

# Ciência Aberta e suas Contribuições para a Educação Aberta<sup>1</sup>

**Rodrigo Esteves de Lima Lopes<sup>2</sup>**

Universidade Estadual de Campinas/CNPq, Campinas, SP, Brasil

(...) openness could refer to pretty much anything: The process of knowledge creation, its result, the researching individual him or herself, or the relationship between research and the rest of society (FECHER; FRIESIKE, 2014).

**Resumo:** Este ensaio tem por objetivo refletir sobre a relação entre a Ciência Aberta e a Educação Aberta. Advogo em favor de um modelo o qual produtos resultantes da nova cultura científica proposta pela Ciência Aberta devem ser considerados Recursos Educacionais Abertos, assim ajudando a transformar os processos de ensino e aprendizagem. Estes estariam baseados em ações de pesquisa, que incluiriam diferentes segmentos da sociedade, além de saberes e conhecimentos locais. Por conseguinte, tal transformação teria consequências para a formação de professores e das práticas pedagógicas no ensino de línguas. Inicia-se o artigo com uma discussão sobre as motivações deste trabalho, discutindo-se, então, as principais características da Ciência Aberta e seu potencial de contribuições para a Linguística Aplicada.

**Palavras-chave:** Recursos Educacionais Abertos; Ciência Aberta; Educação Aberta; Linguística Aplicada

**Title:** Open Science and its Contributions to Open Education

**Abstract:** This essay aims to discuss the relationship between Open Science and Open Education. I advocate a model in which products resulting from the new scientific culture proposed by Open Science should be considered Open Educational Resources, thus helping to transform teaching and learning processes. These would now be based on research activities, including different segments of society and local knowledge. Therefore, such a transformation would affect teacher education and pedagogical practices in language teaching. The article begins with a discussion of the motivations for this work. It then examines the main characteristics of Open Science and its potential for contributions to Applied Linguistics.

**Keywords:** Open Educational Resources; Open Science; Open Education; Applied Linguistics.

<sup>1</sup> Uma versão anterior deste artigo foi apresentada durante a prova didática do concurso para obtenção do título de *Livre Docente em Linguagem e Tecnologia* pela Universidade Estadual de Campinas em agosto de 2022.

<sup>2</sup> Doutor em Linguística Aplicada pela Universidade Estadual de Campinas. Professor associado na Universidade Estadual de Campinas (UNICAMP) no Instituto de Estudos da Linguagem (IEL)/Departamento de Linguística Aplicada (DLA). ORCID: <https://orcid.org/0000-0003-3681-1553> E-mail: [rll307@unicamp.br](mailto:rll307@unicamp.br)

## Introdução

Este ensaio tem por objetivo refletir sobre algumas questões contemporâneas relacionadas às pesquisas em Linguagem e Tecnologia – especialmente a Ciência Aberta (doravante CA) – e suas possíveis contribuições para a discussão já tradicionalmente estabelecida no âmbito dos Recursos Educacionais Abertos, doravante REA (WILEY, 2002, 2014), no contexto dos Estudos da Linguagem.

Na área de Linguística Aplicada brasileira, estudos recentes têm demonstrado o crescimento das pesquisas de REA, seja pela ampliação da sua disponibilização em repositórios abertos nas diversas áreas de ensino (COSTA et al., 2016), seja pelo crescimento no número de teses e dissertações defendidas no contexto brasileiro (LIMA-LOPES; BIAZI, 2021). Trabalhos também exploram a noção de autoria (COSTA et al., 2020; BEVILÁQUA; VETROMILLE-CASTRO; LEFFA, 2021), que se modifica ao pensarmos nas novas perspectivas instauradas por um modelo de produção de materiais revisáveis, reutilizáveis e readaptáveis, como é o caso dos REA. A reflexão sobre a disponibilidade de recursos também tem sido de suma importância, destacando-se o desenvolvimento de sistemas de autoria abertos (BEVILÁQUA et al., 2017; BEVILÁQUA; COSTA; FIALHO, 2020) e o sistemático levantamento das possibilidades para professores no ensino de língua estrangeira (COSTA; PY; FIALHO, 2017).

Os REA têm promovido uma transformação na cultura de ensino, criando de oportunidades de letramento crítico (BEVILÁQUA; LEFFA; KIELING, 2019; BEVILÁQUA; VETROMILLE-CASTRO; LEFFA, 2021) via construção de espaços e formatos abertos (SILVEIRA, 2012), que, ao contrário dos fechados, desafiam os processos ideológicos e mercadológicos. De fato, cabe ressaltar a importância dos REA não apenas para o desenvolvimento de conteúdos localmente relevantes (AMIÉL, 2012), mas também para a atividade docente em rede (PRETTO, 2012). Neste contexto, os professores se tornam protagonistas, construindo materiais relevantes para a situação de ensino e de aprendizagem em que se encontram. Tal protagonismo é, por si, uma relevante instância de valorização docente que extrapola o papel tradicionalmente atribuído ao professor: consumidor e aplicador de materiais fechados e produtos midiáticos (SANTANA, 2012). Esse novo papel propicia o nascimento de reflexões epistemológicas sobre o uso da tecnologia em situações de ensino (KIELING; LEFFA; BEVILÁQUA, 2019), levando à criação de espaços que vão além da simples perspectiva utilitarista.<sup>3</sup>

As inquietações que deram origem a este ensaio surgiram como resultado de minha participação no *Simpósio Descrição Linguística: Gestão de Dados Linguísticos* (FREITAG et al., 2021), promovido pela ABRALIN em 2020 e no *I Encontro Letramento Científico na Educação Básica*, promovido pelo Pós-Graduação em Educação da UFCG em maio de 2022.

---

<sup>3</sup> Neste artigo, utilizamos a definição de Souza (2011, p. 226), para quem uma “visão utilitarista do conhecimento propagada pela pedagogia liberal, em que se destaca a busca de habilidades e competência, visando à produtividade e a competitividade”.

