadas; b) estabelecer cumplicidades cognitivas; c) relacionar e diferenciar informação; d) prever erros cognitivos; e) dar exemplos de boas práticas; f) proporcionar informação enciclopédica ou informação

antecipada; e g) proporcionar informação linguística necessária (Estopà, 2019, p. 256-258). O produto desse trabalho pode ser visto no exemplo a seguir:

Figura 1: Estopà, 2019, verbete *computador*.



Marconi, Ruiz Miyares, Cutugno (2018), ao realizarem o levantamento léxico dos estudantes de Cuba, compilaram um corpus de 700 mil palavras, com 25 mil palavras diferentes, advindas de textos de crianças estudantes entre o 2º e o 6º anos escolares, das zonas urbanas e rurais; são 5800 textos, 1200 amostras gravadas. Os temas dessas produções são orientados pelo currículo escolar do país, de modo que assuntos como *família*, *escola* e *entretenimento* se repetiram (Miyares Bermúdez; Suárez; Márqués; Heredia; Ruiz Miyares, 2006).

Baseado nisso, o uso do corpus LAFEC tem oferecido diferentes produtos lexicográficos no país: *Diccionario Ortográfico Español*, *Diccionario Básico Español*, *Diccionario Español Ilustrado* (DEI, para as crianças do 2º ao 4º ano, o que seria o primeiro ciclo do primário cubano) e *Vocabulario Inverso y Anagramas del Español*. O *Diccionario Básico Español* (DBE) está disponível na versão impressa, na *internet* e em

aplicativos para *smartphones* e tem o objetivo não apenas de uso pedagógico, como também da acessibilidade da obra lexicográfica para toda a cidadania cubana (Miyares Bermúdez; Zubillaga; Loinaz; Iparragirre; Ruiz Miyares; Suárez; Marqués; 2012; Tamayo Lozada; Ruiz Miyares, 2021). No caso do DEI, Miyares Bermúdez (2018) indica que

Este dicionário está dirigido ao primeiro ciclo do ensino primário e entre seus objetivos está alcançar a familiarização dos educandos com o uso desta importante ferramenta linguística. Para assimilar o anterior, o DEI possui características de acordo com o desenvolvimento psicobiológico dos escolares (Miyares Bermúdez, 2018, p. 120, tradução nossa)⁵.

Além disso, a autora lista as características principais da obra:

- As definições, geralmente descritivas, são simples e claras para propiciar a compreensão do significado das palavras.
- As definições sempre aparecem nas orações mediante as quais se pode precisar o significado e a função da palavra.
- Todas as palavras das definições são curtas, não excedem doze letras; isso se teve em conta em todos os casos para facilitar a leitura por parte das crianças dos primeiros anos. [...]
- As entradas em sua maioria foram selecionadas entre as vozes encontradas na investigação sobre o léxico, a essas se adicionaram outras extraídas dos livros didáticos dos primeiros anos do ensino primário. Isso é muito importante, pois as crianças aprenderão outros vocábulos que contribuirão em sua formação cognoscitiva e, por fim, uma melhor interpretação da realidade que os rodeia.
- O DEI é um dicionário linguístico, alfabético e contemporâneo.
- Teve-se em conta critérios psicológicos, educativos, éticos, científicos e patrióticos, que estão presentes nas definições e nas orações que servem de exemplo, as quais são acessíveis às crianças, de acordo com sua idade e grau de escolarização. [...]
- O dicionário está propenso a atingir boa ortografia graças à novidade de destacar em vermelho as letras em cada entrada que

⁵ “Este dicionario está dirigido al primer ciclo de enseñanza primaria y entre sus objetivos está lograr la familiarización de los educandos con el uso de esta importante herramienta lingüística. Para lograr lo anterior, el DEI posee características acordes con el desarrollo psico-biológico de los escolares.”

possam ser motivo de erro ortográfico. (MIYARES BERMÚDEZ, 2018, p. 120, tradução nossa)⁶.

Temos, em síntese, por um lado, um trabalho de terminologia de ciências para o uso em sala de aula que leva em conta a forma com que a criança elege o que seria um termo a ser lematizado e sua definição; e temos, por outro lado, um trabalho cujo corpus se baseia no programa de ensino nacional (de Cuba, no caso) para definir sua compilação e, posteriormente, para definir os produtos lexicográficos que dali serão realizados. Os dois trabalhos se relacionam com as demandas do ensino nacional e com os currículos previstos para a faixa etária e o público.

