



# Gradus

Revista Brasileira de  
Fonologia de Laboratório

Vol. 5, nº 2

Dezembro de 2020

ISSN 2526-2718

# *Gradus*

Revista Brasileira de Fonologia de Laboratório

Vol. 5, nº 2

ISSN: 2526-2718

Curitiba, PR

Dezembro de 2020

DOI <https://doi.org/10.47627/gradus.v5i2>

**Responsável Técnico:** Luiz Cláudio Silveira Duarte

Revista filiada ao  
Programa de Pós-Graduação em Letras da  
Universidade Federal do Paraná  
Edifício D. Pedro I, 10º andar, sala 1018/1019  
Rua General Carneiro, 460  
80060-150 Curitiba, PR – Brasil

*Gradus* é um periódico digital, com duas edições anuais. Todo o conteúdo está integralmente disponível na Internet.

<https://gradusjournal.com/>

Os direitos autorais dos trabalhos publicados pertencem a seus autores. *Gradus* segue a política de acesso aberto (*open access*), com o objetivo de favorecer a disseminação do conhecimento.

Aplica-se a licença *Creative Commons Atribuição-Não Comercial* a todos os textos publicados na *Gradus*.

[https://creativecommons.org/licenses/by-nc/4.0/deed.pt\\_BR](https://creativecommons.org/licenses/by-nc/4.0/deed.pt_BR)

*Gradus* publica artigos acadêmicos sobre fonologia de laboratório. Interessados em submeter trabalhos para publicação devem consultar as diretrizes para autores, disponíveis na página do periódico na Internet.

EX RETE SCIENTIA   
From the Net comes Knowledge

## **Produção e hospedagem**

ERS Academic Services

64 Nile Street

LONDON N1 7SR

United Kingdom

<https://exretescientia.com/>

A diagramação da *Gradus* segue preceitos desenvolvidos por Edward R. Tufte para a diagramação de seus livros sobre a expressão visual da informação.<sup>1</sup>

A principal vantagem desta diagramação é colocar informações complementares diretamente ao lado do texto principal. As notas marginais são usadas para referências bibliográficas, para imagens, para referências internas, e quaisquer outras informações convenientes, tornando assim a leitura mais confortável.

<sup>1</sup> Por exemplo, Edward R. TUFTE (2006). *Beautiful Evidence*. Cheshire (CT): Graphics Press.

# *Equipe editorial*

## Editores

**Dra. Adelaide Hercília Pescatori Silva**  
Universidade Federal do Paraná

**Dr. Ubiratã Kickhöfel Alves**  
Universidade Federal do Rio Grande do Sul

## Conselho Editorial

**Dr. Alessandro Rodrigues Meireles**  
Universidade Federal do Espírito Santo

**Dr. André Nogueira Xavier**  
Universidade Federal do Paraná

**Dra. Carmen Lúcia Barreto Matzenauer**  
Universidade Católica de Pelotas

**Dra. Denise Kluge**  
Universidade Federal do Rio de Janeiro

**Dr. Didier Demolin**  
Université Sorbonne Nouvelle – Paris 3

**Dr. Felipe Flores Kupske**  
Universidade Federal da Bahia

**Dr. Francisco de Oliveira Meneses**  
Universidade Federal de São Carlos

**Dra. Giovana Ferreira-Gonçalves**  
Universidade Federal de Pelotas

**Dra. Izabel Seara**  
Universidade Federal de Santa Catarina

**Dra. Larissa Berti**  
Universidade Estadual Paulista Júlio de Mesquita Filho  
(*campus Marília*)

**Dra. Maria de Fatima de Almeida Baia**  
Universidade Estadual do Sudoeste da Bahia

**Dra. Marisa Cruz**  
Universidade de Lisboa

**Dr. Miguel de Oliveira Jr.**  
Universidade Federal de Alagoas

**Dr. Pedro Luis Luchini**  
Universidad Nacional de Mar del Plata

**Dr. Reiner Vinicius Perozzo**  
Universidade Federal do Rio Grande do Sul

**Dr. Ronaldo Lima Jr.**  
Universidade Federal do Ceará

**Dr. Ryan Keith Shosted**  
University of Illinois

**Dra. Vera Pacheco**  
Universidade Estadual do Sudoeste da Bahia

## *Conteúdo*

<b>Editorial</b>	5
<b>From the Editors</b>	7
<b>Editorial (español)</b>	9
<b>Artigos</b>	
<b>Redução vocálica: a gradiência na reorganização segmental</b>	12
<i>Cecília Valle Souza Toledo, Thaís Cristófaró Alves da Silva</i>	
<b>Debates</b>	
<b>Uma introdução à ciência aberta e ao compartilhamento de dados científicos de pesquisa</b>	33
<i>Karolayne Costa Rodrigues de Lima, Marcos Sfair Sunye</i>	

## Editorial

Apresentamos aos leitores o número 2 do volume 5 da *Gradus*. Num ano desafiador, como foi 2020, em que estivemos impedidos de atividades presenciais diversas – entre elas a coleta de dados para experimentos de produção e percepção de fala – consideramos uma vitória a publicação deste número, sobretudo porque conseguimos manter o padrão de qualidade que a *Gradus* faz questão de buscar.

Nesse sentido, cabe lembrar que, embora não exijamos que os autores que submetem seus trabalhos à *Gradus* sejam doutores, por entendermos que a qualidade de um artigo não está necessariamente atrelada à titulação de seu autor, nossos pareceristas fazem um trabalho primoroso e cuidadoso, de modo que a avaliação cega por pares, a que cada artigo é submetido, garante que tenhamos textos de excelência. E essa estratégia tem levado a *Gradus* a se consolidar como uma revista de referência na área de Fonologia de Laboratório.

As limitações que a pandemia nos impõe, conjugadas à política de publicação de bons trabalhos, resultam num número de apenas dois textos, um artigo e um texto na seção Debates.

O artigo, de autoria de Cecília Toledo e Thaís Cristófar, “Redução vocálica: a gradiência na organização segmental”, descreve a redução e o apagamento envolvidos na produção da vogal alta posterior, em posição átona final de palavras, em contexto seguinte à vogal anterior alta ou *glide* palatal. Os achados levam as autoras a sugerirem que o apagamento é o estágio final de um fenômeno de reorganização segmental, de natureza gradiente: análises acústicas permitem que as autoras identifiquem “vestígios” da vogal posterior alta átona final sobre a estrutura formântica da vogal anterior alta precedente, principalmente sobre a trajetória descendente de  $F_2$ , na porção final da vogal anterior alta. Como decorrência, embora o apagamento não seja audível, ele pode ser caracterizado a partir do detalhe fonético, a exemplo desse mencionado. Ao promover uma reorganização segmental, conforme sustentam as autoras, o apagamento tem impacto sobre a representação fonológica. A partir daí, as autoras argumentam que a Teoria de Exemplares acomoda de maneira adequada o fenômeno e buscam elucidar ao leitor como o apagamento gradiente pode ser tratado à luz desse modelo.

O texto da seção Debates, “Uma introdução à ciência aberta e ao compartilhamento de dados científicos de pesquisa”, de autoria de Karolayne Costa Rodrigues de Lima e Marcos Sunye, não trata

especificamente de Linguística, mas toca numa questão importantíssima, que concerne também à nossa área: a ciência aberta e a divulgação de dados de pesquisa, movimentos que vêm crescendo no exterior e que, aos poucos, são implementados no Brasil, como aliás tem sido a política das publicações da Associação Brasileira de Linguística. Resultado do evento “Ciência Aberta – como fica a Linguística”, realizado em 04 de setembro último<sup>1</sup>, no âmbito do programa de pós-graduação em Letras da UFPR, o texto busca situar os dados científicos no contexto da ciência aberta, apresentando e discutindo conceitos de dados, compartilhamento e reuso de dados científicos. O texto também conduz uma reflexão sobre os desafios impostos pela gestão dos dados científicos, bem como pela tipologia de dados e sobre as vantagens de compartilhamento e reuso de dados científicos. Os desafios, como argumentam os autores, não se resumem à gestão dos dados, mas abarcam, igualmente, as questões computacionais embutidas na construção e implantação de repositórios de dados de pesquisa, em razão da diversidade, volume e tipologia desses dados. Tais reflexões levam os autores a indicarem diretórios de pesquisa de repositórios de dados disciplinares e multidisciplinares. Cabe frisar que a discussão contida no texto pode orientar nossas reflexões sobre armazenamento dos dados que coletamos e produzimos em nossas pesquisas e se constitui num forte argumento sobre a necessidade cada vez mais premente de considerarmos aderir às práticas de ciência aberta.

À luz de discussões como essa, os editores da *Gradus* têm cogitado a possibilidade de tornarem não só abertos os artigos publicados na revista, mas também os pareceres exarados para cada um deles. Essa nos parece uma maneira de reconhecer a importância do trabalho dos pareceristas, já que cogitamos atribuir código DOI também aos pareceres. Esta é uma discussão que precisa ser amadurecida e que informaremos aos leitores conforme avance.

E, por falar em DOI, uma boa notícia: todos os artigos da *Gradus* contam, agora, com um código DOI. Isto vale para todos os artigos, desde o primeiro número do primeiro volume da revista. A implantação do DOI nos artigos possibilitará pleitearmos a indexação da *Gradus* em grandes plataformas de periódicos.

Nós finalizamos este editorial desejando aos caros leitores, autores, pareceristas e membros do conselho editorial boas festas! Que elas possam renovar nossas forças e nossas esperanças e que, em breve, possamos voltar a colher nossos dados e a conversar pessoalmente.

*Adelaide H. P. Silva*  
*Ubiratã Kickhöfel Alves*  
Editores

<sup>1</sup> O vídeo do evento está disponível em <https://www.youtube.com/watch?v=8CoaJVHI2ro&t=661s>

## *From the Editors*

We present Volume 5, issue 2 of *Gradus* – Brazilian Journal of Laboratory Phonology. In a challenging year such as 2020, in which we were prevented from having face-to-face activities – among which the collection of data for production and speech perception experiments – we consider the publication of this issue to be a victory, especially because we managed to maintain the quality standard that *Gradus* strives to pursue.

Although the authors of any submission to *Gradus* are not required to hold a Doctor's degree (since we understand that the quality of an article is not necessarily linked to the title of its author), it is worth highlighting that our reviewers do a careful job in ensuring that we publish excellent articles. This practice has led *Gradus* to consolidate its status of a reference journal in the field of Laboratory Phonology.

This said, the limitations imposed by the pandemic, as well as our policy of publishing good articles, have resulted in the publishing of only two texts, an article and a text in the 'Debates' section.

The article, entitled "*Redução vocálica: a gradiência na organização segmental*" ("Vowel reduction: gradience in segmental organization") and written by Cecília Toledo and Thaïs Cristófaró, describes the processes of reduction and deletion involved in the production of the high back vowel in word-final unstressed position, when this vowel follows a high front vowel or palatal glide. The findings show that deletion is the final stage of a segmental reorganization phenomenon which presents a gradient nature: acoustic analyses show "traces" of the word-final unstressed high back vowel in the formant structure of the preceding high front vowel, mainly in the downward  $F_2$  trajectory in the final portion of the high front vowel. As a result, although deletion is not audible, it can be characterized from phonetic detail. As stated by the authors, by promoting segmental reorganization, deletion has an impact on phonological representations. The authors argue that Exemplar Theory adequately accounts for the phenomenon, as it clarifies how the gradient process of deletion can be explained in light of this model.

The text in the 'Debates' section is entitled "*Uma introdução à ciência aberta e ao compartilhamento de dados científicos de pesquisa*" ("An introduction to open science and the sharing of scientific research data") and is authored by Karolayne Costa Rodrigues de Lima and Marcos Sunye. This text does not specifically deal with Linguistics, but addresses a very important issue that

also concerns our field: the pursuit of an open science and the dissemination of research data, which are movements that have been growing abroad and that have been implemented gradually in the Brazilian scenario, as in the publications by the Brazilian Association of Linguistics (Abralin). This text is a result of the event entitled “*Ciência Aberta – Como fica a Linguística?*” (“Open science – What about Linguistics?”), which took place in the Graduate Program of Linguistics at Universidade Federal do Paraná on September 4<sup>th</sup>, 2020<sup>1</sup>. The text aims to situate scientific data in the context of open science, presenting and discussing concepts related to the sharing and reuse of scientific data. The text also proposes a reflection on the challenges imposed by the management of scientific data, as well as on the type of data and on the advantages of sharing and reusing such data. As argued by the authors, the challenges are not limited to data management, but also cover the computational issues embedded in the construction and implementation of research data repositories, due to the diversity, volume and typology of these data. These reflections lead the authors to suggest research directories of disciplinary and multidisciplinary data repositories. It should be noted that the discussion contained in the text can guide our reflections on the storage of the data that we collect and produce in our research, being a strong argument in favor of adhering to open science practices.

In view of these discussions, the editorial board at *Gradus* has been considering the possibility of opening not only the articles published in the magazine, but also the reviews for each one of the articles. This seems to be a way of recognizing the importance of the work done by our reviewers, as we are considering assigning a DOI code to peer reviews as well. This is a possibility that needs consideration, and thus we will inform our readers of the conclusions we eventually reach.

Speaking of DOI, we have good news: all *Gradus* articles now have a DOI code. This applies to all articles, from the first issue of the first volume on. The implementation of DOI codes in the articles will allow us to apply for indexation in large indexing platforms.

Finally, we take this opportunity to wish a happy holiday to our dear readers, authors, reviewers and members of the editorial board. May these days renew our strengths and our hopes, so we can soon be able to both collect our data and talk in person.

*Adelaide H. P. Silva*  
*Ubiratã Kickhöfel Alves*  
Editors

<sup>1</sup> The video of the event is available at <https://www.youtube.com/watch?v=8CoaJVHI2ro&t=661s>.



## Editorial

Presentamos a los lectores el número 2 del volumen 5 de *Gradus*. En un año desafiante como el de 2020, que nos impidió tener diferentes actividades presenciales – entre ellas la recolección de datos para experimentos de producción y percepción del habla –, consideramos la publicación de este número una gran victoria, especialmente porque logramos mantener el estándar de calidad que *Gradus* se propone mantener.

En este sentido, cabe recordar que, si bien no exigimos que los autores que presentan su trabajo a *Gradus* sean doctores, pues entendemos que la calidad de un artículo no está necesariamente vinculada al título de su autor, nuestros evaluadores hacen un trabajo cuidadoso, para que la evaluación ciega por terceros, a los que se les envía cada artículo, asegure que tengamos textos de gran calidad. Esta estrategia ha llevado *Gradus* a consolidarse como una revista de referencia en el campo de la Fonología de Laboratorio.