Por ocasião do primeiro evento, buscamos refletir sobre um dos problemas básicos da área de Estudos da Linguagem: a reprodutibilidade e o gerenciamento de dados. Durante esta reflexão, observamos ser notório o fato de que parte significativa dos estudos em nossa área não oferece a possibilidade de reprodução dos resultados. É claro que a discussão não pressupõe a replicação exata de estudos ou experimentos, como pode ocorrer em outras áreas do conhecimento, mesmo porque o estudo da linguagem é marcado pelo olhar humano e qualitativo, o qual recai até mesmo sobre dados numericamente representados (BIBER; CONRAD; REPPEN, 1998). Nossa reflexão recai sobre como a indisponibilidade pública e aberta de dados e protocolos de pesquisa impede que novos olhares sejam lançados a contextos já estudados.

Já no segundo, a mesa redonda intitulada *Letramento Científico e Digital na Escola Básica* refletiu sobre como as transformações da sociedade nos últimos 30 anos poderiam contribuir com a ampliação e introdução do letramento científico. A mesa se caracterizou pela troca de experiências sobre como levar a pesquisa na área de estudos da linguagem para aula de línguas, com um óbvio impacto de tal perspectiva para a formação docente. Observamos que boa parte dos nossos documentos oficiais de área, assim como da literatura na área ensino de línguas, não coloca a pesquisa como eixo central.

Assim, neste artigo, busco responder a tais inquietações, pelo menos de forma parcial, propondo uma discussão sobre como a formação científica de professores no contexto do ensino de línguas e linguagens pode contribuir para criação de conhecimentos de modo que a aula de língua materna ou estrangeira seja encarada como um espaço aberto para a construção de conhecimentos científicos sobre a sociedade. De forma mais específica, procuro centrar minhas discussões no conceito de Ciência Aberta, estabelecendo relações com as possibilidades filosóficas e pedagógicas trazidas pelos REAs. Acredito que a criação e disponibilização de bases de dados, além da adoção de uma perspectiva atrelada à Ciência Aberta, podem ser um passo importante para uma transformação na maneira como construímos o conhecimento no âmbito das ciências da linguagem.

### **Ciência Aberta: Caracterização e escolas de pensamento**

De acordo com Silva e Silveira (2019), as origens do conceito de Ciência Aberta (CA) estão no incentivo à transparência científica, abrangendo desde o planejamento da pesquisa, até o uso de programas e bases de dados abertas de forma a possibilitar a livre distribuição e reprodutibilidade do estudo. Tal movimento teria como consequência a possibilidade de colaboração entre pesquisadores de diversas áreas e localidades, além da democratização do conhecimento científico. Apesar de toda a transformação que a Ciência Aberta parece trazer à baila, Bartling e Friesike (2014) observam que a criação aberta e colaborativa do conhecimento é mais antiga do que muitos julgam, sendo a disseminação do conhecimento tão importante quanto sua criação.

Por conta disso, o conhecimento científico e sua divulgação sempre foram dependentes de processos tecnológicos. Criações como a escrita e a agricultura, por exemplo,

geraram profundas transformações, constituindo a base tecnológica daquilo que hoje chamamos de sociedade (BARTLING; FRIESIKE, 2014). Da mesma forma, ao passo que a invenção da imprensa por Johannes Gutenberg levou a mais transformações, as manufaturas de produção caligráfica de livros, comuns nas universidades Europeias da Idade Média, foram abandonadas em favor de prensas, responsáveis por criar um mercado de divulgação científica (KATZENSTEIN, 1986). Gilmont (1999) observa que a prensa foi essencial para a constituição do movimento protestante instituído por Lutero. Entre as consequências deste movimento estão não apenas a livre circulação de suas teses, como também um processo de alfabetização em massa para leitura dos livros sagrados cristãos. Vale destacar o papel da religião nesse processo, pois, de acordo com Bartling e Friesike (2014) foi um importante veículo de disseminação da informação, especialmente por abrigar diversos clérigos produtores e divulgadores de conhecimento científico.

O papel das universidades na constituição do que hoje chamamos de internet é significativo desde seus primórdios. Tal envolvimento está no desenvolvimento de sua infraestrutura, assim como na sua utilização desta para o compartilhamento de produções científicas, ambas culminando no desenvolvimento dos protocolos *www* (NEGROPONTE, 1995), entre outros avanços. Seria esperado, então, que a utilização sistemática das diversas plataformas digitais tivesse criado um espaço de compartilhamento até então inexistente, tornando reais colaborações outrora limitadas pela indisponibilidade de dados e ferramentas de pesquisa. Por conta disso, ao pensarmos em nossa sociedade contemporânea, as portas abertas pelas mídias digitais levam a uma nova transformação.

Como sustentam Fecher e Friesike (2014), apesar de todos dos envolvidos no movimento de Ciência Aberta o entenderem como uma prática de cooperação, divulgação e compartilhamento de dados e resultados, a comunidade acadêmica está longe de estabelecer unanimidade neste conceito. Os autores observam a existência de cinco escolas de pensamento, cada uma carregando concepções específicas para no termo: de pensamento público; democrática; pragmática; de infraestrutura; e de impacto.

A primeira, denominada de *Escola de Pensamento Público*, advoga em favor da necessidade de ampliação da audiência que possui acesso ao conhecimento científico. Um dos pressupostos básicos desta escola é que a chamada *web social* (ou 2.0) poderia contribuir para esta abertura, tanto por permitir aos cientistas que abram suas pesquisas, como por permitir o seu acesso por indivíduos não especialistas. Isso levaria a, pelo menos, dois caminhos paralelos a serem trilhados: um que se preocupa com a acessibilidade das pesquisas e outro que tem seu foco na compreensão destes resultados por um público não acadêmico. Entretanto, parte significativa deste movimento ainda segue uma lógica *top-down*, que atribui ao pesquisador e cientista profissional o poder de análise e concessão de seus resultados. A segunda escola, chamada *Escola Democrática*, está comprometida com a abertura da maneira como o conhecimento é gerado. Ela difere da escola anterior à medida que promove a acessibilidade ao conhecimento por meio do acesso aos produtos de pesquisa, especialmente publicações e os dados nelas utilizados. Nesta escola, duas outras linhas parecem surgir: Publicação Aberta e os Dados Abertos. A primeira, diz respeito ao acesso às publicações

científicas pela sociedade, ao passo que a segunda está relacionada ao acesso direto aos dados (FECHER; FRIESIKE, 2014).