Nesse sentido, organizar um corpus para o caso brasileiro impõe o desafio da natureza continental do país, em primeiro lugar, e, além disso, há o limite sobre as diferenças curriculares de cada Rede de Ensino, já que a BNCC (Brasil, 2018) é bastante abrangente e porosa à diversidade do país. O único material pedagógico que está presente em todo o território nacional de maneira uniforme e com pouca variação regional é o livro didático. Esse *status* do livro didático pode ser bastante profícuo para orientar os critérios de definições da *lemmata* e o tipo de informações que os dicionários para turmas de alfabetização deveriam utilizar.

Metodologia e análise de dados

Para compreender melhor a produtividade do uso do livro didático e estabelecer critérios para a conformação do corpus, optou-se por realizar um corpus experimental com apenas um livro didático do primeiro ano, chamado “Itororó Português” (Jacintho,

⁶ “Las definiciones, generalmente descriptivas, son sencillas y claras para propiciar la comprensión del significado de las palabras. Las definiciones siempre aparecen en oraciones mediante las cuales se puede precisar el significado y la función de la palabra. [...] Todas las palabras de las definiciones son cortas, no exceden las doce letras; esto se tuvo en cuenta en todos los casos para facilitar la lectura por parte de los niños de los primeros grados. [...] Las entradas en su mayoría fueron seleccionadas entre las voces encontradas en la investigación sobre el léxico, a éstas se añadieron otras extraídas de los libros de texto del primer ciclo de enseñanza primaria. Esto es muy importante, pues los niños aprenderán otros vocablos que contribuirán en su formación cognoscitiva y, por ende, a una mejor interpretación de la realidad que los rodea. El DEI es un diccionario lingüístico, alfabético y contemporáneo. Se tuvieron en cuenta criterios psicológicos, educativos, éticos, científicos y patrióticos, los que están presentes en las definiciones y en las oraciones que sirven de ejemplo, las cuales son asequibles a los niños de acuerdo con su edad y grado. [...] El diccionario propende a lograr una buena ortografía gracias a la novedad de destacar en rojo aquellas letras de cada entrada que puedan ser motivo de error ortográfico.”

2019) e utilizado em cinco escolas consultadas na Rede Municipal de Porto Alegre. As unidades de ensino do livro são temáticas (Alimentação, Nomes, Plantas, Animais, etc.) e suas propostas giram em torno de atividades de leitura, de metalinguagem (sinonímia e antonímia, conceito de sílaba, por exemplo), de exposição oral, de produção escrita e de um jogo por unidade.

O estabelecimento de critérios para a compilação dos textos do livro didático não foi transparente, ou seja, para cada decisão foi analisado o que o corpus poderia responder sobre os questionamentos de pesquisa. Inicialmente, pensou-se em compilar todo o livro, mas logo se entendeu que não seria de grande utilidade, já que as ordens dos exercícios são muito recorrentes (“leia”, “escreva”, “desenhe”, “copie”, etc.) e, de acordo com o livro, são lidas preferencialmente pelas professoras⁷. Procuraram-se, então, os textos destinados à leitura compartilhada ou autônoma (realizada pelas crianças) e chegou-se a dois tipos de textos: a) os próprios textos de leitura das unidades de ensino; e b) as listas de palavras que as crianças são encorajadas a lerem sozinhas para fixar famílias silábicas, qualificar a consciência fonológica e/ou compreender convenções ortográficas. Nesse caso, é importante salientar que o critério determinante para a conformação do corpus foi as questões norteadoras: *o que as crianças devem (potencialmente) ler neste livro? E/ou qual é o vocabulário apresentado que as crianças estarão mais expostas?* Essas respostas poderiam ser diferentes para um livro didático cuja organização de unidades de ensino fosse distinta e tivesse outras propostas. Portanto, a primeira decisão sobre a compilação do corpus foi o estabelecimento de questões norteadoras, que orientarão sobre a recolha de textos diferentes em livros didáticos diferentes, considerando a possibilidade de estudos posteriores.

A seguir, a segunda decisão foi estabelecer separação entre textos literários e não literários. A maioria dos textos submetidos para a leitura das crianças são textos literários (parlendas, canções, quadrinhas, etc.) nesse livro didático. O jogo de palavras e as rimas promovem “ampliação do repertório cultural, da consciência fonológica e em múltiplos letramentos” (Conceição, 2020, p. 339). Ao entender que esse não é um recorte funcional para a criança, especialmente após analisar o livro didático e suas propostas de tarefas, textos literários e não literários foram igualmente tratados e agregados ao corpus, mas,

⁷ Neste artigo, opta-se pelo uso da palavra “professora”, no feminino, pois o número de alfabetizadores homens é bastante reduzido. No último Censo Escolar (INEP, 2022), as professoras correspondem a 88% dos docentes das séries iniciais e no período de alfabetização elas correspondem a pouco mais de 90%.

na análise quantitativa, as repetições (de canções, por exemplo) poderiam oferecer “ruído” – o que obrigaria a uma análise qualitativa desses itens.