Las limitaciones que nos impone la pandemia, sumadas a la política de publicación de buenos trabajos, han resultado en un volumen con solo dos textos, un artículo y un texto en la sección ‘Debates’.

El artículo, de Cecília Toledo y Thaïs Cristófar, “*Redução Vocálica: a gradiência na organização segmental*” (“Reducción vocálica: el gradiente en la organización segmental”), describe la reducción y la deleción en la producción de la vocal alta posterior, en una posición final átona de palabra, en un contexto posterior a una vocal anterior o a un *glide* palatal. Los hallazgos llevan a las autoras a sugerir que la deleción es la etapa final de un fenómeno de reorganización segmental, de naturaleza gradual: los análisis acústicos permiten a las autoras identificar “rastros” de la vocal posterior alta final átona en los formantes de la vocal precedente alta anterior, principalmente en la trayectoria descendente de  $F_2$ , en la porción final de la vocal precedente. Como resultado, aunque la deleción no es audible, dicho fenómeno puede ser caracterizado por el detalle fonético, como se mencionó anteriormente. Al promover una reorganización segmental, la deleción tiene un impacto en la representación fonológica. A partir de allí, las autoras argumentan que la Teoría de Ejemplares da cuenta adecuadamente del fenómeno, dado que explica cómo se puede tratar la deleción

gradiente.

El texto de la sección Debates, “*Uma introdução à ciência aberta e ao compartilhamento de dados científicos de pesquisa*” (“Una introducción a la ciencia abierta y al intercambio de datos de investigación científica”), de Karolayne Costa Rodrigues de Lima y Marcos Sunye, no aborda específicamente una cuestión lingüística, pero toca un tema muy importante, que también se discute en nuestra área: la ciencia abierta y la difusión de datos de investigación, movimientos que han crecido en el exterior y que se están implementando gradualmente en Brasil, como se ha visto en las publicaciones de la Asociación Brasileña de Lingüística (Abralín). Resultado del evento “*Ciência Aberta – Como fica a Linguística?*” (“Ciencia Abierta – ¿Cómo queda la Lingüística?”), realizado el 4 de septiembre, en el marco del programa de posgrado en Letras de la Universidade Federal do Paraná<sup>1</sup>, el texto busca situar los datos científicos en el contexto de la ciencia abierta, presentando y discutiendo cómo compartir y reutilizar datos científicos. El texto también realiza una reflexión sobre los desafíos que impone la gestión de datos científicos, así como sobre el tipo de datos y sobre las ventajas de compartir y reutilizar datos científicos. Los desafíos, como argumentan los autores, no se limitan a la gestión de datos, sino que también cubren los problemas computacionales integrados en la construcción e implementación de repositorios de datos de investigación, debido a la diversidad, el volumen y la tipología de estos datos. Tales reflexiones llevan a los autores a indicar directorios de investigación de repositorios de datos disciplinarios y multidisciplinares. Cabe señalar que la discusión contenida en el texto puede orientar nuestras reflexiones sobre el almacenamiento de datos que recopilamos y producimos en nuestras investigaciones, y expresa la necesidad cada vez más urgente de considerar la adhesión a prácticas de ciencia abierta.

A la luz de discusiones como la anterior, los editores de *Gradus* hemos estado considerando la posibilidad de abrir no solo los artículos publicados en la revista, sino también las evaluaciones de cada uno de ellos. Nos parece que esta es una forma de reconocer la importancia del trabajo de los pares evaluadores, ya que también estamos pensando en asignar un código DOI a las evaluaciones. Esta es una discusión que necesita de reflexión y que informaremos a los lectores a medida que experimente avances.

Y hablando de DOI, tenemos buenas noticias: todos los artículos de *Gradus* ahora tienen un código DOI. Esto se aplica a todos los artículos, desde el primer número del primer volumen de la revista. La implementación del DOI en los artículos nos permitirá solicitar la indexación en grandes plataformas de revistas científicas.

Terminamos este editorial deseando unas felices fiestas a nuestros queridos lectores, autores, evaluadores y miembros del conse-

<sup>1</sup> El video del evento está disponible en <https://www.youtube.com/watch?v=8CoaJVHI2ro&t=661s>.

jo editorial. Que estos días renueven nuestras fortalezas y nuestras esperanzas y que, pronto, podamos tanto recopilar nuestros datos como hablar en persona.

*Adelaide H. P. Silva*

*Ubiratã Kickhöfel Alves*

Editores



# Redução vocálica: a gradiência na reorganização segmental

<https://doi.org/10.47627/gradus.v5i2.143>

**Autores**

**Cecília Valle Souza Toledo**

<https://orcid.org/0000-0001-7997-3064>

Universidade Federal de Minas Gerais

**Thais Cristófaros Alves da  
Silva**

<https://orcid.org/0000-0002-9187-4154>

Universidade Federal de Minas Gerais

## Abstract

This paper aims at describing the phenomenon of non-stressed high back vowel reduction and loss in word-final position in Brazilian Portuguese, when it is preceded by a high front vowel or a palatal glide: *tio* [ˈtʃi.ʊ] ~ [ˈtʃiʷ] and *maio* [ˈmaʊ] ~ [ˈmaʊʷ]. We suggest that what is traditionally seen as elision or deletion should be understood as the final stage of segmental reorganisation, in which a segment that was formerly present no longer occurs. We also suggest that segmental reorganisation is gradient and has an impact to phonological representation. Exemplar Theory is the theoretical framework adopted in this study. We claim that, though the high back vowel is not heard in word-final position, it is characterized by phonetic detail after segmental reorganisation. Phonetic detail can be experimentally investigated, and in this paper this investigation was carried out through acoustic analysis. This paper provides evidence for the relevance of experimental studies in phonology.

**Keywords:** vowel reduction; segmental reorganisation; high vowels; gradience; phonetic detail.

## Resumo

Este artigo tem por objetivo descrever o fenômeno de redução e apagamento da vogal alta posterior átona, em final de palavras, quando precedida de vogal alta anterior ou glide palatal, como, por exemplo, em *tio* [ˈtʃi.ʊ] ~ [ˈtʃiʷ] e *maio* [ˈmaʊ] ~ [ˈmaʊʷ]. Sugerimos que o que é tradicionalmente denominado cancelamento, perda ou apagamento vocálico, deve ser compreendido como o estágio final de reorganização segmental, em que um segmento inicialmente presente deixa de ocorrer. Sugerimos também que a reorganização segmental é gradiente e tem impacto na representação fonológica. A Teoria de Exemplares é o modelo teórico adotado neste trabalho. Argumentamos que embora a vogal alta posterior não seja audível em final de palavras, ela é caracterizada de maneira específica a partir do detalhe fonético decorrente da reorganização segmental. O detalhe fonético pode ser experimentalmente investigado e, neste artigo, a investigação foi realizada por análise acústica. Este artigo oferece evidências para a relevância de estudos experimentais em fonologia.

**Palavras-chave:** redução vocálica; reorganização segmental; vogais altas; gradiência; detalhe fonético.

## Histórico do artigo

Recebido 2019-08-15

Revisado 2020-01-07

Aceito 2020-04-07

Publicado 2020-12-23

## Introdução

O enfraquecimento segmental decorrente do desvozeamento de vogais altas pode levar a um estágio em que a vogal pode não mais ocorrer.<sup>1</sup> A breve duração das vogais altas pode contribuir para que ocorra o desvozeamento, sobretudo, em contextos prosódicos átonos.<sup>2</sup> As vogais altas são as que apresentam menor duração dentre as demais vogais, e a breve duração destas vogais pode ser um fator que contribui para o enfraquecimento segmental.<sup>3</sup> Sugerimos que o que é tradicionalmente denominado cancelamento, perda ou apagamento vocálico deve ser compreendido como o estágio final de reorganização segmental, em que um segmento, inicialmente presente, deixa de ocorrer. Esta proposta será investigada neste artigo.

Modelos tradicionais da fonologia assumem que o cancelamento segmental é um fenômeno categórico: o segmento está presente ou ausente na representação.<sup>4</sup> Esta visão dicotômica e categórica em relação à presença ou ausência do segmento decorre, sobretudo, da natureza simbólica adotada nas representações fonéticas e fonológicas, geralmente baseadas em símbolos propostos pela Associação Internacional de Fonética, IPA.<sup>5</sup> Símbolos que representam segmentos sonoros têm limitações descritivas e podem restringir a compreensão dos fenômenos linguísticos decorrentes da dinâmica da fala.<sup>6</sup> Perspectivas atuais em fonologia buscam incorporar a dinâmica da fala na análise e sugerem que a perda segmental é gradiente e pode, eventualmente, ter impacto categórico.<sup>7</sup> Tomando-se a perspectiva atual, este artigo adota a Teoria de Exemplos, para investigar o fenômeno de redução e apagamento da vogal alta posterior átona, em final de palavras, quando precedida de vogal alta anterior ou glide palatal em dois estudos de caso.

<sup>1</sup> Vieira e Silva, “Redução vocálica em postônica final” (2015).

<sup>2</sup> de Souza, “Redução de vogais altas pretônicas no português de Belo Horizonte” (2012).

<sup>3</sup> Beckman, “When is a syllable not a syllable?” (1996).

<sup>4</sup> Lee, “Fonologia gerativa” (2017).

<sup>5</sup> Port, “How are words stored in memory? Beyond phones and phonemes” (2007).

<sup>6</sup> Munson et al., “Deconstructing phonetic transcription” (2010).

<sup>7</sup> Bybee, *Phonology and language use* (2001); Pierrehumbert, “Exemplar dynamics” (2001); Johnson, “Decisions and mechanisms in exemplar-based phonology” (2007).

Estudos de Caso		Exemplo	Representação
a. [i] tônico + [ʊ] átono final	i-tônico	<i>tio</i>	[ˈtʃi.ʊ] ~ [ˈtʃiʊ] ~ [ˈtʃiʊ]
b. vogal + [ɹ] glide + [ʊ] átono final	i-glide	<i>maio</i>	[ˈmaɪʊ] ~ [ˈmaɪʊ]

**Tabela 1:** Estudos de caso analisados.

Os dois estudos de caso listados na tab. 1 apresentam uma vogal alta anterior seguida de vogal alta posterior. A vogal alta anterior pode ser tônica [i], como em (a), ou pode ser um glide palatal [ɹ], como em (b). Faremos referência a casos como os ilustrados em (a) como *i-tônico* e aos casos em (b) como *i-glide*.

Casos do tipo *i-tônico* podem ser produzidos como hiato, como ditongo, e, ainda, a vogal [ʊ] átona final pode não ocorrer. Casos do tipo *i-glide* podem ser produzidos como ditongo ou a vogal [ʊ] átona final pode não ocorrer. A diferença entre os dois estudos de

caso diz respeito à tonicidade (vogal tônica ou átona) e à natureza vocálica (vogal ou glide). A semelhança entre os dois casos é que a vogal átona final [ʊ] pode não ocorrer. Faz-se pertinente investigar se o mesmo tipo de fenômeno opera nos dois casos, em relação ao cancelamento da vogal átona final [ʊ]. Tradicionalmente, ao se assumir que ocorre o apagamento da vogal [ʊ] átona final, espera-se que um fenômeno de mesma natureza ocorra nos exemplos na tab. 1.

Neste artigo, sugerimos que o que é tradicionalmente denominado cancelamento, perda ou apagamento vocálico deve ser compreendido como o estágio final de reorganização segmental em que um segmento, inicialmente presente, deixa de ocorrer. Investigaremos, experimentalmente, por meio de análise acústica, a natureza da redução segmental.

O fenômeno discutido neste artigo foi previamente considerado na literatura.<sup>8</sup> O presente trabalho avança ao realizar uma pesquisa experimental detalhada, ao investigar nomes e verbos (sendo que trabalhos precedentes consideraram apenas nomes) e ao discutir a natureza da perda segmental.

Este artigo tem a seguinte organização. A segunda seção apresenta a perspectiva teórica da Teoria de Exemplos, que é adotada neste trabalho. A terceira seção considera em detalhes o objeto de estudo e as categorias vocálicas que foram analisadas. A quarta seção considera os princípios metodológicos adotados. A quinta seção discute os resultados encontrados e é seguida da conclusão.

## *Perspectiva Teórica*

A Teoria de Exemplos sugere que a experiência linguística tem impacto nas representações mentais. A variação encontrada na fala – dialetal, social, individual – é, portanto, parte do conhecimento fonológico.<sup>9</sup> Para a Teoria de Exemplos, as variantes linguísticas produzidas e percebidas na fala formam feixes de exemplos na representação mental. O mapeamento dos exemplos se dá a partir da experiência do conhecimento linguístico do falante. Cada novo exemplar junta-se aos demais da mesma natureza. Exemplos são abstrações de palavras ou expressões que agrupam som e significado. As diversas produções da palavra *rio* – [hi̯], [hi.ʊ], [hi] –, por exemplo, constituem um feixe de exemplos que acumulam informações fonéticas, semânticas, perceptuais e sociais desta palavra. Dentre estas informações, no caso de *rio*, há a categorização do exemplar como nome ou verbo. Assim, o conhecimento fonológico é multirrepresentado, incluindo o detalhe fonético e toda e qualquer informação decorrente da experiência com a língua. Esta perspectiva exclui repre-

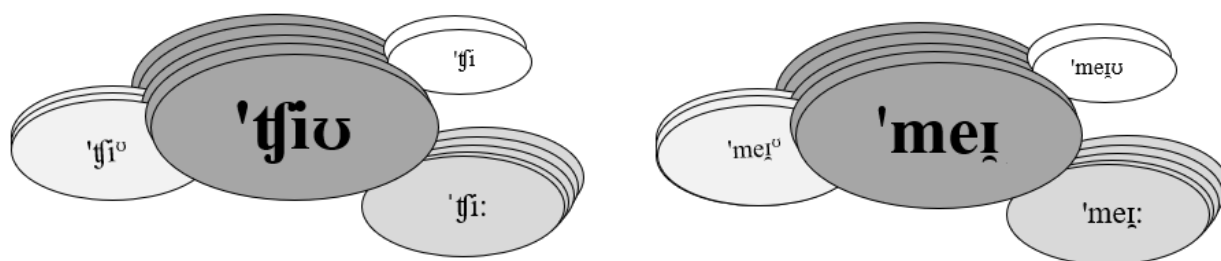
<sup>8</sup> Santiago, “Alteração segmental em sequência de vogais altas no português de Belo Horizonte” (2005); Mendonça, “Segmental reduction in sequences of high vowels” (2010).