Já a terceira, chamada de *Escola Pragmática*, centraliza seus esforços na viabilização das publicações e dados abertos, estando preocupada em otimizar o processo de produção do conhecimento científico por meio do acesso viabilizado aos pares e à sociedade. Como resultado, haveria o envolvimento de pessoas que estariam à margem deste processo, refinando a produção de conhecimento a qual deixa de ser individual para constituir-se colaborativamente. Na quarta perspectiva, chamada de *Escola de Infraestrutura*, a preocupação estaria na infraestrutura necessária para a viabilização dos projetos. Esta escola vê a Ciência Aberta como um desafio tecnológico, buscando desenvolver software, redes computacionais e plataformas que possam tornar as ações de pesquisa possíveis. Por fim, a *Escola de Impacto* busca refletir sobre os sistemas de mensuração de produção acadêmica. Tal preocupação nasce da importância que as métricas acadêmicas possuem para o financiamento de pesquisa e na reputação do pesquisador. Entre as críticas ao modelo tradicional estão: a centralidade dos fatores de impacto, que medem a reputação da revista e não a penetração de um artigo; o tempo e a energia gasto no processo de revisão por pares, que muitas vezes esbarra em preconceitos científicos possibilitados pelo véu da anonimização; além da discussão dos formatos de publicação, muitas vezes atrelada a práticas da era pré-digital (FECHER; FRIESIKE, 2014).

### **Ciência aberta e Linguística Aplicada: algumas questões na contemporaneidade**

Pode-se observar que as escolas propostas por Fecher e Friesike (2014) representam faces complementares, figurando como *calcanhares de Aquiles* daquilo que podemos definir como Ciência Aberta. Isso se dá porque, como apontam Vicente-Saez e Martinez-Fuentes (2018), este movimento está relacionado a transformações culturais e tecnológicas que impactam na forma como nossa pesquisa é planejada, realizada, disponibilizada e avaliada. Assim, apesar da profusão de preocupações e novas possibilidades de Ciência Aberta, acredito que podemos defini-la, à guisa de conclusão, em termos de três eixos principais: a Publicação Aberta — que englobaria o acesso às publicações científicas —; Dados Abertos — que traz questões referentes à disponibilização, incluindo a infraestrutura necessária — e Código Aberto — que demandaria o desenvolvimento de ferramentas de pesquisa e de disponibilização de dados e de publicações (JOMIER, 2017).

A publicação aberta pressupõe o livre acesso aos resultados de pesquisa por meio de livros e revistas acadêmicas que não estejam condicionados a pagamento (*paywalls*). Em muitos casos, mesmo tendo publicado seus artigos/capítulos em edições comerciais, autores compartilham rascunhos “quase-finais” (os chamados *pré-prints*), seja por meio da correspondência individual, seja pela disponibilização em repositórios institucionais comuns em universidades. Geralmente, a publicação aberta tende a ser mais rápida, uma vez que os repositórios não precisam respeitar os prazos de publicação impostos pelos procedimentos editoriais comerciais. Esta prática é chamada pelos periódicos de *self-archiving* e é prevista

por algo entre 70% e 80% das editoras comerciais (SITEK; BERTELMANN, 2014). Parte das editoras comerciais passou a reagir aos processos de autopublicação. Entre as práticas comuns estão: os períodos de embargo, que determina um prazo mínimo de exclusividade para publicação dos originais; a cobrança pela abertura definitiva do original, taxa normalmente paga pelo autor ou financiamento da pesquisa; pagamento por utilização, taxa por acesso único paga pelo utilizador; ou acordo com bibliotecas e agências de fomento (SITEK; BERTELMANN, 2014).

Para além dos *pré-prints*, diversas outras ações possibilitam a publicação livre. Entre elas, está a vinculação dos periódicos a universidades ou associações científicas, e não a editoras comerciais. Esta prática, já comum na Linguística Aplicada brasileira, permite o acesso às publicações sem o pagamento de taxas, ao mesmo tempo em que obriga tais instituições a buscar financiamentos externos para manutenção das publicações. Ainda no contexto brasileiro, a Scielo – iniciativa da FAPESP e do Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde, e, posteriormente, apoiado pelo CNPq –, também pode ser categorizada como uma iniciativa de publicação aberta.

Algo a se observar é que, mesmo com as diversas práticas de resistência às publicações fechadas, boa parte das revistas acadêmicas ainda segue um modelo de publicação pré-digital. Uma prática comum é a disponibilização dos arquivos em PDF paginados em um formato que imita os periódicos em sua versão impressa (BARTLING; FRIESIKE, 2014). Tal disponibilização pode ser considerada uma forma de limitação de acesso, pois exige leitores específicos além dos navegadores. Ele também caracteriza um sub-aproveitamento do potencial de hiperligação, uma vez que as publicações poderiam possuir vídeos, animações e imagens em maior definição, facilitando a compreensão por parte do leitor. Em outras palavras, se o movimento de CA buscar efetivamente romper as barreiras de divulgação de publicações, seria necessária uma ruptura em relação ao sistema de representação textual baseado em tecnologias fechadas e proprietárias (SIDLER, 2014), como é o caso do formato PDF.

Os dados abertos, por seu turno, ocasionariam a disponibilização do conteúdo utilizado em uma pesquisa de forma livre e aberta para a comunidade (JOMIER, 2017). Assim,

(...) The rationale behind Open Data in science is in this case researcher-centric; it is a conjuncture that fosters meaningful data mining and aggregation of data from multiple papers. Put more simply, Open Data allows research synergies and prevents duplication in the collection of data (FECHER; FRIESIKE, 2014, p. 26).