Por fim, outra decisão tomada se refere ao comentário de apresentação dos textos. Esses comentários foram, ao final, excluídos da compilação, porque, no caso desse livro didático, o vocabulário estava mais orientado a uma exploração oral sugerida às professoras, como se observa neste exemplo:

<p>ANTES DE LER</p> <p>VOCÊ VAI LER UMA PARLENDA.</p> <p>A) VOCÊ SABE O QUE É UMA PARLENDA?</p> <p>B) OBSERVE A ILUSTRAÇÃO QUE ACOMPANHA O TEXTO. O QUE SERÁ QUE TEM DENTRO DO COPO?</p>	<p>Orientação de exploração que deve ser apresentada pela professora.</p>
<p>LÁ EM CIMA DO PIANO TEM UM COPO DE VENENO. QUEM BEBEU MORREU, O AZAR FOI SEU!</p> <p>DA TRADIÇÃO POPULAR.</p> 	<p>Texto que a criança deve ou deveria ler.</p>

Para a automatização da compilação dos textos, recorreu-se, inicialmente, a um programa de reconhecimento de textos a partir de fotos das páginas do livro didático. Não houve sucesso, já que menos de 30% do texto foi reconhecido. Em seguida, recorreu-se à leitura dos textos em um programa de reconhecimento de voz; nesse caso o reconhecimento das palavras foi de pouco mais de 50%; o fato de haver muitas ocorrências de algumas palavras como “babaquara” ou “baboseira”, que servem, justamente, para favorecer a questão da sonoridade nos textos, impediu eficiência maior do programa. Em função disso, além da digitação de boa parte dos dados, fez-se, em seguida, correção manual e última conferência dos textos informatizados.

Dessa forma, os textos foram compilados em um arquivo de extensão .txt e fez-se o *upload* para o programa Sketch Engine (<https://app.sketchengine.eu/>). A escolha pelo

website se deu em função de ser uma plataforma de fácil utilização e que não prejudica o desempenho do dispositivo no qual está sendo utilizada, além do fato de esse corpus, que tem apenas caráter experimental, ser pequeno e provisório – o que também é favorecido pela proposta do *website*. Outro fator que contribuiu para a opção pelo Sketch Engine é a facilidade de etiquetamento das ocorrências (de acordo com as classes gramaticais) e o fato de estar disponível no mesmo sítio eletrônico o Corpus Brasileiro (GELC, 2022, composto por 3.896.392.719 de ocorrências, de acordo com o Sketch Engine), que pode ajudar a balizar o corpus do livro didático. É importante ressaltar que o Corpus Brasileiro ranqueia as mil primeiras ocorrências de cada categoria etiquetada (substantivos, adjetivos, verbos, advérbios, etc.) na versão gratuita. A opção pela versão mais leve e gratuita se deu em função de ser um estudo com finalidade exploratória.

Após a compilação do corpus, verificaram-se 2799 palavras registradas e, apesar de ser um corpus pequeno – e experimental –, buscou-se fazer a lista de ocorrências e observar quais eram as primeiras palavras lexicais ranqueadas no programa. Entende-se como palavras gramaticais as que pertencem a “classes de palavras como pronomes, conjunções e preposições, que têm a função de estabelecer relações entre segmentos do enunciado no texto” (Fornari, 2009, p. 169), já as palavras lexicais são aquelas que “contêm em si informação semântica, como os verbos, os substantivos e os adjetivos” (Fornari, 2009, p. 170). O motivo de observância predominante das palavras lexicais deu-se em função do objetivo geral do dicionário.

WORDLIST livro didático

word (120 items | 3,20K total frequency)

Word	Frequency	Word	Frequency	Word	Frequency	Word	Frequency	Word	Frequency
1 .	171	11 tem	43	21 as	21	31 seu	14	41 formiga	9
2 ,	135	12 do	37	22 se	20	32 os	14	42 dois	9
3 que	132	13 um	34	23 não	20	33 lua	14	43 como	9
4 a	111	14 I	33	24 ;	18	34 você	11	44 mas	9
5 de	84	15 para	31	25 sopa	18	35 mais	11	45 a	9
6 o	79	16 da	29	26 _	18	36 muito	10	46 ao	8
7 e	69	17 na	27	27 no	18	37 diz	10	47 quando	8
8 ?	59	18 será	26	28 ninguém	16	38 dos	10	48 quem	8
9 é	46	19 uma	23	29 em	16	39 eu	10	49 *	8
10 -	44	20 com	22	30 por	14	40 sus	9	50 bilhete	8