<sup>9</sup> Bybee, *Phonology and language use* (2001); Cristófar-Silva e Gomes, “Teoria de exemplos” (2017).

sentações únicas e discretas, como proposto por teorias formais da fonologia.

Na Teoria de Exemplares a mudança sonora é entendida como fonética e lexicalmente gradual. À medida que pronúncias inovadoras ocorrem no uso linguístico, as categorias armazenadas para tais formas também mudarão gradualmente.<sup>10</sup> Nessa perspectiva, casos de redução vocálica, como os discutidos neste artigo, têm categorias múltiplas que armazenam tanto as formas reduzidas como as não reduzidas. De acordo com Bybee (2001), exemplares de formas reduzidas serão acumulados na representação fonológica à medida que formas reduzidas se tornarem frequentes na experiência linguística, de maneira que as representações são mais ou menos robustas. Considere a fig. 1.

<sup>10</sup> Bybee, *Phonology and language use* (2001).



**Figura 1:** Feixes de exemplares das palavras *tio* e *meio*.

O diagrama apresentado na fig. 1 ilustra feixes de exemplares que representam a categorização das palavras *tio* e *meio*. O feixe de exemplares, em ambas as palavras, armazena a gradiência fonética existente no percurso redutivo. Com finalidade ilustrativa, a robustez dos exemplares é indicada por cores diferentes e pelo tamanho dos diagramas na fig. 1: quanto mais robusto é o exemplar, mais escuro é o tom de cinza e maior é a sua dimensão. O exemplar mais robusto para a palavra *tio* apresenta um hiato [tʃiʊ]. Isso significa que a forma [tʃiʊ] é mais recorrente na experiência linguística ilustrada na fig. 1 do que as demais formas atestadas para a palavra *tio*: [tʃiʊ̃], [tʃi:] e [tʃi]. Por outro lado, o exemplar mais robusto para a palavra *meio* é [meĩ]. Ou seja, na palavra *meio* a pronúncia sem a vogal átona final é mais recorrente do que as demais pronúncias da palavra: [meiʊ̃]; [meĩʊ̃]; [mei:]. Obviamente, os diagramas da fig. 1 poderiam apresentar exemplares que caracterizassem a palatalização ou não da consoante inicial da palavra *tio*, ou diferenças de pronúncia por faixa etária ou sexo do falante. Contudo, as representações com várias informações são visualmente complexas e, por esta razão, a fig. 1 é restrita aos casos de redução da vogal átona final [ʊ].

Pesquisas que utilizam o aporte teórico da Teoria de Exemplares investem em metodologias experimentais, na busca de evidências

empíricas que reflitam a natureza das representações abstratas.<sup>11</sup> O método experimental é recorrentemente utilizado por fonólogos de laboratório. Fonólogos de laboratório podem discordar quanto à teoria fonológica, mas compartilham uma preocupação em aperfeiçoar a metodologia, a modelagem e o acervo de resultados que fundamentam empiricamente os estudos fonológicos.<sup>12</sup>

Os pressupostos da Fonologia de Laboratório orientaram a construção metodológica deste artigo. Utilizamos a análise acústica na busca por evidências empíricas sobre a natureza da redução da vogal [ʊ] átona final adjacente à vogal alta anterior. Na análise acústica, consideramos a duração e a frequência dos formantes das sequências de vogais. A próxima seção considera o método de categorização das sequências sonoras analisadas: ‘vogal alta anterior + vogal alta posterior’.

## *Categorias vocálicas analisadas*

Vogais altas em sequência são analisadas neste artigo: anteriores + posteriores. Vogais altas anteriores e posteriores compartilham a propriedade acústica de valores baixos de  $F1$ , que é característica de vogais altas. Vogais anteriores apresentam valores altos de  $F2$  – cerca de 2.000–2.500 Hz – e as vogais posteriores apresentam valores baixos de  $F2$  – cerca de 800–1000 Hz.<sup>13</sup> Considerando-se os valores de  $F1$  e de  $F2$  indicados anteriormente, esperamos que quando a sequência de ‘vogal alta anterior + vogal alta posterior’ for produzida, os valores de  $F2$  sejam inicialmente altos – que é característico da vogal alta anterior – e passem a ser baixos ao final da sequência segmental, uma vez que valores baixos de  $F2$  são característicos da vogal alta posterior. Esperamos, portanto, uma trajetória descendente de  $F2$ .

São três as categorias vocálicas a serem analisadas neste artigo:

1. ‘vogal alta anterior + vogal alta posterior’ presentes, [iʊ] ou [ɪʊ];
2. vogal alta anterior decorrente da redução vocálica, [i<sup>ʊ</sup>] ou [ɪ<sup>ʊ</sup>]; e
3. vogal alta anterior ou glide plenos em final de palavras, como [i] ou [ɪ].

Na primeira categoria, as duas vogais altas são audíveis e presentes no sinal acústico pela trajetória descendente de  $F2$ . Na segunda categoria, a vogal posterior no final da palavra, i.e., [ʊ], não é audível e a trajetória descendente de  $F2$  não é atestada no sinal acústico. Este caso reflete a redução vocálica, [i<sup>ʊ</sup>] ou [ɪ<sup>ʊ</sup>], e há estabilidade formântica da vogal final. Finalmente, na terceira categoria a vogal alta anterior ou glide plenos são audíveis e identificados no

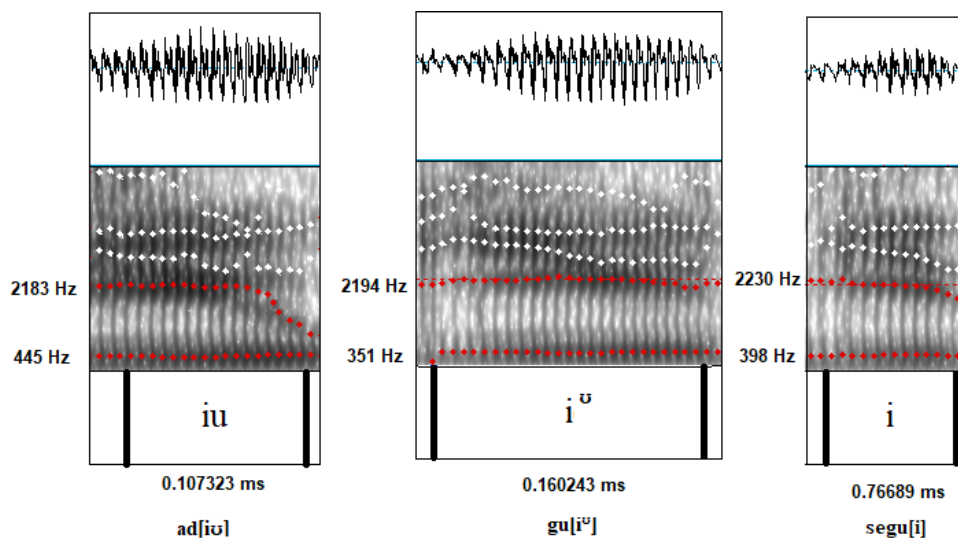
<sup>11</sup> Cristófar-Silva e Gomes, “Teoria de exemplares” (2017).

<sup>12</sup> Pierrehumbert et al., “Conceptual foundations of phonology as a laboratory science” (2000).

<sup>13</sup> Callou et al., “O sistema vocálico do português do Brasil” (2002); Barbosa e Madureira, *Manual de fonética acústica experimental* (2015); Cristófar-Silva et al., *Fonética acústica* (2019).



sinal acústico. Considere a fig. 2.

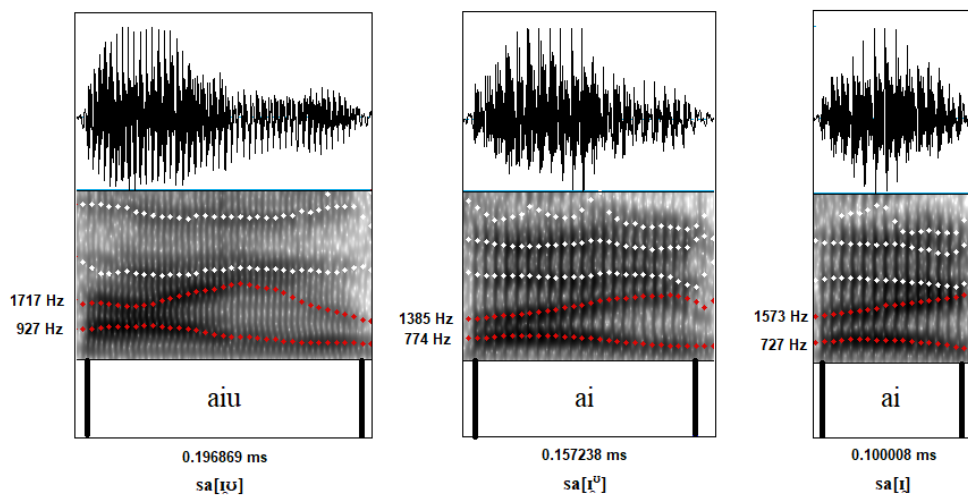


A fig. 2 ilustra três exemplos do estudo de caso *i-tônica*. O exemplo à esquerda ilustra um dado em que as vogais [i<sup>u</sup>] foram produzidas: *adio*. Dados em que ambas as vogais foram produzidas foram anotados como *iu*. A produção das duas vogais é, claramente, identificada pela trajetória descendente de  $F_2$  (linhas pontilhadas em vermelho).

O exemplo ao centro ilustra um dado em que a vogal final [u] não é audível e a trajetória descendente de  $F_2$  não é atestada no sinal acústico: *guio*. Estes casos foram anotados como *i<sup>u</sup>*. Observe que, no exemplo ao centro,  $F_2$  se mantém estável do início ao fim da vogal, o que indica acusticamente que, em princípio, apenas a vogal [i] foi produzida.

O exemplo à direita ilustra um dado produzido com a vogal anterior plena: *seguí*, que foi anotado por *i*. Neste caso, a trajetória formântica é estável, como seria esperado para monotongos. Características acústicas semelhantes são atestadas para o estudo de caso com *i-glide*. Considere a fig. 3.

Figura 2: Etiquetação de dados de *i-tônica*.



**Figura 3:** Etiquetação de dados de *i-glide*.

A fig. 3 ilustra três exemplos do estudo de caso *i-glide*, em que foram considerados os seguintes segmentos: ‘vogal + glide + [ʊ]’. O exemplo à esquerda ilustra um dado em que a sequência de vogais [V̄iʊ] foi produzida: *saiu*. Observa-se o declínio na trajetória de  $F_2$ , o que caracteriza a presença da vogal [ʊ] que é também audível. Estes casos foram anotados com *aiu*, *eiʊ* ou *oiʊ*.

O exemplo ao centro ilustra um dado em que a vogal [ʊ] foi reduzida: *saiu*. Os dados que apresentaram redução de [ʊ] foram anotados com *aiʊ*, *eiʊ* ou *oiʊ*. Observa-se um breve declínio dos valores de  $F_2$  ao final da palavra. Contudo, o declínio de  $F_2$  não expressa a trajetória descendente esperada na transição para a vogal [ʊ], pois é breve e não tem correlato audível. Esses casos foram anotados com *aiʊ*, *eiʊ* ou *oiʊ*.

Finalmente, o exemplo à direita ilustra um dado produzido com o glide [ɪ] pleno, em final de palavra, precedido da vogal [a]: *sai*. Os dados produzidos com o glide [ɪ] pleno foram anotados como *ai*, *ei* ou *oi*.

As categorias apresentadas nesta seção tiveram por objetivo investigar a redução e o apagamento da vogal alta posterior átona [ʊ], em final de palavras, quando precedida de vogal alta anterior ou glide palatal. Nos casos em que a vogal alta posterior ocorre em final de palavra, a vogal alta posterior é audível e tem correlato acústico da trajetória descendente de  $F_2$ : [iʊ] por exemplo, na palavra *tio*, e [V̄iʊ], por exemplo, na palavra *meio*. Nos casos em que a vogal alta posterior não ocorre em final de palavra, ela não é audível e não se observa a trajetória descendente de  $F_2$  esperada se houvesse a vogal [ʊ]. Nestes casos a ausência da vogal alta posterior final foi representada por [ʊ], como, por exemplo, em *macio* [ma'siʊ] e em *meio* [ˈmeɪʊ].

Consideramos também casos em que uma vogal alta anterior ou um glide palatal ocorre em final de palavras, como, por exemplo, *saci* [sa'si] e *lei* [ˈleɪ]. Neste grupo de palavras, a vogal ou glide final

foi considerada como tendo natureza plena. As características da vogal ou glide plenos foram contrastadas com casos em que uma vogal ou glide ocorre em final de palavra após a redução segmental de [ʊ] átono em final de palavra. Por exemplo, tanto em *rio* [ˈhiʊ] quanto em *ri* [ˈhi] ocorre, em final de palavra, uma vogal alta anterior, enquanto que em *caio* [ˈkaɪʊ] e *cai* [ˈkaɪ] ocorre em final de palavra um glide palatal. A nossa investigação pretende avaliar se a vogal ou glide plenos têm características segmentais análogas às da vogal ou glide que ocorre em final de palavra devido à ausência da vogal alta posterior final [ʊ]. Pretendemos avaliar os índices de presença ou ausência da vogal alta posterior final e seus correlatos físicos, por meio dos parâmetros acústicos de duração e estrutura formântica das vogais que são considerados na próxima seção.