A prática de disponibilização exige, todavia, uma reflexão sobre suas implicações e os modelos de viabilização. Diferentes áreas do estudo da linguagem possuem diversas exigências em termos da natureza dos dados e seus processos de consolidação. Tal diversidade também demanda discussões de caráter ético, uma vez que diversas áreas da LA têm suas análises baseadas em dados não anônimos, como é o caso de interações em sala de aula, testes paramétricos, redações e entrevistas. Esses são os motivos pelos quais protocolos claros deveriam ser criados nas diversas subáreas.

Apesar de não ser unânime, a abertura de dados é uma prática comum em algumas áreas do estudo da linguagem, especialmente a Linguística do Corpus, tendo em vista que

diversos corpora são disponibilizados para pesquisa por meio de concordâncias, listas de palavras ou mesmo em formato bruto. Um exemplo de um repositório desta natureza seria o *CorpusData*,<sup>4</sup> que oferece grandes amostras de dados linguísticos em diversas línguas como forma de promoção de seus serviços pagos. Exemplos dentro do contexto brasileiro são: o projeto Teletandem Brasil e o Profis. O corpus do projeto Teletandem Brasil (ARANHA; LEONE, 2017), desenvolvido nos últimos anos por João Telles e Solange Aranha nos diversos campi da UNESP. Neste programa, voltado para o ensino e aprendizagem de português para estrangeiros e inglês para brasileiros, os pesquisadores estão consolidando um corpus multimidiático e multi-dados, no qual sessões gravadas de Teletandem estão associadas às suas transcrições e a tarefas adjacentes, como redações dos alunos. Na UNICAMP, recentemente, está sendo realizado um processo de sistematização do corpus de redações do Profis, chamado PROCORP. O Profis é um projeto de inserção na universidade que visa a formação de alunos da escola pública da região de Campinas. O programa oferece 120 vagas e tem a duração de dois anos, garantindo vagas em cursos de graduação na UNICAMP aos seus aprovados. Os dados compreendem redações realizadas pelos alunos no âmbito do programa desde 2011, sua data de início. Vale notar, entretanto, que o acesso aos dados vai além dos diversos corpora, pois ela pode também estar relacionada à disponibilização de tabelas consolidadas, como foi realizado por Lima-Lopes e Biazzi (2021).<sup>5</sup>

Outro ponto não menos importante diz respeito à tutela destes dados, ou seja, como sugerido por Freitag e colaboradores (2021), associações de área e universidades poderiam promover tal tutela, dada a sua característica perene. Um exemplo desta prática seria o REDU, Repositório de Dados da UNICAMP,<sup>6</sup> que abriga o insumo de pesquisas realizadas no contexto da instituição.

O código aberto, por fim, estaria relacionado ao movimento de software livre (BRETTHAUER, 2001) cujas origens estariam nos processos de resistência aos softwares fechados e comerciais e das restrições impostas pelos valores, políticas de atualização. Entre as principais vantagens do uso de software livre está, por conseguinte, a possibilidade de utilização por recursos técnicos de forma mais generalizada, uma vez que, por incentivar a colaboração na comunidade de programadores e usuários, sua utilização é relevante para os contextos de pesquisa, ensino e aprendizagem. Infelizmente, o uso de software livre não é uma prática de pesquisa comum na área de Linguística Aplicada.

Apesar de a discussão sobre processos metodológicos na LA ser algo salutar, ela parece estar pouco presente nas publicações da área. Em importantes periódicos da área, como a *Trabalhos em Linguística Aplicada*,<sup>7</sup> a *Revista Brasileira de Linguística Aplicada*<sup>8</sup> e a *Revista*

---

<sup>4</sup> Disponível em: <https://www.corpusdata.org/>. Acesso em 02 de agosto de 2022.

<sup>5</sup> Lima-Lopes e Biazzi (2021) disponibilizaram os dados tabulados de seu levantamento em um repositório aberto. Disponível em <http://www.github.com/rl307>. Acesso em 02 de agosto de 2022.

<sup>6</sup> Disponível em: <https://redu.unicamp.br/>. Acesso em 02 de agosto de 2022.

<sup>7</sup> Disponível em: [https://www.scielo.br/scielo.php?script=sci\\_serial&pid=0103-1813&lng=en&nrm=iso](https://www.scielo.br/scielo.php?script=sci_serial&pid=0103-1813&lng=en&nrm=iso). Acesso em 16 de dezembro de 2022.

<sup>8</sup> Disponível em: [https://www.scielo.br/scielo.php?script=sci\\_serial&pid=1984-6398&lng=en&nrm=iso](https://www.scielo.br/scielo.php?script=sci_serial&pid=1984-6398&lng=en&nrm=iso). Acesso em 16 de dezembro de 2022.

*Linguagem e Ensino*,<sup>9</sup> observa-se um total de 127 artigos que possuem a palavra *metodologia* em seus metadados. Deles, apenas 13 são dedicados a reflexões com o foco central no processo metodológico de coleta e análise de dados. Os demais discutem o processo de ensino ou aprendizagem de línguas, incluindo suas mais diversas áreas e abordagens.

De fato, duas podem ser as motivações para tais resultados. Na primeira, ressalta-se que o ensino de línguas é uma área que se confunde com o próprio surgimento da LA (CELANI, 1992; MCNAMARA, 2012), ao passo que, na segunda, a LA tem desenvolvido uma tradição de análise qualitativa de dados que parece prescindir do uso de computadores (MOITA LOPES, 1994). Como já colocado por pesquisadores no contexto do GT de Linguagem e Tecnologias da ANPOLL (FINARDI et al., 2020), o movimento metodológico da LA tem se caracterizado pela valorização de processos interpretativos que não utilizam sistemas automáticos ou semiautomáticos para a análise e classificação, tendendo a desvalorizá-los. O uso de programas é mais comum em abordagens que se filiam a discussões quantitativas ou baseadas em métodos mistos (JOHNSON; ONWUEGBUZIE; TURNER, 2007; O'HALLORAN et al., 2018), com pouca representação na LA.

