

Contribuições da ortografia para a análise de fenômenos fonológicos: o caso de sequências #(i)sC no português brasileiro

The contributions of orthography in comprehending phonological phenomena: the case of #(i)sC sequences in Brazilian Portuguese

DOI: <https://doi.org/10.24206/lh.v7i2.40702>

Matheus Freitas

Doutorando do Programa de Pós-Graduação em Estudos Linguísticos da Universidade Federal de Minas Gerais com bolsa da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. Mestre em Estudos Linguísticos pela mesma instituição.

E-mail: matheusfgomes@outlook.com

ORCID: <https://orcid.org/0000-0001-9395-015X>

RESUMO

Este artigo trata de sequências formadas por sibilante [s] + consoante que podem ou não ser precedidas por uma vogal [i] em início de palavra no português brasileiro. Essas sequências são referidas como sequências #(i)sC. Exemplos de palavras que apresentam as sequências #(i)sC investigadas neste estudo são *escola*: [iskɔla] ~ [skɔla] e *spray*: [isprej] ~ [sprej]. As palavras analisadas compreendem padrões ortográficos distintos: <#ESC> e <#SC>. Foi testada a hipótese de que cada padrão ortográfico se relaciona a diferentes taxas de ocorrência da vogal e a diferenças no detalhe fonético em sequências #(i)sC por meio de uma investigação experimental. Os resultados demonstram que a ocorrência da vogal inicial é favorecida em palavras com o padrão <#ESC> em comparação a palavras com o padrão <#SC>. Também foram atestadas diferenças na duração da vogal, mais longa em palavras com o padrão <#ESC> do que em palavras com o padrão <#SC>. É discutido como que padrões ortográficos, que têm trajetória histórica distinta na língua, podem se relacionar a diferentes realizações fonéticas. Também é explorado o papel que o detalhe fonético tem em caracterizar diferentes conjuntos de palavras. Argumenta-se que a forma ortográfica das palavras pode representar evidência útil para a análise de fenômenos fonológicos. Para isso, a análise é realizada à luz dos modelos teóricos da Teoria de Exemplares e da Integração de Múltiplos Padrões.

Palavras-chave: Sibilantes. Ortografia. Redução vocálica. Detalhe fonético. Teoria de Exemplares.

ABSTRACT

This paper addresses word-initial sequences formed by a sibilant [s] + a consonant that may or may not be preceded by an [i] vowel in Brazilian Portuguese. These sequences are referred to as #(i)sC sequences. Examples of words that present the #(i)sC sequences investigated in this study are *escola*: [iskɔla] ~ [skɔla] 'school' and *spray*: [isprej] ~ [sprej]. The investigated words comprise two different orthographic patterns: <#ESC> and <#SC>. We tested the hypothesis that each orthographic pattern is related to different rates of vowel occurrence and to differences in the phonetic detail in #(i)sC sequences, following an experimental approach. Results indicate that vowel occurrence is favoured in <#ESC> words rather than in <#SC> words. Differences in vowel duration were also identified, the vowel is longer in <#ESC> words than in <#SC> words. We discussed how orthographic patterns, which have unique historical trajectories in the lexicon, might be related to different phonetic

realisations. The role of phonetic detail in characterising different sets of words is also explored. We claim that spelling can be useful evidence to analyse phonological phenomena. We investigated the data in light of Exemplar Models and Integration of Multiple Patterns models.

Keywords: Sibilants. Orthography. Vowel weakening. Phonetic detail. Exemplar Models.

Introdução¹

Este artigo tem como objetivo avaliar sequências formadas por sibilante [s] + consoante (C: [p, t, k]), que podem ou não ser precedidas por uma vogal alta anterior [i] em início de palavra, no português brasileiro (PB), o que caracteriza a alternância entre as formas #isC ~ #sC. Doravante, essas sequências serão referidas como sequências #(i)sC. Exemplos de palavras que apresentam as sequências #(i)sC investigadas neste estudo são *escola*: [iskola] ~ [skola] e *spray*: [isprej] ~ [sprej], que apresentam padrões ortográficos distintos: respectivamente <#ESC> e <#SC>. Pretende-se discutir como esses padrões ortográficos, historicamente associados a conjuntos diferentes de palavras, se relacionam a diferentes realizações de sequências #(i)sC, que apresentam aparentemente a mesma sequência segmental.

Como hipótese, foi assumido que palavras que apresentam diferentes padrões ortográficos podem se associar a diferentes taxas de ocorrência e a diferentes valores duracionais da vogal inicial, ainda que correspondam a uma mesma sequência segmental. Os pressupostos que fundamentam a discussão partem dos modelos da Integração de Múltiplos Padrões (TREIMAN; KESSLER, 2014; TREIMAN, 2017) e da Teoria de Exemplares (JOHNSON, 1997, 2007; PIERREHUMBERT, 2001, 2003; BYBEE, 2001, 2006, 2010). Foi realizada uma investigação experimental, avaliando a sequência sonora inicial #(i)sC nos dois padrões ortográficos, <#ESC> (ex.: *escola*, *espaço*, *estrada*) e <#SC> (ex.: *skol*, *spa*, *stress*). A decisão por uma investigação experimental se ancora nos preceitos teórico-metodológicos da Fonologia de Laboratório (PIERREHUMBERT; BECKMAN; LADD, 2011).