## Metodologia

As palavras analisadas neste estudo foram selecionadas a partir de consulta ao *corpus* do Projeto Aspa<sup>14</sup> e ao Novo Dicionário Aurélio da Língua Portuguesa.<sup>15</sup> Um total de 48 palavras foi analisado, sendo 24 palavras para o estudo de caso *i-tônico* e 24 palavras para o estudo de caso *i-glide*. As palavras analisadas são apresentadas em ordem alfabética na tabela que segue:

<i>i-tônico</i>		<i>i-glide</i>	
adio	envio (nome)	apoio (nome)	leio
aprecio	envio (verbo)	apoio (verbo)	maio
arrepio (nome)	guio	boio	meio
arrepio (verbo)	início	caio	odeio
cio	macio	cheio	passeio (nome)
chio	navio	creio	passeio (verbo)
copio	plantio	desmaio (nome)	raio
desvio (nome)	rio (nome)	desmaio (verbo)	recreio
desvio (verbo)	rio (verbo)	ensaio (nome)	saio
doentio	tio	ensaio (verbo)	shoyo
elogio (nome)	trio	freio (nome)	sorteio (nome)
elogio (verbo)	vario	freio (verbo)	sorteio (verbo)

As palavras da tab. 2 estão apresentadas em ordem alfabética. Ao selecionarmos estas palavras, tínhamos em mente um outro grupo de palavras que terminassem em vogal alta anterior. Pretendíamos comparar o conteúdo segmental das palavras terminadas em ‘vogal alta anterior + [ʊ] átono final’ em que a vogal átona final [ʊ] não ocorre – como, por exemplo, em *rio* [hiʊ] ou *caio* [kaɪʊ] – com o conteúdo segmental das palavras terminadas em vogal alta

<sup>14</sup> <http://projetoaspa.org> (hoje desativado); acesso em abril de 2018. Cristóvão-Silva et al., “ASPA” (2005).

<sup>15</sup> Ferreira, *Novo dicionário Aurélio da língua portuguesa* (2009).

**Tabela 2:** Palavras terminadas em (vogal alta anterior + [ʊ] átono final).

anterior – como, por exemplo, em *ri* [hi], ou *cai* [kaɪ]. Um total de 37 palavras foi considerado no segundo grupo, sendo 20 palavras para o estudo de caso *i-tônico final* e 17 palavras para o estudo de caso *i-glide final*.<sup>16</sup> Considere a tab. 3.

<sup>16</sup> Um total de onze palavras foi repetido, sendo quatro nos casos de i-pleno e sete nos casos de glide-pleno.

<i>i-tônico final</i>		<i>i-glide final</i>	
adio	pedi	apoio (nome)	boi
aprecio	mereci	apoio (verbo)	constrói
arrepio (nome)	tupi	boio	dói
arrepio (verbo)	rompi	caio	cai
chio	mexi	cheio	rei
cio	saci	creio	lacrei
copio	cuspi	desmaio (nome)	samurai
desvio (nome)	davi	desmaio (verbo)	vai
desvio (verbo)	revi	ensaio (nome)	samurai
doentio	parati	ensaio (verbo)	sai
elogio (nome)	maragogi	freio (nome)	frei
elogio (verbo)	elegi	freio (verbo)	decifrei
envio (nome)	tupi	leio	falei
envio (verbo)	fervi	maio	pai
guio	segue	meio	lei
início	cresci	odeio	nadei
macio	saci	passeio (nome)	lei
navio	davi	passeio (verbo)	passei
plantio	piti	raio	pai
rio (nome)	mi	recreio	rei
rio (verbo)	ri	saio	sai
tio	piti	shoyo	boi
trio	tri	sorteio (nome)	rei
vario	feri	sorteio (verbo)	votei

**Tabela 3:** Palavras terminadas em (vogal alta anterior).

As primeira e terceira colunas da tab. 3 reproduzem as palavras listadas na tab. 3. As segunda e quarta colunas da tab. 3 apresentam as palavras com i-pleno ou glide-pleno em final de palavra. Essas palavras foram consideradas para a avaliação da natureza da vogal alta anterior e do glide palatal em final de palavra decorrente da redução vocálica.

Dois critérios guiaram a seleção de palavras com i-pleno ou glide-pleno: cadeia segmental semelhante e/ou número de sílabas equivalente. Por exemplo, para a palavra *macio* [ma'si<sup>o</sup>], foi selecionada a palavra *saci* [sa'si]. Palavras dissílabas, como *nadei* [na'deɪ], foram comparadas às palavras que se tornam dissílabas após a redução vocálica, como em *odeio* [o'deɪ<sup>o</sup>]. Note que tanto em *macio-saci* quanto em *odeio-nadei*, a consoante que precede a vogal ou ditongo a ser analisada é igual: [s] no primeiro par de palavras e [d] no segundo par de palavras. Ou seja, a consoante que precede a sequência segmental analisada foi controlada, de

maneira que fosse a mesma para cada par de palavras selecionado. Tal controle fez-se apropriado, uma vez que há evidências na literatura de que consoantes adjacentes e número de sílabas são fatores que podem interferir na duração e na qualidade de vogais.<sup>17</sup>

Foi então elaborado um experimento de leitura de sentenças. Embora a leitura de sentenças leve a um estilo de fala formal e menos sujeito à redução segmental, a vantagem foi a de se ter o controle dos dados. Assumimos que em estilo informal, o índice de redução segmental deverá ser superior ao encontrado no estilo de leitura formal. As sentenças foram elaboradas de maneira que as palavras a serem analisadas fossem seguidas de uma das oclusivas desvozeadas: [p, t, k], como, por exemplo, na frase “Ana tem um tio [p]esquisador”, elaborada para o estudo de caso *i-tônico*, e na frase “O ônibus cheio [t]entou correr”, elaborada para o estudo de caso *i-glide*. A opção por restringir o contexto seguinte a [p, t, k] visou à objetiva identificação do término da sequência de ‘vogal alta anterior + vogal alta posterior’ no sinal da fala, uma vez que as consoantes oclusivas apresentam ausência de energia como parâmetro acústico. Optamos por restringir o contexto seguinte a uma consoante desvozeada, pois a literatura reporta que a redução vocálica é favorecida quando a vogal se encontra adjacente a sons desvozeados.<sup>18</sup> Assim, o impacto do estilo formal de leitura adotado no experimento teria um contraponto favorável ao objeto de estudo, devido ao contexto favorável à redução vocálica.

Um total de 85 sentenças foi obtido para cada um dos 10 participantes.<sup>19</sup> As sentenças foram apresentadas aos participantes no programa POWER POINT, na tela de um computador. Cada sentença foi exibida por cinco segundos. O limite de tempo colocado na exibição das sentenças teve o objetivo de controlar a duração das gravações. Ao restringir o tempo de leitura de cada sentença, conseguimos que cada gravação coletada tivesse 11 minutos em média. Os slides com as sentenças foram aleatorizados a cada coleta de dados para cada participante. Assim, o efeito de ordem de apresentação dos dados foi excluído. Cinco participantes eram do sexo masculino e cinco participantes eram do sexo feminino. Todos os participantes são naturais da região metropolitana de Belo Horizonte, com idade entre 18 e 29 anos. A coleta de dados foi realizada em uma cabine com isolamento acústico no Laboratório CEFALA da Escola da Engenharia da UFMG. A gravação foi realizada em um gravador M-áudio micro-track II, configurado em uma taxa de amostragem de 44,1 KHz, a 16 bits.

As gravações foram organizadas para fins de análise acústica. O programa PRAAT foi utilizado para fins de etiquetagem e análise dos dados.<sup>20</sup> No primeiro grupo de palavras, foram identificados os segmentos vocálicos a serem analisados: vogais altas anteriores (tônica ou glide) e vogais altas posteriores em final de palavra (quando ocorreram). No segundo grupo de palavras, foram iden-

<sup>17</sup> Ferreira-Silva e Pacheco, “Características da duração do ruído das fricativas de uma amostra do português brasileiro” (2012).

<sup>18</sup> Meneses, “As vogais desvozeadas no português brasileiro” (2012); de Assis, “A emergência de consoantes finais no português brasileiro na microrregião de Araguaína/ Tocantins” (2017).

<sup>19</sup> Para cada um dos 10 participantes, foram formuladas sentenças com o seguinte número de palavras: 24 palavras com [iɔ] + 24 palavras [iɔ] + 24 palavras com i-pleno + 24 palavras glide-pleno – 11 palavras repetidas = 85 palavras.

<sup>20</sup> <https://www.fon.hum.uva.nl/praat/>.

tificadas as vogais altas anteriores plenas ou glide pleno no final da palavra. Foram analisados os seguintes parâmetros acústicos para as vogais: duração relativa e frequências dos dois primeiros formantes.

Um desafio imposto na análise dos dados foi quanto a caracterização formântica para se investigar a qualidade vocálica, considerando-se a variabilidade dos falantes. Para a análise articulatória, decidimos trabalhar com a diferença dos valores de  $F1$  e  $F2$ : ( $F2 - F1$ ), para normalizar os dados dos falantes. O processo de normalização envolve tipicamente uma transformação na escala dos formantes.<sup>21</sup> Suponhamos que, em uma situação hipotética, dois falantes com tubos de ressonâncias de tamanhos distintos produzam a vogal [i] com os seguintes valores de  $F1$  e  $F2$ : Falante 1 ( $F1 = 450$  Hz;  $F2 = 2100$  Hz). Falante 2 ( $F1 = 250$  Hz;  $F2 = 1900$  Hz). O Falante 1 possui frequências de ressonância maiores do que o Falante 2, mas ambos os falantes possuem diferença  $F2 - F1 = 1650$  Hz. A diferença de  $F2 - F1$  foi utilizada para caracterizar a qualidade da parte final da vogal, em casos de redução vocálica e em casos de vogais ou glides plenos. Medimos, portanto, os valores de  $F1$  e  $F2$  em 100 pontos temporais na porção final – 80 a 90 % – do intervalo segmentado (os intervalos segmentados estão expostos nas figs. 2 e 3). Posteriormente, calculamos a média dos 100 pontos obtidos. As medidas dos formantes foram extraídas do PRAAT com o auxílio de um *script*.<sup>22</sup>

Além das medidas de  $F2 - F1$ , consideramos a duração relativa das vogais. A duração relativa permite que efeitos que afetam a duração intrínseca dos segmentos sejam normalizados.<sup>23</sup> Por meio da duração relativa, obtivemos o espaço (percentual) ocupado pelas sequências ‘vogal alta anterior + vogal alta posterior’ em cada palavra. Neste trabalho, a duração relativa foi obtida pelo seguinte cálculo:

$$\text{Duração relativa} = \frac{\text{duração absoluta da sequência}}{\text{duração absoluta da palavra}} \times 100$$

As durações absolutas da sequência e da palavra foram calculadas por meio de um *script*.<sup>24</sup> O resultado do cálculo da duração relativa mostrou o espaço (percentual) ocupado pelas sequências [io], [V̥i̯] e [i̯] dentro de cada palavra.

A análise da duração relativa permitiu avaliar a reorganização temporal das sequências ‘vogal alta anterior + vogal alta posterior’ quando a vogal [i̯] está ou não presente. Sugerimos que a investigação da duração relativa e dos valores dos formantes permite caracterizar a reorganização segmental em que um segmento, inicialmente presente, deixa de ocorrer.

Os dados obtidos para a duração relativa e para a diferença entre  $F2 - F1$  foram exportados para um arquivo EXCEL. O arquivo

<sup>21</sup> Kent e Read, *Análise acústica da fala* (2015).

<sup>22</sup> Toledo, “Redução da vogal [u] adjacente à vogal alta anterior” (2019).

<sup>23</sup> Ferreira-Silva e Pacheco, “Características da duração do ruído das fricativas de uma amostra do português brasileiro” (2012).

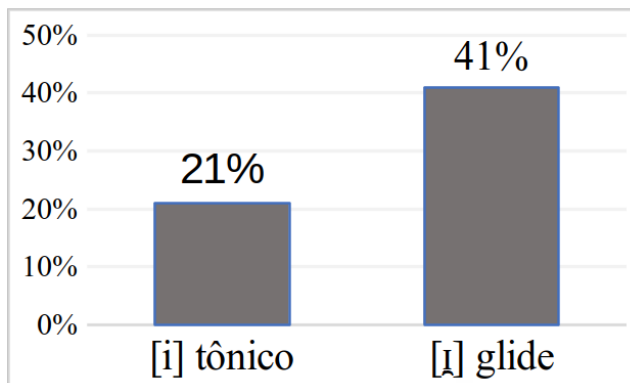
<sup>24</sup> Mietta Lennes, *Calculate segment durations*, [http://www.helsinki.fi/~lennes/praat-scripts/public/calculate\\_segment\\_durations.praat](http://www.helsinki.fi/~lennes/praat-scripts/public/calculate_segment_durations.praat).

EXCEL foi convertido para arquivo de texto .CSV, para ser reconhecido pela linguagem R.<sup>25</sup> Os resultados obtidos são apresentados na próxima seção.

<sup>25</sup> <https://www.r-project.org/>. O R é um ambiente projetado em torno de uma linguagem computacional que auxilia os pesquisadores na análise estatística e na plotagem de gráficos.

## Análise e discussão dos resultados

O total de dados esperado era de 850 (85 palavras x 10 participantes). Contudo, 5 dados foram descartados, pois a leitura da palavra-alvo na sentença foi equivocada. O *corpus*, após a exclusão dos cinco dados, contém 845 dados. A primeira análise teve por objetivo caracterizar os índices da redução vocálica da vogal átona final [ʊ]. Considere a fig. 4:



**Figura 4:** Índices de ausência da vogal átona final [ʊ].

A fig. 4 apresenta os índices de ausência da vogal final [ʊ] nos casos de *i-tônico* e *i-glide*. Os resultados indicaram que a redução e apagamento da vogal final [ʊ] ocorre tanto em nomes quanto em verbos, sem haver diferença estatisticamente significativa nos índices de ausência da vogal final [ʊ] entre as duas categorias. Os resultados indicam que o fenômeno é variável. Observa-se que, nos dados analisados, a ausência da vogal [ʊ] tem índices maiores, i.e., 41%, nos casos de *i-glide*, do que nos casos de *i-tônico*: 21%. Este resultado é relevante para compreendermos a evolução do fenômeno de redução e apagamento da vogal final [ʊ].<sup>26</sup> Sugerimos que a ausência da vogal [ʊ] ilustrada na fig. 4 reflete o estágio final da redução vocálica. Assim, a redução da vogal final [ʊ] se encontra em estágios mais avançados quando a vogal alta anterior é um glide [ɪ] (41%) do que quando a vogal alta anterior é tônica (21%). Portanto, a qualidade da vogal alta anterior, quando produzida como vogal ou glide, tem impacto na evolução da redução de [ʊ].

<sup>26</sup> Santiago, "Alteração segmental em sequência de vogais altas no português de Belo Horizonte" (2005); Mendonça, "Segmental reduction in sequences of high vowels" (2010).