Entre as diversas ferramentas utilizadas em pesquisas, gostaria de destacar o AntConc, desenvolvido por Laurance Antony (2022); o CATMA, desenvolvido por Jan Meister e, atualmente, liderado por Evelyn Gius (GIUS et al., 2022); o ELAN (2022), desenvolvido pelo Max Planck Institute for Psycholinguistics; o Language Archive; e os repositórios de *scripts* em linguagem R. O primeiro é um software de Linguística do Corpus que realiza o levantamento de dados tais como concordâncias, listagem de palavras, cálculo de palavras-chave, plotagem de ocorrência, cálculo de colocados e levantamento de n-gramas.<sup>10</sup> Apesar de ser livre para a utilização sem custos e ser multiplataforma, o AntConc não disponibiliza seu código fonte sem solicitação prévia.

Já o CATMA, por seu turno, é um programa de análise qualitativa e marcação textual, oferecendo um modelo de anotação qualitativo adaptável a diversas línguas e teoricamente agnóstico. Ou seja, com este programa seria possível marcar exemplos do texto que satisfaçam uma ou mais categoria de análise, sendo que tais marcações são posteriormente recuperadas, possibilitando o levantamento sistemático de exemplos. Por ser acessado via navegador, não são necessárias instalações ou compiladores, apesar de o código fonte estar disponível. O ELAN é um software de transcrição multimodal que permite anotação em áudio, vídeo e imagem de forma simultânea. Ele se organiza no formato de uma linha do tempo, algo deveras similar a um software de edição de vídeo, permitindo que o usuário sobreponha informações sobre os diversos modos de linguagem analisados. Permite-se, assim, a visualização dessas anotações concomitantemente à exibição da peça audiovisual.

Por fim, há um movimento de disponibilização de *scripts* computacionais utilizados em pesquisas na área de Linguística Aplicada por programadores utilizando a linguagem R (R CORE TEAM, 2021) promovendo integração com a área de Humanidades Digitais; exemplo disso é o

---

<sup>9</sup> Disponível em: <https://periodicos.ufpel.edu.br/ojs2/index.php/rle>. Acesso em 16 de dezembro de 2022.

<sup>10</sup> Para uma discussão sobre esta metodologia, recomendo a leitura de Berber Sardinha (2004) e Lima-Lopes (2016).

repositório do Grupo MiDiTeS,<sup>11</sup> sediado na UNICAMP. Apesar de exigirem conhecimento prévio na área de programação, este tipo de iniciativa ajuda a divulgar procedimentos metodológicos para a sociedade como um todo, uma vez que o código é de acesso totalmente aberto.

### **Ciência aberta e suas perspectivas para Linguística Aplicada**

De fato, a academia é uma das instituições com discurso mais hermético. Devido a isso, uma importante consequência da adoção de um modelo de ciência aberta reside na evolução da forma como a sociedade interage com o conhecimento. Tradicionalmente, ciência é algo construído por pesquisadores ou cientistas, pessoas que tomam o fazer científico como profissão (BARTLING; FRIESIKE, 2014). A possibilidade de abertura desta *caixa de pandora* levaria a novas frentes de colaboração, envolvendo segmentos da sociedade sistematicamente excluídos da produção do saber científico e estendendo as redes de colaboração para além dos muros de nossos departamentos e institutos. De nada implicaria uma transformação na forma e velocidade que os conhecimentos são publicados se eles não estiverem disponíveis para serem replicados ou aplicados em novos contextos.

As possibilidades de replicação que nascem da CA são importantes para a LA. Primeiramente, a replicabilidade exige a documentação precisa dos passos e protocolos de pesquisa. Tal documentação ganha traços relevantes por deixar mais claro o processo de geração de dados, sua tutela e análise, desenvolvendo credibilidade da comunidade (MEIE, 1995), algo que também se aplica a estudos qualitativos (KING, 1995).

Outras questões podem estar relacionadas à quebra de uma tradição, que, desde os meados do século XIV — época em que o atual modelo de Ciência começa a tomar forma — está centrada na competição individual e no produto científico privado com restrições de uso (ALBAGLI, 2015; ARZA; FRESSOLI, 2018). Este modelo impõe traços mercadológicos a conhecimentos que poderiam ser incluídos em uma reserva social, causando dependência de editoras privadas as quais se tornam detentoras dos direitos autorais, vendidos em forma de assinaturas a altos custos. Ao mesmo tempo, os sistemas do capitalismo de rede e vigilância (ZUBOFF, 2019) passam a aferir lucros por meio da informação que seria, tecnicamente, oferecida de forma gratuita pelos usuários. Desse modo, de um lado, há a necessidade de proteção do indivíduo da continuidade do capitalismo de rede e, de outro, há a necessidade cada vez maior de trazer o conhecimento científico para um lugar mais próximo da sociedade (ALBAGLI, 2015) e, no caso do contexto da área de Estudos da Linguagem, também da formação de professores.

Uma proposta de utilização de Ciência Aberta dentro do âmbito da LA, aplicada de forma a ampliar e discutir as experiências já existentes, deveria levar em conta, pelo menos três perspectivas:

#### **i) Construção do conhecimento linguístico local**

---

<sup>11</sup> Acessível em: <https://github.com/rll307>. Acesso em 16 de dezembro de 2022.

Assim como os Recursos Educacionais Abertos, os processos baseados em Ciência Aberta têm um importante papel na construção de uma educação voltada não apenas para pautas gerais e que refletem uma macropolítica educacional, mas também que valorizam conhecimentos e valores locais. Em outras palavras, parte significativa de um trabalho de CA no contexto da LA estaria na coleta e análise de dados coletados localmente. Tal possibilidade teria um importante papel na construção de uma consciência das diversas variações de linguagem as quais, quando contrastadas à norma culta, podem trazer reflexões importantes oportunidades de aprendizado e construção de identidade.