De acordo com a frequência, as primeiras palavras lexicais que ocorrem são os verbos “é”, “tem” e “será”, ou seja, duas formas do verbo “ser” e uma forma do verbo “ter”; verbos de uso abundante na Língua Portuguesa e de conhecimento das crianças. As ocorrências seguintes são “não”, “sopa” e “neném”. No caso de “sopa” e “neném”, a frequência alta ocorre em função da repetição em um único texto, pois um dos materiais

sobre alimentação é uma canção de Sandra Peres, do Grupo Palavra Cantada, chamada “Sopa”, em que há repetição do verso “O que que tem na sopa no neném?”. “Sopa” ocorre mais do que “neném” porque é o título da música e também é encontrada no título de outro texto – uma receita de sopa de legumes. A lista de ocorrências por *lemma*⁸, função que o programa dispõe, aparece um pouco diferente, indicando os verbos no infinitivo e reduzindo a quantidade de palavras gramaticais, como se vê na imagem:

WORDLIST livro didático

lemma (170 items | 3,200 total frequency)

Lemma	Frequency	Lemma	Frequency	Lemma	Frequency	Lemma	Frequency	Lemma	Frequency
o	206	ter	52	se	20	todo	13	formiga	9
,	171	-	44	;	18	2	12	mas	9
de	169	a	39	sopa	18	estar	12	quando	8
.	135	l	33	...	18	você	12	3	8
que	132	para	31	fazer	17	saber	11	marta	8
ser	93	seu	30	muito	17	mais	11	quem	8
em	72	ir	24	por	17	ficar	11	ogara	8
e	69	com	22	neném	16	eu	10	bilhete	8
?	59	dizer	22	ele	14	água	10	*	8
um	56	não	20	lua	14	como	9	pedrar	7

“Sopa” e “neném” continuam com ocorrência alta, mas houve depuração de dados, em que se observa um número maior de ocorrências de verbos: “ser”, “ter”, “ir”, “dizer”, “fazer”, “estar”, “saber”, “ficar”. O verbo “pedrar” é listado, mas, além de não existir, não ocorre no livro; ao observar as linhas de concordância, percebe-se que o programa extrai o verbo “pedrar” de várias ocorrências da palavra “Pedro”:

CONCORDANCE livro didático

CQL [lemma=="pedrar"] 7
2,115.44 per million tokens 0.21%

Details

	Left context	KWIC	Right context
1	doc#0 s correções necessárias. </s><s> Se quiser, use a moldura que está na p. 211 </s><s> A poção mágica Marta e	Pedro	estão brincando de preparar poções mágic
2	doc#0 inta de lua – diz Marta, apontando um parágrafo do livro. </s><s> – E casca de ovo amassada... – acrescenta	Pedro	</s><s> – Vamos colocar tinta nanquim,
3	doc#0 nta Pedro. </s><s> – Vamos colocar tinta nanquim, que é mais fácil – sugere Marta </s><s> – Está bem! – diz	Pedro	– Agora vamos repassar os ingredientes.
4	doc#0 Foi difícil deixar meu cachorro com raiva! – diz Marta, rindo. </s><s> – Cinco lágrimas de tristeza... – continua	Pedro	, lembrando de quando tiraram o brinqued
5	doc#0 tenária... – diz Marta, olhando as azeitonas que flutuam no caldeirão. </s><s> – Areia do deserto... – continua	Pedro	, lembrando de como ficou o forno quando
6	doc#0 unhas posições por toda a casa. </s><s> – Na verdade, mais da metade dos ingredientes foram trocados – diz	Pedro	</s><s> – Não importa, o que está feito, €
7	doc#0 ii experimentar primeiro? </s><s> Os dois ficam olhando a poção por um instante, enquanto cheiram a mistura.	pedro	se levanta, pega o caldeirão e diz, baixinhc

Observou-se que a classe de palavras lexicais que mais ocorre, portanto, são os verbos; em segundo lugar, os substantivos e, em terceiro lugar, os adjetivos. Extraiu-se essa informação das ocorrências dos primeiros duzentos *lemmas* mais frequentes, já

⁸ Optou-se por utilizar a nomenclatura do programa, embora exista a palavra “lema” em português, pois, no programa, ela aparece como uma função do aplicativo.

que se observou que a função *lemma* foi mais produtiva do que a função *wordlist* sem nenhum critério prévio. Em função dessas ocorrências, fez-se nova lista de frequência baseada nas classes gramaticais: verbo (*verb*), substantivo (*noun*) e adjetivo (*adjective*).