Duas questões podem ser colocadas:

1. Por que a redução da vogal final [ʊ] é favorecida nos casos de *i-glide*?

2. Por que os índices de ausência da vogal [ʊ] diferem para *i-glide* e *i-tônico*?

Em relação à primeira questão, podemos sugerir que o contexto postônico no PB favorece o enfraquecimento segmental.<sup>27</sup> Uma vez que tanto o glide quanto a vogal [ʊ] se encontram em posição postônica nos casos de *i-glide*, seria a atonicidade que favoreceria o fenômeno de enfraquecimento segmental. Contudo, esta abordagem não explica por que é a vogal [ʊ], e não o glide, que é enfraquecida o suficiente para ser cancelada. Do ponto de vista articulatorio, as vogais são mais robustas do que os glides. Portanto, seria esperado o cancelamento do glide e não da vogal.

O glide é parte do ditongo, enquanto a vogal final [ʊ] reflete, nos nomes, a vogal temática que tipicamente marca o gênero masculino: *macio* x *macia*, ou a pessoa nos verbos: (*eu*) *caio*. Uma interpretação possível é de que a vogal [ʊ] pode ser omitida nos nomes para o masculino, porque o gênero da palavra seria expresso por artigos e também pela concordância nominal. Em relação aos verbos, a vogal final [ʊ] reflete a pessoa verbal e esta também poderia ser preservada em pronomes e pela concordância verbal. Esta explicação, pautada na morfologia, contudo, não explica por que em formas de plural a vogal [ʊ] não é omitida: *ensaios*, não \**ensaís*. Se a explicação fosse morfológica, não seria afetada pelo contexto segmental: final de palavra, mas não em casos de plural.

Possivelmente, o cancelamento da vogal [ʊ] pode ter relação com a assimetria observada entre sequências de vogais altas anteriores e posteriores no PB. Um dos aspectos desta assimetria é a variabilidade acentual em palavras como *gratuito*, *druída* ou *fluido*, que podem apresentar o acento tônico, tanto na vogal alta posterior, quanto na vogal alta anterior. Por outro lado, em palavras como *ciúme* ou *miúdo* não há variabilidade acentual em português: *ciúme* ou *míudo*. Casos que apresentam um ditongo formado por vogal alta anterior seguida de glide posterior no português são oriundos da vocalização da lateral, como em *Dilma* ou *Silva*. Ditongos formados por vogal alta anterior seguida de glide posterior ocorrem somente em posição final de palavras e em formas verbais: *caiu*, *sumiu*, *partiu*, etc.<sup>28</sup> Os fatores de assimetria entre sequências de vogais altas anteriores e posteriores no PB ainda precisam ser mais bem compreendidos. O fato é que o enfraquecimento segmental é favorecido em posição átona, em final de palavra e nos casos de *i-glide*.

Quanto à segunda questão formulada, devemos entender por que os índices de ausência da vogal [ʊ] diferem para *i-glide* e *i-tônico*. Caso o enfraquecimento da vogal [ʊ] seja interpretado como um fenômeno fonológico geral, seria esperado que esse se aplicasse sempre que o contexto vogal [ʊ] átona em final de palavra precedida de vogal alta anterior fosse encontrado. A noção de processo fonológico prevê que sempre que o contexto estrutural é encon-

<sup>27</sup> de Assis, "A emergência de consoantes finais no português brasileiro na microrregião de Araguaína/ Tocantins" (2017).

<sup>28</sup> O nome 'Piu piu' é uma exceção, pois apresenta um ditongo formado por vogal alta anterior seguida de glide posterior em final de palavra.



trado, o fenômeno se aplica e, portanto, seria esperado comportamento análogo em relação ao enfraquecimento e perda segmental em sequências segmentais similares. O argumento de que os índices diferentes de cancelamento da vogal [ʊ] para *i-glide* e *i-tônico* decorrem da natureza da vogal alta ser ou não um glide, ou ser tônica ou átona, tira o foco do elemento cancelado – vogal [ʊ] – para o elemento preservado: vogal alta anterior. Ademais, se o argumento para índices diferentes de cancelamento decorresse da vogal alta ser ou não um glide, a regra que expressaria o processo fonológico deveria ser desmembrada em relação ao contexto de aplicação: [ʊ] → ∅ / [i]\_\_ e também [ʊ] → ∅ / [ɪ]\_\_. Certamente, esta alternativa leva à perda da generalização atestada, que é: cancelamento de vogal [ʊ] átona em final de palavra quando precedida de vogal alta anterior, seja glide ou vogal. Esse fenômeno é tão generalizado que afeta até casos em que há vocalização da lateral palatal: *velho* [ˈvɛɪʊ] ~ [ˈvɛɪɪ].

Sugerimos que a diferença dos índices de cancelamento da vogal [ʊ] nos casos de *i-glide* e *i-tônico* possa ser melhor compreendida pela temporalidade da evolução do fenômeno. Possivelmente, o cancelamento da vogal [ʊ] átona em final de palavra teve início em casos como *i-glide*, em que as duas vogais altas são átonas, e por ter mais tempo de implementação, apresenta maiores índices do que os casos de *i-tônico*, que foram implementados posteriormente. Sugerimos ainda que a implementação do fenômeno de enfraquecimento e perda segmental seja gradualmente implementado, como previsto pela Teoria de Exemplares. Em seguida, investigaremos duas hipóteses relacionadas à implementação gradual da reorganização segmental: reorganização temporal e reorganização articulatória.

A primeira hipótese sugere que a redução da vogal [ʊ] final resulta em reorganização temporal da sequência de ‘vogal alta anterior + vogal alta posterior’. A reorganização temporal será avaliada pela investigação de valores da duração relativa. Espera-se maior duração relativa nos casos em que as duas vogais são produzidas: [iʊ]. Nos outros dois casos a serem analisados, apenas a vogal alta anterior foi audível. O primeiro caso agrupa exemplos em que a vogal [ʊ] final era inicialmente pronunciada e não mais ocorre, representados por [i<sup>ʊ</sup>]. O segundo caso agrupa exemplos de vogal ou glide plenos em final de palavra: [i] ou [ɪ]. Sugerimos que a duração relativa da vogal alta anterior nos casos em que a vogal ocorreu inicialmente, mas não é mais pronunciada, ou seja, [i<sup>ʊ</sup>] e [ɪ<sup>ʊ</sup>], será maior do que a duração relativa da vogal ou glide plenos. A relação esperada da duração relativa é: [iʊ] > [i<sup>ʊ</sup>] > [i]. Sugerimos que a maior duração da vogal alta anterior nos casos em que [ʊ] é ausente – como em [maˈsi<sup>ʊ</sup>] *macio* – expressa o detalhe fonético que caracteriza o fenômeno de redução segmental. Ou seja, há reorganização temporal da sequência ‘vogal alta anterior + vogal alta posterior’ que reflete a perda segmental.

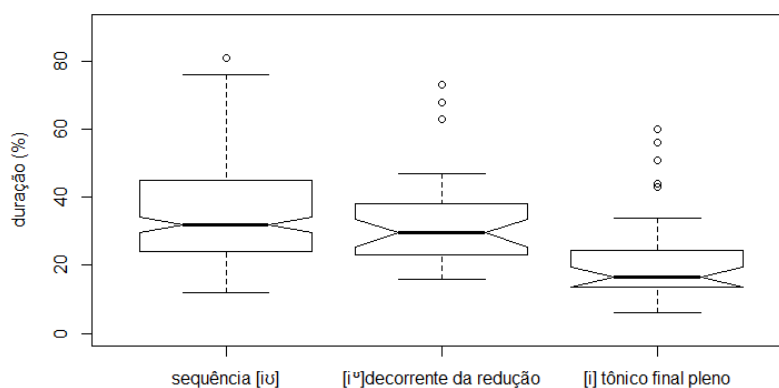
A segunda hipótese a ser testada é de que a reorganização articulatória afeta a qualidade da vogal alta anterior adjacente ao [ʊ]. Casos de vogal plena, como em [sa'si] *saci*, apresentam valores de formantes típicos de uma vogal alta anterior. Por outro lado, em casos em que a vogal final [ʊ] é cancelada, como [ma'si<sup>ʊ</sup>] *macio*, os valores dos formantes da vogal alta anterior deverão ser estatisticamente diferentes de vogais altas anteriores plenas. Ou seja, há reorganização articulatória da vogal alta anterior que reflete a perda segmental nos casos em que a vogal posterior [ʊ] deixa de ocorrer no final de palavras.

As duas hipóteses sugerem que o percurso de redução vocálica é gradiente e envolve reorganização temporal e articulatória da cadeia segmental como previsto pela Teoria de Exemplos.<sup>29</sup> Cada hipótese será analisada nas próximas páginas, visando a compreender a implementação gradual da reorganização segmental: reorganização temporal e reorganização articulatória.

Com a finalidade de testar a primeira hipótese, consideramos os valores da duração relativa em três casos:

1. as duas vogais foram produzidas, i.e., [iʊ], como em [ma'siʊ] *macio*;
2. a vogal átona final foi cancelada, i.e., [i<sup>ʊ</sup>], como em [ma'si<sup>ʊ</sup>] *macio* e
3. vogais plenas, como em [sa'si] *saci*.

A hipótese que formulamos sugere que a duração relativa será estatisticamente diferente em cada um dos três casos analisados. Considere a fig. 5.



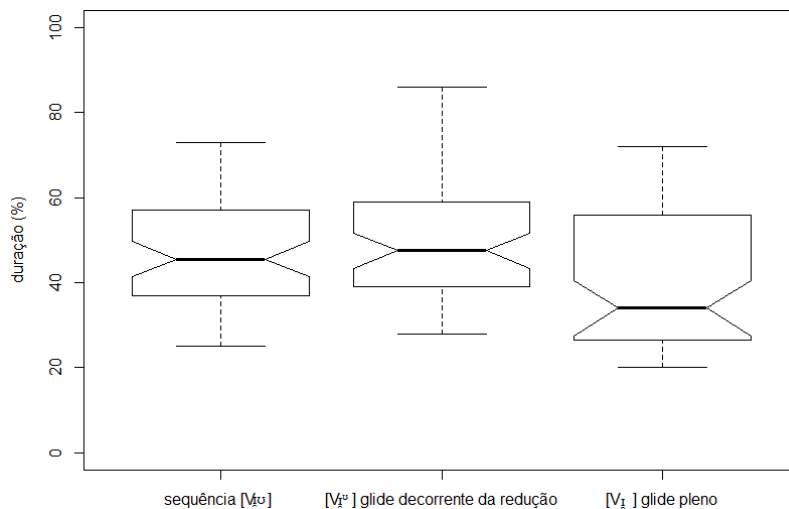
**Figura 5:** Duração relativa de [iʊ], de [i<sup>ʊ</sup>] e de [i] tônico final pleno.

A fig. 5 reflete resultados da duração relativa de casos de *i-tônico*. O *boxplot* à esquerda caracteriza a duração relativa de casos em que a vogal final [ʊ] foi produzida. Neste caso, duas vogais foram articuladas: [iʊ]. O *boxplot* central exibe a duração relativa de casos em que a vogal [ʊ] não foi produzida. Neste caso, uma única

<sup>29</sup> Bybee, *Phonology and language use* (2001); Pierrehumbert, "Exemplar dynamics" (2001); Foulkes e Docherty, "The social life of phonetics and phonology" (2006); Johnson, "Decisions and mechanisms in exemplar-based phonology" (2007).

vogal foi articulada: [i<sup>u</sup>]. Finalmente, o *boxplot* da direita apresenta a duração relativa da vogal [i] tônica plena.

A hipótese formulada sugeriu a seguinte relação esperada para a duração relativa nos dados que foram considerados: [i<sup>u</sup>] > [i<sup>u</sup>] > [i]. Os resultados obtidos indicaram que a duração relativa da sequência [i<sup>u</sup>] não foi estatisticamente diferente da duração relativa da vogal [i<sup>u</sup>] decorrente da redução ( $W=3.103$ ,  $p=0,3048$ ). Este resultado sugere que tanto a produção das duas vogais [i<sup>u</sup>] quanto a produção da vogal reduzida [i<sup>u</sup>] sejam englobadas em uma única categoria. Ao investigar os valores de duração relativa entre [i<sup>u</sup>] e [i], a análise estatística indicou que a duração relativa da vogal reduzida [i<sup>u</sup>] é significativamente maior do que o [i] tônico final pleno ( $W=807.5$ ,  $p=7.366e-05$ ). Este resultado corrobora a proposta de que a vogal [i<sup>u</sup>] é significativamente maior do que a vogal [i] plena. Argumentamos que a maior duração da vogal [i<sup>u</sup>] expressa o detalhe fonético que caracteriza o fenômeno de redução segmental, ou seja, há reorganização temporal da sequência ‘vogal alta anterior + vogal alta posterior’, que reflete a perda segmental do [u] átono final. Podemos sugerir a seguinte relação para a duração relativa nos casos de *i-tônico*: [i<sup>u</sup>] ~ [i<sup>u</sup>] > [i] e também [i<sup>u</sup>] > [i]. Considere a fig. 6.



**Figura 6:** Duração relativa das sequências [V̄īu], [V̄īu] e de [V̄ī] glide final pleno.

A fig. 6 reflete resultados da duração relativa de casos de *i-glide*. Neste caso, a duração relativa foi medida para a vogal e o glide correspondente, bem como para a vogal [u] final quando esta ocorreu. Foi necessário agregar a vogal que precede o glide nas medidas de duração relativa, uma vez que a transição formântica em ditongos dificultaria a segmentação somente do glide. O *boxplot* à esquerda caracteriza a duração relativa de casos em que a vogal final [V̄īu] foi produzida. O *boxplot* central exibe a duração relativa de casos em que a vogal final [u] não foi produzida: [V̄īu]. Finalmente, o *boxplot* da direita apresenta a duração relativa da

vogal [V<sub>1</sub>] glide final plena.

De maneira análoga aos resultados da fig. 5, os resultados expressos na fig. 6 mostram que a diferença entre a duração das sequências [V<sub>1</sub>ɔ] e [V<sub>1</sub><sup>ɔ</sup>] não é estatisticamente significativa, pois  $p > 0,05$  ( $W=1809,5$ ,  $p=0,4829$ ). Por outro lado, os dados evidenciam que há diferença significativa entre a duração da sequência [V<sub>1</sub><sup>ɔ</sup>] decorrente da redução vocálica e a duração da sequência [V<sub>1</sub>] com glide pleno ( $W=1950$ ,  $p=0,004258$ ). A duração relativa nos casos de redução vocálica, [V<sub>1</sub><sup>ɔ</sup>], é significativamente maior do que a duração relativa nos casos em que ocorre um glide pleno: [V<sub>1</sub>].