### **ii) Inserção da pesquisa como forma de construção de conhecimento**

A análise e coleta de dados locais levaria à necessidade de inserção de componentes de pesquisa tanto na formação de professores, quanto nos processos metodológicos e materiais de ensino. Tal possibilidade dá outra dimensão à aula de língua, uma vez que o livro didático e o conhecimento advindo das políticas educacionais passam a ser um dos pontos de comparação em uma teia de variantes e estruturas linguísticas muito mais complexas. Compreender as diferenças, semelhanças e o lugar social ocupado por cada uma delas seria o resultado esperado deste tipo de abordagem, na qual os aprendizes construíram suas próprias versões de uma gramática social da língua (CHAMBERS, 2010). Naturalmente, esta abordagem não poderia prescindir da exploração de bases de dados e programas de análise linguística (tanto qualitativa(o)(s) quanto quantitativa(o)(s)).

### **iii) Formação tecno-filosófica de professores**

Nesse sentido, professores precisam de uma formação que parta de perspectivas analíticas de linguagem embasadas em uma perspectiva crítica acerca dos meios. Precisamos ir além de uma perspectiva utilitarista do uso da tecnologia e partir para a compreensão de seus impactos na sociedade e nos meios de produção de capital e de linguagem. Para além disso, vale ressaltar que a formação tecnológica deve caminhar por duas searas. A primeira estaria em preparar professores para o inesperado, uma vez que a mutabilidade dos processos tecnológicos faz com que a existência de *blueprints* se torne impossível. A segunda estaria relacionada ao olhar crítico perante o inerente processo de plataformização (O'NEIL, 2016; ZUBOFF, 2019), que leva à reflexão sobre processos de colonização tecnológica.

## **Considerações finais**

Este ensaio teve por objetivo discutir o conceito de Ciência Aberta – e suas possíveis contribuições – para a discussão já tradicionalmente estabelecida no âmbito dos Recursos Educacionais Abertos no contexto dos Estudos da Linguagem. O trabalho se iniciou com a definição de CA, passando por uma análise sobre suas características básicas, com especial ênfase em suas escolas de definição. Em um segundo momento, discutiu-se a CA e sua relação com a LA, passando-se, então, para uma reflexão sobre possíveis contribuições e transformações que seu modelo poderia trazer para nossa área.

Em uma associação direta entre as principais características de ambos os modelos, podemos observar que a CA pode contribuir para a construção de experiências de Educação

Aberta, em especial se a encarmos como um conjunto de Recursos Abertos. Quanto ao R(euso), a reutilização de dados é um pressuposto da CA, que pode partir de dados já disponíveis para ampliação e novas reflexões. Além disso, a prática dos dados abertos permite a comparação entre resultados obtidos em diversas realidades, um ponto relevante para o reconhecimento dos fenômenos da linguagem. A R(evisão) também é algo natural à CA, uma vez que a possibilidade de revisão de resultados é implícita a um modelo que pressupõe a reutilização de dados, que agora podem ser observados a partir de novas subjetividades. No caso do (r)emix, uma das possibilidades trazidas pela CA está em comparar resultados provenientes de diversas bases de dados disponíveis online ou localmente produzidas. Esta comparação, necessariamente, leva a uma nova visão dos dados disponibilizados. Já a (r)etenção de dados e artigos distribuídos no modelo da CA, comumente, continuam à disposição para novos estudos. Talvez r(edistribuir) seja uma das questões mais sensíveis quando pensamos em associar modelos de CA ao trabalho com REA. Isso porque, muitas vezes, os dados não podem ser redistribuídos diretamente, por questões referentes à ética em pesquisa. Tal questão traz à baila, mais uma vez, algo já colocado em momentos anteriores deste trabalho, a necessidade de desenvolvimento de protocolos de pesquisa e distribuição de dados a serem seguidos pelos pesquisadores em LA.

Como proposto por Albagli (2015), a CA pode ser vista como uma reação a um modelo de capitalismo que se ancora na apropriação de informação e conhecimento produzidos coletivamente. Tal tensão se estabelece como uma reação ao modelo de propriedade intelectual, que se caracteriza por uma escassez artificial, que regula algo disponível e fértil, o resultado das diversificadas interações marcadas pela produção de conhecimento. Tal perspectiva, como colocada por Albagli (2015), ecoa o que Lessig (2004) caracteriza como um fechamento da cultura, ao construir uma massa de direitos autorais baseados em informações de origem livre, está-se impondo um modelo que utiliza a cultura como forma de construção de bens privados.

Concordo com o defendido por Albagli (2015) e Lessig (2004), pois são necessárias reações ao processo de captura e privatização do coletivo. Acredito que a Ciência Aberta tem este potencial, não apenas por abrir a *caixa de pandora* dos métodos e resultados científicos, mas também por promover discussões mais amplas com a sociedade. Presumivelmente, tal discussão incluiria outros pesquisadores que, por meio do reuso e revisão, por exemplo, poderiam ampliar o escopo de pesquisas já existentes. Além disso, a CA tem o potencial de incluir a sociedade mais ativamente na produção da ciência. É natural pensar que a divulgação livre e aberta dos resultados de pesquisa poderia caracterizar tal inclusão. Todavia, acredito que a CA pode ir além ao tornar livres bases de dados e ferramentas que podem ser utilizadas por qualquer pessoa. O resultado, por conseguinte, poderia ser uma quebra no sistema hierárquico que hoje caracteriza a produção do conhecimento científico.

Ao imaginar este cenário, naturalmente, deve-se observar a necessidade de inclusão do pensamento científico como a base da construção do conhecimento (individual e coletivo) sobre a linguagem. Uma consequência seria a integração do cabedal de métodos e perspectivas oferecidas pela CA na formação de professores, nas práticas de sala de aula e na

elaboração de documentos oficiais, criando-se um contexto no qual a linguagem (e, por pressuposto, a língua) não seria simplesmente um bem a ser aprendido, mas também conhecimento a ser construído.

### Agradecimentos

Gostaria de agradecer ao CNPq pelo financiamento desta pesquisa.