Dessa maneira, os verbos, substantivos e adjetivos mais frequentes, em comparação com o Corpus Brasileiro, são, também, abundantes na Língua Portuguesa de forma geral e essa informação pôde ser observada através da consulta ao corpus geral de Língua Portuguesa que está disponível no próprio website do Sketch Engine, no Corpus Brasileiro (GELC, 2022). Seguem as comparações de frequência dos verbos, substantivos e adjetivos.

VERBOS					
ordem de ocorrência	verbo no corpus do livro didático	número de ocorrências		ordem de ocorrência	Verbo no Corpus Brasileiro
1	ser	93		1	ser
2	ter	52		3	ter
3	ir	24		2	ir
4	dizer	22		7	dizer
5	fazer	17		5	fazer
6	estar	12		4	estar
7	saber	11		14	saber
8	ficar	11		12	ficar
9	pedrar	7		*	pedrar
10	ver	6		11	ver
11	parar	6		134	parar
12	querer	6		15	querer
13	deixar	5		19	deixar
14	escrever	5		66	escrever
15	haver	5		8	haver
16	vir	5		16	vir
17	dar	5		10	dar

18	contar	5	31	contar
19	poder	5	6	poder
20	responder	5	131	responder

*Embora a palavra “pedrar” ocorra no Corpus Brasileiro, sempre relacionada a “Pedro”, assim como no corpus do livro didático, não está nas primeiras mil ocorrências, que é o número máximo de ocorrências permitido pelo Sketch Engine na versão gratuita.

No caso dos verbos, houve coocorrência nos dois corpora e certa semelhança na abundância de uso desses verbos assinalados, conforme se pode observar na tabela. Já no caso dos substantivos, a coocorrência não foi tão próxima nos dois corpora, embora, das vinte palavras mais frequentes no corpus do livro didático, nenhuma ocorreu menos de 500 vezes no Corpus Brasileiro. Com exceção de “vitória-régia”, todas têm mais de cinco mil ocorrências, fato que comprovaria que são palavras bastante utilizadas na Língua Portuguesa. É importante, contudo, refletir o fato de que os substantivos são as palavras que mais definem a temática de um texto (Parejo, 1997) e, em função de a BNCC (Brasil, 2018) sugerir a diversidade de textos, inclusive diversidade temática, é esperado que nem todos os substantivos mais frequentes do livro didático coocorram com as primeiras mil palavras mais frequentes do Corpus Brasileiro. Para o ranqueamento do corpus do livro didático, excluíram-se os substantivos próprios.

SUBSTANTIVOS					
ordem de ocorrência	substantivo no corpus do livro didático	número de ocorrências		ordem de ocorrência	Substantivo no Corpus Brasileiro
1	sopa	18		**	sopa
2	neném	16		**	neném
3	lua	14		**	lua
4	água	10		80	água
5	formiga	9		**	formiga
6	cigarra	8		**	cigarra
7	bilhete	8		**	bilhete
8	criança	6		63	criança
9	flor	6		**	flor

10	estrela	6	867	estrela
11	índio	6	**	índio
12	gente	6	119	gente
13	urso	6	**	urso
14	homem	6	65	homem
15	vitória-régia	6	**	vitória-régia
16	tribo	5	**	tribo
17	tempo	5	6	tempo
18	árvore	5	890	árvore
19	amigo	5	106	amigo
20	noite	5	155	noite

** Não consta entre as mil primeiras ocorrências de substantivos.

Para os adjetivos, realizou-se o mesmo procedimento. Ranquearam-se, portanto, os vinte adjetivos mais frequentes e também foi feita a comparação com o Corpus Brasileiro. Ocorrências que poderiam ser substantivos ou adjetivos, como “branco” e “mágico”, por exemplo, foram consideradas adjetivos após análise qualitativa, na ferramenta *concordance*, ou seja, no contexto. Para situações semelhantes, no caso dos substantivos, também houve a mesma conferência. Em comparação aos substantivos e verbos, as ocorrências de adjetivos diminuíram muito, ou seja, o uso de adjetivos neste corpus do livro didático *Itororó* é bastante reduzido.