Os resultados apresentados nas figs. 5 e 6 confirmam a hipótese de reorganização temporal decorrente da perda segmental do [ɔ] átono final em sequências de ‘vogal alta anterior + vogal alta posterior’. Sugerimos que este resultado indica a natureza gradiente da perda segmental, como previsto pela Teoria de Exemplares. A maior duração relativa nos casos decorrentes da redução vocálica quando comparados com as vogais ou glides plenos é um correlato físico que oferece evidências sobre a perda segmental gradiente.

A seguir, apresentamos os resultados obtidos na investigação da segunda hipótese: a reorganização articulatória afeta a qualidade da vogal alta anterior adjacente ao [ɔ]. A hipótese de reorganização articulatória também é motivada pela premissa da Teoria de Exemplares, que sugere que a redução vocálica seja foneticamente gradiente. Os dados considerados foram obtidos pela diferença  $F2 - F1$  extraída da parte final das vogais em questão. Comparamos os valores de  $F2 - F1$  da vogal alta anterior decorrente da redução de [ɔ] com os valores de  $F2 - F1$  da vogal alta anterior plena. A predição é de quanto mais alto é o valor de  $F2 - F1$ , mais anterior é a vogal. Espera-se diferença significativa dos valores de  $F2 - F1$  nos casos em que ocorreu o cancelamento da vogal átona final [ɔ] e nos casos de vogal alta anterior plena. A diferença esperada é de que os casos de cancelamento da vogal átona final [ɔ] apresentem valores mais baixos de  $F2 - F1$  do que casos com vogal plena. Os valores baixos de  $F2 - F1$  indicariam que a vogal [ɔ], que não é mais produzida, altera a qualidade da vogal alta anterior adjacente. Assim, mesmo após a perda segmental, encontraríamos evidências do detalhe fonético da [ɔ] nos valores de  $F2 - F1$  da vogal [i]. Considere as figs. 7 e 8.



mais anteriores do que as vogais altas anteriores decorrentes da redução. Isso porque os casos de redução vocálica – [i<sup>v</sup>] ou [V̥<sup>v</sup>] – apresentaram valores menores de  $F2 - F1$  do que de vogais e glides plenos, o que confirma a hipótese formulada.

Os resultados apresentados nas figs. 7 e 8 oferecem evidências de que o percurso de cancelamento da vogal [ɔ] átona final é gradiente, e não categórico. Evidência para a natureza gradiente da redução da vogal [ɔ] átona final decorre dos valores de  $F2 - F1$  que caracterizam o detalhe fonético de uma vogal posterior previamente existente nos casos em que houve a redução vocálica. A perda segmental é compreendida como sendo gradiente e caracterizada por parâmetros de descrição acústica.

## Conclusão

Este artigo teve o objetivo de descrever o fenômeno de redução e apagamento da vogal alta posterior átona, em final de palavras, quando precedida de vogal alta anterior ou glide palatal. Por exemplo, em *tio* ['tʃi.ɔ] ~ ['tʃi<sup>v</sup>] e *maio* ['maɪɔ] ~ ['maɪ<sup>v</sup>]. Sugerimos que o que é tradicionalmente denominado cancelamento, perda ou apagamento vocálico, é melhor compreendido como o estágio final de reorganização segmental em que um segmento, inicialmente presente, deixa de ocorrer. Buscamos também demonstrar que a reorganização segmental, decorrente da redução segmental, é gradiente. O presente trabalho avançou em relação a estudos precedentes ao realizar pesquisa experimental detalhada na investigação da redução vocálica em sequências de vogais altas distintas e, adicionalmente, por considerar nomes e verbos.

Os resultados obtidos indicam que a qualidade da vogal alta anterior tem impacto na evolução da redução de [ɔ]. Palavras terminadas em [ɪ] glide + [ɔ] apresentaram maiores índices de redução de [ɔ] do que palavras terminadas em [i] tônico + [ɔ]. Os resultados mostraram também que há reorganização temporal e articulatória da vogal alta anterior precedente ao [ɔ] reduzido. A reorganização temporal é expressa pela maior duração da vogal [i] e do glide [ɪ] quando comparadas com a vogal alta anterior ou glide palatal plenos. A reorganização articulatória mostrou que há tendência à centralização da vogal alta anterior. Os valores de  $F2 - F1$  da vogal [i] e do glide [ɪ], de palavras como *rio* ['hi<sup>v</sup>] e *saio* ['saɪ<sup>v</sup>], se mostraram menores do que os valores de  $F2 - F1$  da vogal [i] e do glide [ɪ] pleno – *ri* e *sai*. Sugerimos que a diferença encontrada na duração e na trajetória dos formantes das vogais altas anteriores reflita o detalhe fonético da vogal [ɔ] não mais produzida.

Argumentamos também que o fenômeno de enfraquecimento e perda segmental é gradualmente implementado, tanto em casos *i-*

*tônico* como em casos *i-glide*, e que há diferença de temporalidade na implementação: o cancelamento da vogal [ʊ] átona em final de palavra teve início em casos como *i-glide* e, por essa razão, sequências '[ɪ] glide + [ʊ]' se encontram em um estágio mais avançado no percurso evolutivo do fenômeno de redução do [ʊ] átono em final de palavras.

Esperamos que o presente artigo incentive a formulação de estudos sobre a redução segmental em sequências vocálicas. Trabalhos futuros podem também investigar a assimetria entre sequências de vogais altas anteriores e posteriores no PB. A compreensão dessa assimetria pode contribuir com a reflexão sobre o favorecimento da redução de [ʊ] em relação ao [i]. Ademais, trabalhos futuros podem investigar se há labialização da vogal [i] precedente ao [ʊ] reduzido, por meio da análise do F3 da vogal alta anterior. Finalmente, esperamos que este trabalho contribua com o debate sobre fenômenos de variação e mudança linguística e com o avanço de pesquisas no âmbito da Fonologia de Laboratório e da Teoria de Exemplos.

## Referências

- de Assis, Alessandra Mara (2017). "A emergência de consoantes finais no português brasileiro na microrregião de Araguaína/ Tocantins". Tese de doutorado. Universidade Federal de Minas Gerais.
- Barbosa, Plínio Almeida e Sandra Madureira (2015). *Manual de fonética acústica experimental: aplicações a dados do português*. São Paulo: Cortez. ISBN: 9788524924217.
- Beckman, Mary E. (1996). "When is a syllable not a syllable?" In: *Phonological structure and language processing: cross-linguistic studies*. Editado por Takashi Otake e Anne Cutler. Berlin: De Gruyter, pp. 95–123. ISBN: 9783110815825.
- Bybee, Joan L. (2001). *Phonology and language use*. Cambridge: Cambridge University Press. ISBN: 9780521533782.
- Callou, Dinah, Yonne Leite e João Moraes (2002). "O sistema vocálico do português do Brasil: caracterização acústica". In: *Gramática do português falado: vol. V – convergências*. Editado por Mary Aizawa Kato. Campinas: Unicamp, pp. 33–52. ISBN: 9788526806139.
- Cristófaros-Silva, Thaís, Leonardo S. de Almeida e Thiago Fraga (2005). "ASPA: a formulação de um banco de dados de referência da estrutura sonora do português contemporâneo". *Proceedings of the XXV Congress of Brazilian Society of Computing Science*, pp. 2268–2277.
- Cristófaros-Silva, Thaís e Christina Gomes (2017). "Teoria de exemplos". In: *Fonologia, fonologias: uma introdução*. Editado por Dermeval da Hora e Carmen Lúcia Matzenauer. São Paulo: Contexto, pp. 157–168. ISBN: 9788572449830.
- Cristófaros-Silva, Thaís, Izabel Christine Seara, Adelaide P. Silva, Andreia Schurt Rauber e Maria Cantoni (2019). *Fonética acústica: os sons do português brasileiro*. São Paulo: Contexto. ISBN: 978852000792.
- Ferreira, Aurélio Buarque de Holanda, editor (2009). *Novo dicionário Aurélio da língua portuguesa*. São Paulo: Fundação Dorina Nowill para cegos.
- Ferreira-Silva, Audinéia e Vera Pacheco (2012). "Características da duração do ruído das fricativas de uma amostra do português brasileiro". *Estudos da Língua(gem)* 10.1, pp. 9–28. DOI: 10.22481/e1.v10i1.1167.

- Foulkes, Paul e Gerard Docherty (2006). "The social life of phonetics and phonology". *Journal of Phonetics* 34.4, pp. 409–438. DOI: 10.1016/j.wocn.2005.08.002.
- da Hora, Dermeval e Carmen Lúcia Matzenauer, editores (2017). *Fonologia, fonologias: uma introdução*. São Paulo: Contexto. ISBN: 9788572449830.
- Johnson, Keith (2007). "Decisions and mechanisms in exemplar-based phonology". In: *Experimental approaches to phonology*. Editado por Maria Josep Solé, Patrice Speeter Beddor e Manjari Ohala. Oxford: Oxford University Press, pp. 25–40. ISBN: 9780199296675.
- Kent, Ray D. e Charles Read (2015). *Análise acústica da fala*. Traduzido por Alexsandro Meireles. São Paulo: Cortez Editora. ISBN: 9788524923852.
- Lee, Seung Hwa (2017). "Fonologia gerativa". In: *Fonologia, fonologias: uma introdução*. Editado por Dermeval da Hora e Carmen Lúcia Matzenauer. São Paulo: Contexto, pp. 31–45. ISBN: 9788572449830.
- Mendonça, Gustavo (2010). "Segmental reduction in sequences of high vowels". São Paulo School of Advanced Studies in Speech Dynamic. URL: <http://200.144.254.127:8080/speechdynamics.pdf>.
- Meneses, Francisco (2012). "As vogais desvozeadas no português brasileiro: investigação acústico-articulatória". Dissertação de mestrado. Campinas: Universidade Estadual de Campinas.
- Munson, Benjamin, Jan Edwards, Sarah K. Schellinger, Mary E. Beckman e Marie K. Meyer (2010). "Deconstructing phonetic transcription: covert contrast, perceptual bias, and an extraterrestrial view of Vox Humana". *Clinical Linguistics & Phonetics* 24.4–5, pp. 245–260. DOI: 10.3109/02699200903532524.
- Pierrehumbert, Janet B. (2001). "Exemplar dynamics: word frequency, lenition and contrast". In: *Frequency and the emergence of linguistic structure*. Editado por Joan L. Bybee e Paul J. Hopper. Amsterdam: John Benjamins, pp. 137–158.
- Pierrehumbert, Janet B., Mary E. Beckman e D. Robert Ladd (2000). "Conceptual foundations of phonology as a laboratory science". In: *Phonological knowledge: Conceptual and empirical issues*. Editado por Noel Burton-Roberts, Philip Carr e Gerard Docherty. Oxford: Oxford University Press, pp. 273–304. ISBN: 9780199245772.
- Port, Robert (2007). "How are words stored in memory? Beyond phones and phonemes". *New Ideas in Psychology* 25.2, pp. 143–170. DOI: 10.1016/j.newideapsych.2007.02.001.
- Santiago, Eliana Castelli (2005). "Alteração segmental em sequência de vogais altas no português de Belo Horizonte". Dissertação de mestrado. Belo Horizonte: Universidade Federal de Minas Gerais.
- de Souza, Ricardo Fernandes Napoleão (2012). "Redução de vogais altas pretônicas no português de Belo Horizonte: uma abordagem baseada na gradiência". Dissertação de mestrado. Belo Horizonte: Universidade Federal de Minas Gerais. URL: <https://repositorio.ufmg.br/handle/1843/LETR-96LR2F>.
- Toledo, Cecília Valle Souza (2019). "Redução da vogal [u] adjacente à vogal alta anterior: uma investigação sobre a implementação da redução vocálica". Tese de mestrado. Belo Horizonte: Universidade Federal de Minas Gerais. URL: <http://hdl.handle.net/1843/LETR-BAFH56>.
- Vieira, Maria José Blaskovski e Thaís Cristófaros Silva (2015). "Redução vocálica em postônica final". *Revista da ABRALIN*, pp. 379–406. DOI: 10.5380/rabl.v14i1.42822.





## DEBATES:

# Uma introdução à ciência aberta e ao compartilhamento de dados científicos de pesquisa

<https://doi.org/10.47627/gradus.v5i2.165>

## Abstract

Open Science is defined as an ecosystem that houses the various movements around open access to scientific information. Scientific data are the basic nuclear unit that supports scientific discoveries. In this sense, the purpose of this paper is to situate scientific research data in the context of Open Science, presenting the definitions of data, sharing and reuse of scientific data. It presents an overview of the challenges regarding the management of scientific data, the type of data and the advantages of sharing and reuse. It presents the computational challenges to the construction of research data repositories. We conclude that diversity, typology and volume of data represent the major challenges in the implementation of repositories and indicate a research directory of disciplinary and multidisciplinary data repositories.

**Keywords:** open science; research data; data sharing; data reuse; data repository.

## Resumo

A Ciência Aberta é definida como um ecossistema que abriga os diversos movimentos em torno do acesso aberto à informação científica. Os dados científicos são a unidade nuclear básica que suporta as descobertas científicas. Nesse sentido, o objetivo deste trabalho é situar os dados científicos de pesquisa no contexto da Ciência Aberta, apresentando as definições de dados, compartilhamento e reuso de dados científicos. O trabalho apresenta um apanhado sobre os desafios quanto à gestão dos dados científicos, à tipologia de dados e às vantagens do compartilhamento e reuso. Apresenta os desafios computacionais para a construção de repositórios de dados de pesquisa. Conclui-se que a diversidade, a tipologia e o volume de dados representam os grandes desafios na implantação dos repositórios e indicam-se diretórios de pesquisa de repositórios de dados disciplinares e multidisciplinares.

**Palavras-chave:** ciência aberta; dados científicos de pesquisa; compartilhamento de dados; reuso de dados; repositório de dados.