### Referências

- ALBAGLI, S. Ciência aberta em questão. Em: ALBAGLI, S.; MACIEL, M. L.; ABDO, A. H. (Ed.). *Ciência aberta, questões abertas*. Brasília: IBICT/UNIRIO, 2015. p. 9–26.
- AMIEL, T. Educação aberta: configurando ambientes, práticas e recursos educacionais. Em: SANTANA, B.; PRETTO, C.; LUCCA, N. (Ed.). *Recursos Educacionais Abertos: práticas colaborativas políticas públicas*. Salvador: Edufba, 2012.
- ANTHONY, L. *AntConc (Version 4.1.0) [Computer Software]* Tokyo: Waseda University, , 2022. . Disponível em: <<https://www.laurenceanthony.net/software>>. Acesso em: 4 ago. 2022.
- ARANHA, S.; LEONE, P. The development of DOTI (Data of oral teletandem interaction). Em: FISHER, D.; BEIBWENGER, M. (Ed.). *Investigating computer-mediated communication: corpus-based approaches to language in the digital world*. Ljubljana: University of Ljubljana, Faculty of Arts. [s.l: s.n.]1p. 172–192.
- ARZA, V.; FRESSOLI, M. Systematizing benefits of open science practices. *Information Services & Use*, v. 37, n. 4, p. 463–474, 8 jan. 2018.
- BARTLING, S.; FRIESIKE, S. Towards Another Scientific Revolution. Em: BARTLING, S.; FRIESIKE, S. (Ed.). *Opening Science*. Cham: Springer International Publishing, 2014. p. 3–15.
- BERBER SARDINHA, T. *Linguística de Corpus*. São Paulo: Manole, 2004.
- BEVILÁQUA, A. F. et al. Ensino de línguas online: um sistema de autoria aberto para a produção e adaptação de recursos educacionais abertos. *Calidoscópico*, v. 15, n. 1, p. 190–200, 3 maio 2017.
- BEVILÁQUA, A. F.; COSTA, A. R.; FIALHO, V. R. Made in Brazil: Inovações tecnológicas no âmbito de um sistema de autoria aberto para o ensino de línguas. *Revista Linguagem em Foco*, v. 12, n. 2, p. 8–29, 21 set. 2020.
- BEVILÁQUA, A. F.; LEFFA, V. J.; KIELING, H. dos S. Espanhol como língua estrangeira (e/le), acción poética e xenofobia: uma experiência pedagógica com um Recurso Educacional Aberto na perspectiva dos letramentos críticos. *Trabalhos em Linguística Aplicada*, v. 58, p. 759–778, 23 set. 2019.
- BEVILÁQUA, A. F.; VETROMILLE-CASTRO, R.; LEFFA, V. J. Eu não sou trapaceiro(a): a produção de REA para Letramentos Críticos e Competência Simbólica. *Revista Brasileira de Linguística Aplicada*, v. 21, p. 929–954, 14 maio 2021.
- BIBER, D.; CONRAD, S.; REPPEN, R. *Corpus linguistics: investigating language structure and use*. Cambridge ; New York: Cambridge University Press, 1998.

- BRETTTHAUER, D. Open Source Software: A History. *Published Works*, v. 7, p. 1–23, 26 dez. 2001.
- CELANI, M. A. A. Afinal, o que é Linguística Aplicada? Em: PASCHOAL, M. S. Z.; CELANI, M. A. A. (Ed.). *Linguística Aplicada: da Aplicação da Linguística à Linguística Transdisciplinar*. São Paulo: EDUC, Editora da PUC-SP, 1992.
- CHAMBERS, A. What is data-driven learning. Em: O'KEEFFE, A.; MCCARTHY, M. (Ed.). *The Routledge handbook of corpus linguistics*. London: Routledge, 2010. p. 345–358.
- COSTA, A. R. et al. Contribuindo com o estado da arte sobre Recursos Educacionais Abertos para o ensino e a aprendizagem de línguas no Brasil. *Veredas - Revista de Estudos Linguísticos*, v. 20, n. 1, p. 1–20, 16 ago. 2016.
- COSTA, A. R. et al. Recursos educacionais abertos no Brasil: “plágio” e “(co)autoria” nas concepções de professores de línguas. *REVISTA DIÁLOGO E INTERAÇÃO*, v. 14, n. 1, p. 370–395, 3 nov. 2020.
- COSTA, A. R.; PY, L. O.; FIALHO, V. R. Opções em recursos educacionais abertos para o ensino de Espanhol no Brasil. *Hipertextus - Revista Digital*, v. 17, n. 1, 2017. Disponível em: <<https://periodicos.ufpe.br/revistas/hipertextus/article/view/247956>>. Acesso em: 6 jul. 2022.
- ELAN [Computer software]. NijmegenMax Planck Institute for Psycholinguistics, The Language Archive, 2022. Disponível em: <<https://archive.mpi.nl/tla/elan>>. Acesso em: 8 abr. 2022.
- FECHER, B.; FRIESIKE, S. Open Science: One Term, Five Schools of Thought. Em: BARTLING, S.; FRIESIKE, S. (Ed.). *Opening Science*. Cham: Springer International Publishing, 2014. p. 17–47.
- FINARDI, K. R. et al. Uma rodada de perguntas com os membros do Grupo de Trabalho Linguagem e Tecnologias (ANPOLL). *Revista Linguagem em Foco*, v. 12, n. 2, p. 1–31, 28 ago. 2020.
- FREITAG, R. M. K. et al. Desafios da gestão de dados linguísticos e a Ciência Aberta. *Cadernos de Linguística*, v. 2, n. 1, p. 1–19, 26 abr. 2021.
- GILMONT, J.-F. Protestan Reformations and Reading. Em: CAVALLO, G.; CHARTIER, R. (Ed.). *A history of reading in the West*. Studies in print culture and the history of the book. Tradução Lydia G. Cochrane. Amherst: University of Massachusetts Press, 1999.
- GIUS, E. et al. CATMAZenodo, 6 abr. 2022. Disponível em: <<https://zenodo.org/record/1470118>>. Acesso em: 4 ago. 2022.
- JOHNSON, R. B.; ONWUEGBUZIE, A. J.; TURNER, L. A. Toward a Definition of Mixed Methods Research. *Journal of Mixed Methods Research*, v. 1, n. 2, p. 112–133, abr. 2007.
- JOMIER, J. Open science – towards reproducible research. *Information Services & Use*, v. 37, n. 3, p. 361–367, 7 nov. 2017.
- KATZENSTEIN, U. E. *A Origem do Livro: da idade da Pedra ao advento da impressão tipográfica no Ocidente*. [s.l.] Hucitec / Instituto Nacional do Livro/Fundação Nacional Pró-Memória, 1986.
- KIELING, H. D. S.; LEFFA, V. J.; BEVILÁQUA, A. F. A implementação do ensino híbrido no ensino de inglês durante a formação docente. *Caderno Seminal*, v. 33, n. 33, 30 out. 2019. Disponível em: <<https://www.e-publicacoes.uerj.br/index.php/cadernoseminal/article/view/41868>>. Acesso em: 9 jul. 2020.