ADJETIVOS					
ordem de ocorrência	adjetivo no corpus do livro didático	número de ocorrências		ordem de ocorrência	Adjetivo no Corpus Brasileiro
1	bom	5		4	bom
2	cheio	4		150	cheio
3	velho	4		84	velho
4	necessário	3		23	necessário
5	profundo	3		190	profundo
6	belo	3		76	belo

7	importante	3
8	vazio	2
9	jovem	2
10	limpo	2
11	forte	2
12	branco	2
13	pobre	2
14	leve	2
15	grande	2
16	aquático	2
17	mágico	2
18	gentil	2
19	inteligente	2
20	magro	2

14	importante
509	vazio
134	jovem
271	limpo
43	forte
156	branco
189	pobre
233	leve
2	grande
792	aquático
467	mágico
***	gentil
329	inteligente
858	magro

*** Não consta entre as mil primeiras ocorrências de adjetivos.

Ainda que os vinte adjetivos mais frequentes do corpus do livro didático coocorram distribuídos entre as primeiras mil ocorrências do Corpus Brasileiro, pode-se dizer que são palavras muito frequentes da Língua Portuguesa, pois o adjetivo que está na milésima posição de frequência tem mais de 39 mil ocorrências no Corpus Brasileiro, o que é um número considerável.

Considerações Finais

O principal intento deste trabalho foi desenvolver critérios para a compilação de um corpus que pudesse servir como orientação para um dicionário destinado a turmas em fase de alfabetização. Esse corpus compilado de um livro didático para o primeiro ano do Ensino Fundamental mostrou-se bastante produtivo, pois o processo de estabelecer quais textos e de que forma esses textos seriam aproveitados no corpus formou parte integrante do experimento. Outro ganho com esse processo foi a necessidade de manuseio do próprio livro didático, das suas propostas de atividades e

de estabelecer diferenciação de textos organizados para a leitura da criança e de textos organizados para a leitura da professora.

Além disso, o uso de um corpus de referência de língua geral – o Corpus Brasileiro – foi uma escolha útil para ranquear e para analisar o que é próprio do vocabulário de sala de aula e dos textos oferecidos para as crianças em fase de alfabetização.

A opção por tratar das palavras lexicais e não das gramaticais se mostrou consonante com o vocabulário no qual a criança está sendo exposta no livro didático. A análise de ocorrências de verbos, substantivos e adjetivos foi bastante produtiva para aventar as possibilidades futuras de pesquisa, uma vez que demonstrou que: a) os verbos apresentados no livro didático são muito recorrentes na Língua Portuguesa, de acordo com o corpus de referência; b) os substantivos que mais ocorrem no corpus do livro didático, por serem as palavras mais abundantes na língua e por construírem os significados textuais, embora muito frequentes em coocorrências, não coocorrem na lista dos mil substantivos mais frequentes do Corpus Brasileiro; e c) quase todos os adjetivos mais frequentes no livro didático constam entre os mil adjetivos de maior ocorrência no Corpus Brasileiro, ainda que os números de ocorrências dos adjetivos no corpus do livro didático sejam baixos.

Por fim, reitera-se que este foi um estudo exploratório, com o objetivo de conjecturar os próximos passos para a compilação do corpus para um dicionário destinado a alfabetizandos. Se, por um lado, o número de palavras que compõe esse corpus é bastante reduzido, por outro lado contribuiu para o estabelecimento de procedimentos e critérios para a compilação de um corpus mais significativo. Os limites deste trabalho residem, justamente, nessa questão, já que se privilegiou a apropriação metodológica à obtenção de resultados definitivos.

REFERÊNCIAS

ALMEIDA, Lucas Rezende. A funcionalidade da linguística de *corpus* no ensino de língua estrangeira: oportunidades e desafios. In: BALLBAKI, Angela; CARDOSO, Janaína; ARANTES, Poliana; BERNARDO, Sandra (Org.). *Linguagem: Teoria, Análise e Aplicações* (8). Rio de Janeiro: UERJ - Programa de Pós-graduação em Letras, 2015. p. 649-665.

ÁVILA MARTÍN, M^a del C. *El diccionario en el aula*. Granada: Edita: Editorial Universidad de Granada, 2000.