## Autores

**Karolayne Costa Rodrigues de Lima**

<https://orcid.org/0000-0002-6311-8482>

Universidade Federal do Paraná

**Marcos Sfair Sunye**

<https://orcid.org/0000-0002-2568-5697>

Universidade Federal do Paraná

## Histórico do artigo

Recebido 2020-12-01

Aceito 2020-12-07

Publicado 2020-12-23

## Introdução

A Ciência Aberta é um ecossistema<sup>1</sup> que reúne as diversas iniciativas, planos e ações em torno da abertura das atividades, processos, métodos, avaliações e produções científicas. O foco da ciência aberta é a abertura de todas as etapas do ciclo de vida da atividade científica e apresenta temas ou macrossistemas que regem as políticas de acesso aberto, metodologias ativas e abertas, avaliação aberta, ferramentas abertas e dados abertos. Cada um desses grandes temas expandem-se em subsistemas vascularizados que particularizam determinada prática da atividade científica, formando, assim, uma interligada rede que tal como um ecossistema biológico interage entre si e com outros.

A Ciência Aberta, de forma ampla, é conceituada como um princípio por meio do qual “o conhecimento científico deve ser livre para as pessoas usarem, reutilizarem e distribuírem sem restrições legais”;<sup>2</sup> tornando o conhecimento acessível para a sociedade, de modo a fornecer subsídios para o desenvolvimento de uma ciência cidadã onde os indivíduos possam se apropriar do conhecimento gerado pela academia e/ou instituições e centros de pesquisa para modificar sua realidade e retroalimentar a ciência.

As práticas da ciência aberta têm por característica o uso massivo da tecnologia para conectar e inter-relacionar as etapas do ciclo de vida da atividade científica devido à disponibilização de informações e dados em todas as etapas da pesquisa científica, partindo do seu planejamento na etapa inicial com o delineamento de um Plano de Gestão de Dados (PGD) até o reuso de dados científicos de pesquisa.

A Ciência Aberta não é um tema novo, visto que sua abordagem está intimamente relacionada com a metodologia da pesquisa. Entretanto, o termo “ciência aberta” ou “*Open Science*” tornou-se comum a partir da discussão da mudança de paradigma científico do modelo computacional da última década, focado na simulação de modelos complexos, para a chamada *e-Science*,<sup>3</sup> baseada na exploração de dados com uso massivo de tecnologia e internet. Os dados explorados neste novo paradigma são aqueles gerados pela atividade científica, que dão embasamento aos resultados das pesquisas.

Nesse sentido, o objetivo geral desta comunicação é conceituar os dados científicos de pesquisa no âmbito da ciência aberta, detalhando as boas práticas de disponibilização e reuso de dados em diversas áreas do conhecimento.

<sup>1</sup> Albagli et al., *Ciência aberta, questões abertas* (2015).

<sup>2</sup> <http://www.isitopendata.org/>.

<sup>3</sup> Hey et al., *The fourth paradigm: data-intensive scientific discovery* (2009).

## Contexto da ciência aberta

No âmbito internacional, as primeiras iniciativas da ciência aberta surgiram em meados de 1980 nos Estados Unidos, com os movimentos do *software* livre e do acesso aberto à informação científica. Desses movimentos, destacam-se o Manifesto GNU em 1985, a licença tipo *Copyleft* e a *General Public Licence* (GPL), com as discussões em torno da abertura de licenciamento para sistemas *open source*.<sup>4</sup> Em relação ao acesso aberto, destacam-se a Conferência de Santa Fé em 1999, no estabelecimento de padrões para documentos eletrônicos, e os princípios da *Open Archives Initiative* (OAI); a Declaração de Budapeste em 2002, com a definição de estratégias para o acesso aberto e vias para atingi-lo; a Declaração de Bethesda em 2003, expandindo orientações e recomendações sobre acesso aberto para entidades acadêmicas e científicas; e a Declaração de Berlin, também em 2003, reforçando o compromisso com o acesso livre, pela internet, da informação e do conhecimento científico.<sup>5</sup>

Esses movimentos e declarações foram manifestações a favor da ruptura ao modelo tradicional de comunicação científica, que restringia o acesso aos resultados de pesquisas a revistas fechadas (acessíveis mediante assinatura). Essas manifestações, de forma prática, resultaram na criação dos primeiros periódicos eletrônicos e repositórios digitais que, suportados por *softwares open source* e pela internet, representaram o início de um processo de abertura gradual da atividade científica.<sup>6</sup> Como exemplo de repositório, destaca-se o ArXiv (1990), disponibilizando *preprints* nas áreas de física, matemática e ciência da computação, e englobando os princípios da OAI.

No contexto internacional, em 1998 ocorreu a criação da The Scholarly Publishing and Academic Resources (Sparc), como um acordo universitário em apoio para melhorar o acesso aberto. Em 2001, foi criada a Public Library of Science (PLOS), como um pacto envolvendo diversos países para ofertar artigos científicos em acesso aberto.

No âmbito nacional, o lançamento da *Scientific Eletronic Library* (SciELO), em 1998, destaca-se como um exemplo das primeiras iniciativas em ciência aberta nacional, devido ao oferecimento de uma plataforma de periódicos científicos, revisados por pares, disponível de forma aberta pela internet para toda a sociedade. Em 2005, o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) lançou um manifesto em prol do acesso aberto.

O manifesto do IBICT em 2005 ecoou as iniciativas nacionais existentes, como, por exemplo, o lançamento, em 2004, do Repositório Digital Institucional (RDI) da UFPR,<sup>7</sup> composto por três bibliotecas digitais (teses e dissertações, periódicos eletrônicos

<sup>4</sup> Stallman, *Software livre para una sociedad libre* (2004).

<sup>5</sup> Rios et al., “Manifestos do movimento de acesso aberto” (2019).

<sup>6</sup> Setenareski et al., *A dinâmica competitiva do mercado mundial de publicações científicas* (2020).

<sup>7</sup> Sunye et al., “A experiência da UFPR na construção de repositórios digitais, a implantação integrada das ferramentas DSpace, Open Journal System” (2009).

institucionais e biblioteca de digital de imagem e som) e suportado pelos *softwares* livre *DSpace* e *Open Journal System* (OJS), esse último elaborado e mantido pela *Public Knowledge Project* (PKP), entidade sem fins lucrativos pioneira em soluções de *software* para o acesso aberto.

Na esteira do ArXiv, surgiram outras iniciativas de publicação em acesso aberto, tanto com periódicos quanto com a disponibilização integral de teses e dissertações, *preprints* e outros tipos de publicações acadêmicas, científicas e culturais. Nesse sentido, o surgimento dos repositórios digitais conferiu agilidade e gratuidade na obtenção do texto completo de teses e dissertações, fazendo com que serviços pagos de solicitação de documentos como o COMUT, por exemplo, perdessem relevância.

Na *e-Science*, o desenvolvimento da prática científica é suportado pelas tecnologias de informação e comunicação dentro de uma infraestrutura computacional que oferece ao pesquisador plataformas, ferramentas e ambiente para o desenvolvimento da pesquisa de forma integrada e colaborativa com outros pesquisadores e atores. À medida que o desenvolvimento computacional avança, as barreiras ao acesso aberto diminuem e a disponibilidade de publicações aumenta, o que leva a um novo patamar dentro do acesso aberto: a abertura e exploração de dados científicos de pesquisa, unidade nuclear, básica e fundamental da ciência.

## *Dados científicos de pesquisa: disponibilização, compartilhamento e reuso*

Os dados científicos de pesquisa podem ser entendidos como um registro pontual e preciso que representa o resultado de uma manifestação, ação, reação e/ou acontecimento. Dito de outro modo, dados científicos de pesquisa são os dados coletados, organizados e analisados a fim de fornecerem uma informação sobre determinado fato.

A literatura da área indica que não há ainda um consenso sobre a conceituação de dados científicos de pesquisa.<sup>8</sup> A definição de um conceito para dados deve ser precedida de contexto e abordagem, pois o dado pode representar informações distintas em cada área do conhecimento.<sup>9</sup> Uma visão sobre os dados é que eles são considerados a “força vital da ciência e tecnologia” e devem ser processados por humanos e máquinas.<sup>10</sup>

Os dados científicos podem se caracterizar como texto, vídeos, imagens, objetos tridimensionais, objetos digitais e impressos, nu-

<sup>8</sup> Borgman, “The conundrum of sharing research data” (2012).

<sup>9</sup> da Silva, *Gestão de dados científicos* (2019).

<sup>10</sup> Frederick, “Data, open science and libraries: the data deluge column” (2016).

méricos, morfológicos, taxonômicos, amostras biológicas, químicas e físicas, de ressonância, códigos, tabelas, gráficos, protocolos e demais variedades que dependem de contexto e abordagem. Logo, “os dados científicos são todas as evidências de que um investigador necessita para validar suas conclusões após uma pesquisa”.<sup>11</sup>

Em sua tipologia, os dados científicos classificam-se como do tipo observacional, computacional e experimental. Os dados observacionais são provenientes da observação direta, em tempo real e, por isso, únicos, e assim podem ser qualificados como registros históricos, como, por exemplo, dados meteorológicos ou dados de uma observação etnográfica. Os dados computacionais provêm da execução de simulações e modelos computacionais que podem ser reproduzidos se preservadas as condições de *hardware* e *software*. Como exemplo de dados computacionais, têm-se dados advindos de simulações em realidade virtual. Os dados do tipo experimental decorrem da execução de procedimentos controlados para comprovação de fenômenos e estão sujeitos à reprodução se também preservadas as condições de ambiente, padrões e procedimentos. Os dados experimentais podem ser resultados de estudos laboratoriais, ou ainda dados provenientes de estudos comportamentais.

Assim, tem-se então que os dados científicos são dados primários que servem para comprovar ou refutar hipóteses, atestar ou não a validade de dados. Entretanto, quanto ao uso, os dados ou conjuntos de dados (*datasets*) podem ser classificados em abertos ou restritos. A abertura e restrição é designada pelo(s) autor(res) ou pela instituição financiadora da pesquisa, sendo o uso, compartilhamento e reuso regulamentados conforme o tipo de licença de direito autoral e pela documentação atrelada aos *datasets*, e/ou às políticas dos repositórios ou ambientes onde os dados estão armazenados.

A abertura irrestrita dos dados no âmbito da ciência aberta significa a possibilidade de exploração de dados no sentido de uso, reuso, compartilhamento e modificações. Os Princípios FAIR e os Princípios Panton são duas iniciativas que auxiliam os pesquisadores na conceitualização dos dados abertos, recomendando formas de fazê-lo.

Os Princípios FAIR<sup>12</sup> preconizam que um dado aberto é um dado encontrável (*Findable*), acessível (*Accessible*), interoperável (*Interoperable*) e reusável (*Reusable*), sendo que cada princípio possui requisições específicas para cada função. Por exemplo, o princípio de reusabilidade requer que os dados possuam metadados (descrições sobre dados) completos com atributos relevantes e precisos por meio de licenciamento claro e específico, associado a uma procedência detalhada e seguindo padrões relevantes de domínio. Ou seja, para ser FAIR, um dado deve ser recuperável por máquina, estar acessível, poder ser recuperado por meio de

<sup>11</sup> da Silva, *Gestão de dados científicos*, p. 21 (2019).

<sup>12</sup> Wilkinson et al., “The FAIR guiding principles for scientific data management and stewardship” (2016).

protocolos de interoperabilidade, como o *Open Archives Initiative Protocol Metadata Harvesting* (OAI-PMH) ou o *Open Archive Initiative Object Reuse and Exchange* (OAI-ORE), e ser reusável por meio da aplicação de licenciamento aberto do tipo *Creative Commons* (CC). O tipo de licença de direito autoral mínima recomendada é a cc-by, que representa, no mínimo, a obrigatoriedade de créditos ao autor original.

Os princípios Panton<sup>13</sup> representam a abertura irrestrita dos dados, de modo que um dado possa ser usado apropriado (inclusive comercialmente), compartilhado, reusado e modificado sem que exista a necessidade de conferir créditos ao autor original do dado ou *dataset*. O tipo de licença recomendado pela Panton é a licença do tipo CCZero (renúncia aos direitos autorais).

Nessa acepção, observa-se que tornar os dados científicos disponíveis de forma aberta e irrestrita significa colocar à disposição da comunidade científica e da sociedade a possibilidade de obter novas descobertas e análises, complementar ou ressignificar outras pesquisas, tudo a partir do compartilhamento e reuso de dados.

A disponibilização aberta, o compartilhamento e o reuso de dados científicos apresentam vantagens e benefícios para a ciência em si, para os pesquisadores e para a sociedade, tais como: validar os processos científicos e seus resultados; promover a reprodutibilidade de pesquisas; promover a preservação digital de longo prazo, mantendo a integridade dos dados; proporcionar economia de recursos na fase de coleta de dados; promover novas análises e ressignificações de pesquisas; promover a transparência na aplicação de recursos e financiamento; promover visibilidade dos pesquisadores na forma de citações, convites para colaboração, ampliação das redes de pesquisa; e tornar visível e acessível para a sociedade as atividades das universidades, centros e institutos de pesquisa.<sup>14</sup>

Em uma abordagem simplista, o compartilhamento de dados representa a liberação dos dados para outros, enquanto o reuso é a ação de utilizar os dados disponibilizados abertamente por outros em uma pesquisa secundária. Entretanto, tanto o compartilhamento quanto o reuso de dados enfrentam barreiras e desafios que residem nas seguintes dimensões: política, organizacional, sócio-cultural, estrutural-tecnológica, econômica e legal.<sup>15</sup>

A dimensão política está relacionada com elaboração de políticas públicas de informação pelas entidades governamentais e/ou financiadoras que contemplem planejamento, estratégias, objetivos e recursos, a fim de incentivar e viabilizar a disponibilização, acesso, preservação, compartilhamento e reuso de dados.

A dimensão organizacional está relacionada com a forma pela qual as organizações definem suas estruturas e viabilizam a publicação dos dados de pesquisa, seja tanto no sentido de conscientização dos benefícios e vantagens da disponibilização, com-

<sup>13</sup> Murray-Rust et al., *Panton principles* (2010).

<sup>14</sup> Tenopir et al., "Data sharing by scientists" (2011).

<sup>15</sup> Bueno de la Fuente, *Challenges and strategies for the success of open science* (2016); de Oliveira e da Silva, "Ciência aberta" (2016).

partilhamento e reuso, quanto no oferecimento de condições estruturais de trabalho aos pesquisadores.

A dimensão sócio-cultural está ligada à forma como o pesquisador observa e compreende suas atividades de pesquisa e ao quanto está disposto a compartilhar sua produção com outros. Esta dimensão apresenta-se com um caráter pessoal, pois é relacionada com o apego do autor em relação aos seus dados de pesquisa. Dito de outro modo, a dimensão pessoal diz respeito a quanto o pesquisador está disposto a conceder o uso e reuso dos dados produzidos em sua pesquisa. Essa dimensão divide espaço com a dimensão organizacional da feita de que é preciso um trabalho organizacional conjunto e coletivo para incentivar o pesquisador a compartilhar seus dados.