- KING, G. Replication, Replication. *PS: Political Science and Politics*, v. 28, n. 3, p. 444–452, 1995.
- LESSIG, L. *Free culture: how big media uses technology and the law to lock down culture and control creativity*. [s.l.] Petter Reinholdtsen, 2004.
- LIMA-LOPES, R. E. de. Oficina de Introdução à Linguística do Corpus: Relato de Experiência. Em: Anais do VIII Escola Brasileira de Linguística Computacional & XIII Encontro de Linguística de Corpus, João Pessoa, Brasil. *Anais...* Em: VIII ESCOLA BRASILEIRA DE LINGUÍSTICA COMPUTACIONAL & XIII ENCONTRO DE LINGUÍSTICA DE CORPUS. João Pessoa, Brasil: Editora Edgard Blücher, mar. 2016. Disponível em: <10.5151/sosci-viiiieblc-xiii-elc-05\_artigo\_02>. Acesso em: 13 jul. 2020.
- LIMA-LOPES, R. E. de; BIAZI, T. M. D. Cartografia de pesquisas nacionais sobre REA. *Revista Linguagem em Foco*, v. 12, n. 3, p. 172–196, 22 jan. 2021.
- MCNAMARA, T. Poststructuralism and Its Challenges for Applied Linguistics. *Applied Linguistics*, v. 33, n. 5, p. 473–482, 2012.
- MEIE, K. J. Replication: A View From the Streets. *PS: Political Science and Politics*, v. 28, n. 3, p. 456–459, 1995.
- MOITA LOPES, L. P. da. Pesquisa interpretativista em Linguística Aplicada: a linguagem como condição e solução. *DELTA: Documentação e Estudos em Linguística Teórica e Aplicada*, v. 10, n. 2, 1994.
- NEGROPONTE, N. *Being digital*. 1. ed. New York: Knopf, 1995.
- O'HALLORAN, K. L. et al. A Digital Mixed Methods Research Design: Integrating Multimodal Analysis With Data Mining and Information Visualization for Big Data Analytics. *Journal of Mixed Methods Research*, v. 12, n. 1, p. 11–30, jan. 2018.
- O'NEIL, C. *Weapons of math destruction: how big data increases inequality and threatens democracy*. First edition ed. New York: Crown, 2016.
- PRETTO, N. Professores-autores em rede. Em: SANTANA, B.; PRETTO, C.; LUCCA, N. (Ed.). *Recursos Educacionais Abertos: práticas colaborativas políticas públicas*. Salvador: Edufba, 2012.
- R CORE TEAM. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2021.
- SANTANA, B. Materiais didáticos digitais e recursos educacionais abertos. Em: SANTANA, B.; PRETTO, C.; LUCCA, N. (Ed.). *Recursos Educacionais Abertos: práticas colaborativas políticas públicas*. Salvador: Edufba, 2012. p. 133–142.
- SIDLER, M. Open Science and the Three Cultures: Expanding Open Science to all Domains of Knowledge Creation. Em: BARTLING, S.; FRIESIKE, S. (Ed.). *Opening Science*. Cham: Springer International Publishing, 2014. p. 81–85.
- SILVA, F. C. C. da; SILVEIRA, L. da. O ecossistema da Ciência Aberta. *Transinformação*, v. 31, p. e190001, 2019.
- SILVEIRA, S. A. da. Formatos abertos. Em: SANTANA, B.; PRETTO, C.; LUCCA, N. (Ed.). *Recursos Educacionais Abertos: práticas colaborativas políticas públicas*. Salvador: Edufba, 2012. p. 109–120.
- SITEK, D.; BERTELMANN, R. Open Access: A State of the Art. Em: BARTLING, S.; FRIESIKE, S.

(Ed.). *Opening Science*. Cham: Springer International Publishing, 2014. p. 139–153.

SOUZA, Elisabete Gonçalves. Sociedade da informação e reestruturação produtiva: crítica à dimensão utilitarista do conhecimento. *Transinformação*, v. 23, p. 219–226, Dez 2011.

VICENTE-SAEZ, R.; MARTINEZ-FUENTES, C. Open Science Now: A Systematic Literature Review for an Integrated Definition. *Journal of Business Research*, v. 88, p. 428–436, jul. 2018.

WILEY, D. *The Access Compromise and the 5th R – improving learning*. Disponível em: <<https://opencontent.org/blog/archives/3221>>. Acesso em: 26 fev. 2021.

WILEY, D. A. Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. Em: WILEY, D. A. (Ed.). *The Instructional Use of Learning Objects: Online Version*. Bloomington, Ind: Agency for Instructional Technology/Association for Educational Communications & Technology, 2002.

ZUBOFF, S. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. Main edição ed. [s.l.] Profile Books, 2019.

Recebido em: 26/08/2022.

Aceito em: 11/12/2022.