- BANERJI, N., GUPTA, V., KILGARRIFF, A., & TUGWELL, D. *The Oxford Children's Corpus: a corpus of children's writing, reading and education*. 2013. Disponível em: https://www.sketchengine.eu/wp-content/uploads/Beebox_2013.pdf. Acesso em: 13 mai. 2022.
- BATLINER, A.; BLOMBERG, M.; D'ARCY, S.; ELENIUS, D.; GIULIANI, D.; GEROSA, M.; HACKER, C.; RUSSELL, M.; WONG, M. The PF STAR Children's Speech Corpus, *Proc. Interspeech 2005*, Lisbon, Portugal, p. 2761-2764, 2005.
- BRASIL. Ministério da Educação. *Base nacional comum curricular: educação é a base*. Brasília: MEC, 2018.
- BRASIL. MEC. Secretaria de Educação Básica. *Com direito à palavra: dicionários em sala de aula (elaboração Egon Rangel)*. Brasília: Ministério da Educação, Secretaria de Educação Básica, 2012.
- CHAPMAN, Robin S. Children's language learning: an interactionist perspective. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 41(1), 2000; 33-54.
- CIL: *Corpus Infantil Longitudinal* – Grupo de pesquisa de corpus sobre aquisição da língua materna brasileira. Disponível em: <http://corpusinfantil.com.br>. Acesso em: 9 mai. 2022.
- CONCEIÇÃO, Andreia Vieira da. Por entre parlendas, quadrinhas, cordéis e poemas: tecendo saberes em práticas de alfabetização e letramento. *Revista De Estudos Em Educação E Diversidade - REED*, 1(2), 2020, 339-359.
- CORPUS Escrita: banco de dados preparado por SANTOS, R. S. e TONI, A. (2020). Disponível em: <http://leal.fflch.usp.br/CorpusEscrita>. Acesso em: 9 mai. 2022.
- DE SALLES, Jerusa Fumagalli; PARENTE, Maria Alice Mattos Pimenta. Relação entre desempenho infantil em linguagem escrita e percepção do professor. *Cadernos de Pesquisa*, v. 37, n. 132, set./dez. 2007, p. 687-709.
- DURAN, M. S. Lexicografia e Lingüística de Corpus. In: MAGALHÃES, J. S.; TRAVAGLIA, L. C. (Org.). *Múltiplas Perspectivas em Lingüística*. Uberlândia: Edufu, 2008. p.1793-1800.
- ESTOPÀ, Rosa. Lexicografía especializada escolar: construyendo definiciones a partir de un corpus de definiciones escolares deconstruidas. *Normas. Retos y avances en lexicografía: los diccionarios del español en el eje de la variación lingüística*. Anejo 10, 2019, 249-262.
- ESTOPÀ, Rosa; BEVILACQUA, Cleci Regina. *Meu primeiro dicionário de ciências / Mi primer diccionario de ciencias / My first dictionary of science*. Florianópolis: Arte&Livros, 2022.
- FAAß, Gertrud. Lexicography and corpus linguistics. *The Routledge Handbook of Lexicography*, p. 123-425, 2017.
- FORNARI, Michelle Kühn. O tratamento lexicográfico das palavras gramaticais: discussão teórica e análise de verbetes. *Travessias*, Cascavel, v. 3, n. 3, 2009.
- FUERTES-OLIVERA, Pedro A. A corpus-based view of lexical gender in written Business English. *English for Specific Purposes*, 26.2, 2007, p.219–234.

- GEIGER-JAILLET, Anemone; ROSS, Kristel. Des pratiques langagières mixées en préscolaire à leur visée sociale. In: *Bulletin VALS-ASLA*, 2017, vol. N° spécial, no. 2, p. 77-88.
- GELC (Grupo de Pesquisas em Linguística de Corpus). *Corpus Brasileiro*. In: <http://corpusbrasileiro.pucsp.br/cb/Inicial.html>. Acesso em 10 mai. 2022.
- INEP. *Censo Escolar 2021: divulgação dos resultados*. Diretoria de Estatísticas Educacionais. Brasília, 31 de janeiro de 2022.
- JACINTHO, Mônica Franco (editora responsável). *Itororó: Português*; obra coletiva concebida, desenvolvida e produzida pela Editora Moderna. São Paulo: Moderna, 2019.
- LIMA, Rogério Rodrigues de. O dicionário escolar e a prática docente: possibilidades de utilização e contribuições no processo de leitura e produção escrita. *COIPESU*: João Pessoa, 2015.
- MARCONI, Lucía; RUIZ MIYARES, Leonel; CUTUGNO, Paola. Características distributivas do Espanhol de Cuba a partir da análise de dois estudos linguísticos. *Cadernos de tradução* (Porto Alegre). Porto Alegre, RS, 2018.
- MARIAN, Jane. O estudo da linguística de corpus para a tradução especializada: elaboração de um glossário da área da informática – manutenção de Computadores. *Cultura e Tradução*, João Pessoa, v. 3, n. 1, p. 197-209, 2014. Disponível em: periodicos.ufpb.br/index.php/ct/article/viewFile/21567/12170. Acesso em: 15 ago. 2021.
- MIYARES BERMÚDEZ, Eloína; SUÁREZ, Nancy C. Álamo; MÁRQUÉS, Celia Pérez; HEREDIA, Katia González; RUIZ MIYARES, Leonel. *Léxico activo-funcional del escolar cubano*. Santiago de Cuba: Centro de Lingüística Aplicada, 2006.
- MIYARES BERMÚDEZ, Eloína; ZUBILLAGA, Xabier Artola; LOINAZ, Iñaki Alegría; IPARRAGIRRE, Xabier Arregi; RUIZ MIYARES, Leonel Ruiz; SUÁREZ, Cristina Álamo; MARQUÉS, Celia Pérez. *Las últimas ediciones del Diccionario Básico Escolar de Cuba*. Nomdedeu Rull, Antoni, Esther Forgas Berdet and Maria Bargalló Escrivà (Eds.). 2012. *Avances de lexicografía hispánica I*: 201-213. Tarragona: Universidad Rovira i Virgili.
- MUKHERJEE, Joybrato. Corpus linguistics and language pedagogy: The state of the art and beyond. In: Braun, S., Kohn, K. and Mukherjee, J. (eds.), *Corpus technology and language pedagogy: New resources, new tools, new methods*. Frankfurt: Peter Lang, 2006, 5–24.
- NOMDEDEU-RULL, Antoni & TARP, Sven. Hacia un modelo de diccionario en línea para aprendices de español como LE/L2. *Journal of Spanish Language Teaching* 5 (1), 2018: 50–65.
- OLIVEIRA, Lúcia Pacheco de. Linguística de Corpus: teoria, interfaces e aplicações. *Matraga*, Rio de Janeiro, v.16, n.24, jan./jun. 2009. p. 48-76.
- PAREJO, Isabel García. *El campo semántico “placer” en español*: Estudio de la estructura de los sustantivos del campo léxico “placer” en español y análisis de la misma en el idiolecto literario de doce autores españoles de los siglos XIII e XIX. Tesis doctoral dirigida por el Dr. D. Gregorio Salvador Caja. Universidad Complutense. Facultad de Filología; Madrid; 1997.