A dimensão estrutural-tecnológica representa os desafios computacionais e de infraestrutura tecnológica que oportunizem ao pesquisador um ambiente adequado ao trabalho com dados em todas as fases da pesquisa, oferecendo, também, ferramentas para manipulação dos dados e repositórios de dados para armazenamento, curadoria, preservação a longo prazo, acesso e reuso de dados.

A dimensão econômica está interligada com a parte política e organizacional, pois ainda que a disponibilização, o uso e o reuso de dados sejam abertos e gratuitos, existem custos de armazenamento, manutenção e suporte dos dados.

A dimensão legal representa o tipo de licenciamento atrelado aos *datasets* e às condições pré-estabelecidas pelo autor original ou pelo repositório que abriga o dado. Essa dimensão deve atender às novas requisições da Lei Geral de Proteção de Dados (LGPD)<sup>16</sup> no tocante aos dados pessoais sensíveis, consentimento e anonimização.

<sup>16</sup> Lei nº 13.709/18, Lei geral de proteção de dados pessoais (2018).

As dimensões elencadas acima representam uma parte dos desafios quanto ao uso, compartilhamento e reuso de dados científicos de pesquisa. Outra dimensão relevante consiste na elaboração dos repositórios de dados, que são as plataformas por meio das quais serão disponibilizados os dados para acesso. Nesse tocante, desafios como escalabilidade, heterogeneidade e complexidade e qualidade devem ser considerados na elaboração dos repositórios.

## *Repositório de Dados Científicos de Pesquisa: desafios*

Os repositórios de dados são plataformas digitais tecnicamente estruturadas para armazenar e disponibilizar os dados e seus atri-

butos, como: objeto digital, metadados (os atributos dos dados, como autor, título do objeto, tipo, proveniência, entre outros), texto legal sobre licenciamento, documentação e identificador persistente. Os repositórios de dados então cumprem as funções de armazenar, disponibilizar, coletar, preservar e realizar curadoria dos dados. São estruturas hierarquizadas e modulares que devem atender a padrões internacionais de descrição, preservação e interoperabilidade de forma a proporcionar o uso e o reuso de dados sob condições legais.

A concepção e implantação de repositórios de dados científicos é um dos maiores desafios para a Ciência Aberta. Dados científicos possuem dois estados principais: crus (*raw data*) ou já processados. Alguns autores falam em “*dark data*”, que são dados brutos que o pesquisador sequer utilizou em seus estudos.<sup>17</sup>

Dados crus, ou ainda não processados, correspondem exatamente àquilo que foi observado, coletado ou gerado em um experimento científico. Esses dados podem ser fundamentais para a reprodutibilidade de um experimento, ou mesmo para assegurar sua confiabilidade. Os dados processados já passaram por algum processo de transformação ou mesmo simplificação, e também são necessários para assegurar a confiabilidade da pesquisa científica.

Sendo assim, um repositório de dados científicos deve, ao mesmo tempo em que descreve e modela os dados, também descrever seus estados e transformações.

A complexidade da construção de Repositórios de Dados Científicos leva em conta ainda novos desafios, como: a escalabilidade que é a capacidade de tratar o tamanho e a velocidade de crescimento da massa de dados; a heterogeneidade e complexidade dos dados, que variam muito em cada subárea do conhecimento; e a qualidade dos dados produzidos pelo pesquisador.

## Escalabilidade

O avanço da precisão e do detalhamento dos equipamentos utilizados em experimentos científicos faz com que o volume de dados gerados a cada experimento cresça exponencialmente, gerando problemas de escala no volume de dados armazenados. Um exemplo do volume de dados que podem ser gerados por equipamentos de pesquisa é o acelerador de partículas *Large Hadron Collider* (LHC), da Organização Europeia para Pesquisa Nuclear (Cern), que captura 25 *petabytes* de dados todos os anos.

Planejar o crescimento de um repositório científico cujo potencial de crescimento cresce a cada nova geração de equipamentos é um desafio permanente de planejamento financeiro e tecnológico.

<sup>17</sup> Schembera e Durán, “Dark data as the new challenge for big data science and the introduction of the scientific data officer” (2020).



## Heterogeneidade e complexidade

Os equipamentos de pesquisa, além de evoluírem rapidamente na capacidade de coleta e análise e, portanto, na quantidade de dados gerados, evoluem também no tipo de informação que é gerada. Um equipamento que produz apenas espectros em uma determinada análise pode passar a gerar imagens ou imagens 3d, modificando a estrutura do dado gerado.

A própria definição do que é dado cru ou dado processado pode mudar com a evolução do equipamento que pode passar a processar os dados, ao mesmo tempo em que realiza a coleta da massa de informação. Repositórios de dados científicos, mesmo que sejam segmentados em áreas e subáreas, são heterogêneos, e dois *datasets* de um mesmo equipamento podem ter estruturas completamente diferentes.

Encontrar modelos de descrição dos dados em um repositório de dados científicos que consiga levar em conta a complexidade estrutural dos dados é também um desafio permanente.

## Qualidade

Os dados gerados pelos pesquisadores, independentemente de serem dados crus ou já processados, podem ter baixa qualidade. Tanto o dado coletado como capturado em algum experimento está sujeito a erros. Por exemplo, a técnica utilizada no uso do equipamento pode ser inadequada, ou a distribuição amostral de dados coletados pode ser distorcida.

Assim, uma solução possível é a exigência de uma análise por pares dos dados gerados antes de sua submissão a um repositório de dados. O processo de revisão por pares, ainda que dentro de um processo de revisão aberta, como preconiza a Ciência Aberta, iria conferir, neste caso, qualidade, confiabilidade e autenticidade aos dados submetidos.

Os desafios acima descritos visam sanar as lacunas estruturais na construção de repositórios de dados para garantir a reprodutibilidade, que é um dos objetivos da Ciência Aberta. E a reprodutibilidade é compreendida não apenas para o uso, a modificação e a republicação dos dados científicos, mas também para a replicação dos procedimentos, códigos, equipamentos e ferramentas utilizados na prática científica. A reprodutibilidade de dados e procedimentos requer a definição de estratégias computacionais e metodológicas para tratar os processos de análise e experimentação de dados de forma que esses possam ser validados e passíveis de reaproveitamento.

Assim, os repositórios de dados devem ser implementados considerando-se a diversidade da tipologia de dados, o volume de dados, a defini-

ção de estratégias de preservação digital (para além de *backup*), de forma a garantir a conservação e acesso aos dados a longo prazo, preservando-se, assim, a memória científica. Os repositórios devem possuir políticas e diretrizes claras de submissão de dados descrevendo seus padrões, critérios, serviços oferecidos, termos de uso e suporte.

Os repositórios de dados concentram apenas parte da infraestrutura sobre dados científicos de pesquisa. No entanto, é talvez uma das partes relevantes, visto que representa o ponto de acesso por meio do qual os dados tornam-se visíveis e acessíveis e, por conseguinte, prontos para o reuso. O planejamento de repositório de dados deve considerar a área ou subáreas do conhecimento, respeitando os distintos padrões de descrição entre as áreas, de modo que seja possível a recuperação de informação por máquinas e humanos.

O *Re3data* é um diretório confiável de repositório de dados disciplinares e multidisciplinares por meio do qual é possível pesquisar os repositórios por área do conhecimento, tipo de dados, país, assunto, entre outros. Hoje, o *Re3data* aponta a existência de 11 repositórios de dados no Brasil, sendo a maioria da área multidisciplinar. O *Re3data* indica a existência de quatro repositórios de dados na área de Humanidades/Ciências Sociais Aplicadas e Linguística, por exemplo.<sup>18</sup>

<sup>18</sup> <https://www.re3data.org>.

## Considerações finais

A temática da Ciência Aberta e a exploração de dados científicos de pesquisa, embora não sejam assuntos novos, ainda são áreas em exploração no Brasil, carecendo de estudos quanto a definições, estruturação de dados, reusabilidade e repositório de dados. O objetivo deste texto foi situar os dados científicos no âmbito da Ciência Aberta, definindo os principais conceitos, bem como sua relação com o ambiente acadêmico, e o uso de ferramentas *open source*.

Os principais desafios com relação ao uso, ao compartilhamento e ao reuso de dados foram explicitados, de forma a assinalar ao pesquisador possibilidades e vantagens quanto à disponibilização de seus dados, bem como na exploração secundária de dados primários.

No contexto internacional, a abertura dos dados tornou-se ação mandatária nas políticas das agências de fomento e financiamento à pesquisa com dois objetivos distintos: tornar disponível e acessível a produção científica (publicações, dados e demais etapas do ciclo de vida da pesquisa) para fora do meio acadêmico, possibilitando a apropriação do conhecimento pela sociedade

de modo geral; e buscar promover transparência na aplicação de recursos, sejam públicos, privados ou mistos.

A transparência quanto à aplicação de recursos públicos, em se tratando da abertura de dados, permite o controle social em relação às ações de entidades governamentais e/ou com aporte público, e ainda a elaboração de indicadores sobre o investimento em ciência.

No Brasil, embora ocorram discussões sobre a abertura de dados em ações coordenadas pelo IBICT, Rede Nacional de Ensino e Pesquisa (RNP) e Ministério de Ciência, Tecnologia, Inovações e Comunicações (MCTIC) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), a abertura dos dados é apenas incentivada e não requerida obrigatoriamente. A ampla maioria das agências de fomento nacionais, excetuando a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), não possuem políticas e/ou indicações mandatárias em seus editais quanto à abertura dos dados e armazenamento em repositórios.<sup>19</sup> Nesse sentido, deve haver um estreitamento da discussão entre as entidades coordenadoras de políticas nacionais de informação científica com as agências de financiamento para que ocorra adequação dos editais de apoio a pesquisa e inclusão de ações de disponibilização e abertura de dados científicos.

A construção de repositório de dados constitui um dos maiores desafios atuais para a Ciência Aberta e é discussão relevante e necessária em prol do desenvolvimento da atividade científica e na transparência científica.

<sup>19</sup> de Lima, “Impacto percebido pelos pesquisadores quanto ao reuso de dados científicos de pesquisa em repositório de dados brasileiros” (2020).

## Referências

- Albagli, Sarita, Maria Lucia Maciel e Alexandre Hannud Abdo, editores (2015). *Ciência aberta, questões abertas*. Brasília: IBICT. ISBN: 9788570131096. URL: <http://livroaberto.ibict.br/handle/1/1060>.
- Borgman, Christine L. (2012). “The conundrum of sharing research data”. *Journal of the American Society for Information Science and Technology* 63.6, pp. 1059–1078. DOI: 10.1002/asi.22634.
- Bueno de la Fuente, Gema (2016). *Challenges and strategies for the success of open science*. URL: <https://www.fosteropenscience.eu/node/1424>.
- Frederick, Donna Ellen (2016). “Data, open science and libraries: the data deluge column”. *Library Hi Tech News* 33.8, pp. 11–16. DOI: 10.1108/LHTN-09-2016-0040.
- Hey, Tony, Stewart Tansley e Kristin Tolle, editores (2009). *The fourth paradigm: data-intensive scientific discovery*. Redmond: Microsoft Corporation. ISBN: 9780982544204.
- de Lima, Karolayne Costa Rodrigues (2020). “Impacto percebido pelos pesquisadores quanto ao reuso de dados científicos de pesquisa em repositório de dados brasileiros”. Dissertação de mestrado. Curitiba: Universidade Federal do Paraná. URL: <https://www.prppg.ufpr.br/signa/visitante/trabalhoConclusaoWS?idpessoal=64041&idprograma=40001016058P1&anobase=2020&idtc=50>.

- Murray-Rust, Peter, Cameron Neylon, Rufus Pollock e John Wilbanks (2010). *Panton principles: principles for open data in science*. URL: <https://pantonprinciples.org/>.
- de Oliveira, Adriana Carla Silva e Edilene Maria da Silva (2016). “Ciência aberta: dimensões para um novo fazer científico”. *Informação & Informação* 21.2, pp. 5–39. DOI: 10.5433/1981-8920.2016v21n2p5.
- Rios, Fahima, Elaine Oliveira Lucas e Igor Soares Amorim (2019). “Manifestos do movimento de acesso aberto: análise de domínio a partir de periódicos brasileiros”. *Revista Brasileira de Biblioteconomia e Documentação* 15.1, pp. 148–169.
- Schembera, Björn e Juan M. Durán (2020). “Dark data as the new challenge for big data science and the introduction of the scientific data officer”. *Philosophy & Technology* 33.1, pp. 93–115. DOI: 10.1007/s13347-019-00346-x.
- Setenareski, Ligia E., Walter Shima e Marcos S. Sunye (2020). *A dinâmica competitiva do mercado mundial de publicações científicas: tendências e alternativas do acesso aberto*. Curitiba: Appris. ISBN: 9788547334963.
- da Silva, Fabiano Couto Corrêa (2019). *Gestão de dados científicos*. Rio de Janeiro: Interciência. ISBN: 9788571934351.
- Stallman, Richard M. (2004). *Software libre para una sociedad libre*. URL: [https://www.gnu.org/philosophy/fsfs/free\\_software.es.pdf](https://www.gnu.org/philosophy/fsfs/free_software.es.pdf).
- Sunye, Marcos, Ligia E. Setenareski, Fabiano Couto Corrêa da Silva, Edson Ramiro e Lucio Foltran (2009). “A experiência da UFPR na construção de repositórios digitais, a implantação integrada das ferramentas DSpace, Open Journal System”. In: *Implantação e gestão de repositórios institucionais: políticas, memória, livre acesso e preservação*. Editado por Luis Sayão, Lídia Brandão Toutain, Flavia Garcia Rosa e Carlos Henrique Marcondes. Salvador: EDUFBA, pp. 107–122. ISBN: 9788523206550.
- Tenopir, Carol, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Mari-beth Manoff e Mike Frame (2011). “Data sharing by scientists: practices and perceptions”. *PloS one* 6.6, e21101. DOI: 10.1371/journal.pone.0021101.
- Wilkinson, Mark D. et al. (2016). “The FAIR guiding principles for scientific data management and stewardship”. *Scientific Data* 3.1, pp. 1–9. DOI: 10.1038/sdata.2016.18.

# G