- PASCUAL, R. M.; GUEVARA, R. C. L. Developing a children's Filipino speech corpus for application in automatic detection of reading miscues and disfluencies. In: *TENCON 2012 IEEE Region 10 Conference*, 2012, pp. 1-6.
- PEREIRA, Aden Rodrigues. *Análise de base em córpus da tradução de expressões multipalavra no par linguístico português-espanhol*. Tese: Programa de Pós-Graduação em Estudos da Tradução, da Universidade Federal de Santa Catarina. Florianópolis, 2016.
- RANGEL, Ergon. *Com direito à palavra: dicionários em sala de aula*. Brasília: Ministério da Educação, Secretaria de Educação Básica, 2012.
- RIBEIRO, Cacildo Geraldino; DE PAULA, Maria Helena. Dicionários escolares: contribuições no ensino-aprendizagem do léxico. *Anais do SIELP*. Volume 2, Número 1. Uberlândia: EDUFU, 2012.
- RIBEIRO, Cacildo Geraldino; DE PAULA, Maria Helena. O uso do dicionário na escola. *Léxico: investigação e ensino*, p. 33, 2014.
- SEGBERS J.; SCHROEDER S. How many words do children know? A corpus-based estimation of children's total vocabulary size. *Language Testing*. 2017; 34(3): 297-320.
- SINCLAIR, J. M. Corpus and Text - Basic Principles. In: WYNNE, M. (Ed.). *Developing Linguistic Corpora: a Guide to Good Practice*. Oxford: Oxbow Books, 2005.
- SOUSA, Alexandre Melo de. *O uso do dicionário na sala de aula*. III Jornada Nacional de Linguística e Filologia da Língua Portuguesa: Rio de Janeiro, novembro de 2008.
- SUPPES, Patrick. The semantics of children's language. *American Psychologist*, 29(2), 1974, 103–114.
- TALJARD, Elsabé; PRINSLOO, Danie. African language dictionaries for children—a neglected genre. *Lexikos*, v. 29, p. 199-223, 2019.
- TAMAYO LOZADA, J., & RUIZ MIYARES, L. El Diccionario básico escolar en móviles y tabletas con sistema operativo Android. *Serie Científica De La Universidad De Las Ciencias Informáticas*, 14(7), 54-66, 2021.
- TARP, Sven. & FUERTES-OLIVERA, Pedro A. Advantages and Disadvantages in the Use of Internet as a Corpus: The Case of the Online Dictionaries of Spanish Valladolid-Uva. *Lexikos*, 2016, 26: 273-296.
- DINI, Jurnal Pendidikan Anak Usia. e-Dictionaries-based Semantic Gradient: Assisting Preschool Children Connect between Known and New Vocabulary. *Jurnal Obsesi: Jurnal Pendidikan Anak Usia Dini*, v. 6, n. 3, p. 1259-1276, 2022.
- WILD, Kate; KILGARRIFF, Adam; TUGWELL, David. The Oxford Children's Corpus: Using a Children's Corpus. *Lexicography, International Journal of Lexicography*, Volume 26, Issue 2, June 2013, Pages 190–218.