Para além da introdução, este trabalho se divide em cinco seções. Na primeira seção, é explorado o histórico dos padrões ortográficos <#ESC> e <#SC> no sistema de escrita da língua portuguesa. Pretende-se apresentar que, embora correspondam à mesma sequência segmental, os padrões <#ESC> e <#SC> têm trajetórias específicas na língua portuguesa e estão associados a diferentes grupos de palavras no léxico. Na segunda seção, é discutida a interação entre padrões ortográficos e propriedades linguísticas, com especial foco nos impactos de formas ortográficas na adaptação de empréstimos. Nessa seção, são explorados pressupostos teóricos do modelo de Integração de Múltiplos Padrões. Argumenta-se que a forma escrita das palavras pode ser evidência relevante para a análise de fenômenos fonológicos. Na terceira seção, são revisados o caráter gradiente do fenômeno investigado

¹ Este artigo é fruto de uma adaptação de partes da dissertação de mestrado de Freitas (2019), intitulada “A redução segmental em sequências #(i)sC no português brasileiro”, realizada sob orientação da Profa. Dra. Thaís Cristófaros Alves da Silva e disponível em: <http://www.poslin.letras.ufmg.br/defesas/1994M.pdf>. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001. Os dados coletados para este estudo foram discutidos sob um outro viés, voltado especificamente ao estatuto representacional das sequências #sC, no artigo de Cristófaros Silva e Freitas (2020), intitulado “sC-clusters in Brazilian Portuguese”, disponível em: <http://doi.org/10.5334/jpl.231>.

neste estudo no PB e os pressupostos da Teoria de Exemplares, a fim de motivar que o detalhe fonético pode caracterizar sequências #(i)sC de maneira distinta em diferentes grupos de palavras. Na quarta seção, são expostos os procedimentos metodológicos adotados para a presente investigação. Por fim, na quinta e última seção, são apresentados os resultados alcançados e uma discussão orientada a partir dos objetivos apresentados nesta introdução. As seções são seguidas por considerações finais, em que se resumem aspectos essenciais deste trabalho.

1. A história de <#ESC> e <#SC> na língua portuguesa

O sistema de escrita adotado para a representação da língua portuguesa é um sistema alfabético e ortográfico. Enquanto um sistema alfabético, o sistema de escrita prevê uma estreita correspondência entre seus grafemas e os fonemas do sistema fonológico do português (SCLIAR-CABRAL, 2003). Comparado a um sistema sônico ou fônico, cuja correspondência sonora-visual é atestada a partir de sons ou fones, o alfabeto é econômico e “satisfaz as necessidades de representação básica (fonológica) com pouco mais de vinte unidades” (GONÇALVES, 2010, p. 8). Por outro lado, se comparado a um sistema de base silábica ou ideográfica, baseados em sílabas ou unidades de sentido, o alfabeto se aproxima “de uma correspondência mais transparente com o sistema fonético/fonológico da língua” (GONÇALVES, 2010, p. 5).

Ademais, enquanto um sistema ortográfico, o sistema de escrita da língua portuguesa apresenta correspondências fonema-grafema inconsistentes, reguladas por um grau de arbitrariedade convencional, afastando-se da biunivocidade entre as unidades sonoras e gráficas que o caracterizariam como um sistema inteiramente transparente. Na condição de falta de transparência de um sistema alfabético, Soares (2018, p. 288) assume que “a palavra *ortografia* adquire outro significado, o de *grafia correta*, [...] a grafia de acordo com normas que estabelecem a escrita convencional em casos de correspondências que fogem à biunivocidade”. Um sistema ortográfico, como o da língua portuguesa, opera em um funcionamento próprio que, em certa medida, transcende as características do sistema sonoro no qual se baseia.

Para que se entenda a sistematicidade da escrita da língua portuguesa no que diz respeito a sua ortografia, é essencial discorrer sobre como esse sistema foi concebido e implantado. A história da ortografia da língua portuguesa pode ser segmentada em três períodos históricos (cf. GONÇALVES, 2010): os períodos fonético, pseudoetimológico e contemporâneo.

O período fonético abrange o intervalo que vai do século XII, marco dos primeiros textos escritos em língua portuguesa, até o século XVI, sendo caracterizado pela falta de sistematicidade dos padrões gráficos. Ao contrário do que pode sugerir a sua denominação, o estabelecimento das formas

gráficas no período fonético não se pauta em propriedades fonéticas, mas sim em formas arbitrárias, caracterizando o período como desprovido de regras ortográficas.

Entre o século XVI e o início do século XX, demarca-se o período pseudoetimológico da ortografia da língua portuguesa, caracterizado tanto pela busca por padrões ortográficos baseados nos étimos latinos e gregos quanto por comparações sonoras e gráficas entre o português e outras línguas românicas. É em tal intervalo que se tem início a publicação dos primeiros manuais de ortografia em língua portuguesa. Madureira Feijó (1734), ortografista do século XVIII, assume que uma ortografia baseada na pronúncia traria danos à identificação das palavras, devendo as prescrições sobre ortografia ser respaldadas pelos estudos consagrados sobre as ortografias grega e latina. Diferentemente do período fonético, passam a ser prescritas formas ortográficas tidas como “adequadas”. Ainda assim, as produções em tal período apresentam grande assistemática e até mesmo incongruências etimológicas, garantindo sua denominação como *pseudoetimológico*.

Com a publicação da *Ortografia Nacional* de Gonçalves Viana, no ano de 1904, tem-se início o período contemporâneo, ou período das reformas ortográficas, da ortografia da língua portuguesa. Na *Ortografia Nacional*, é proposto um sistema de escrita que, dentro do possível, contempla todas as variedades da língua portuguesa e que prioriza seu próprio funcionamento interno enquanto um sistema, demarcando contrastes sonoros na escrita e descartando padrões etimológicos que não se justifiquem pela língua falada. As regras propostas por Gonçalves Viana fundamentam, até os dias de hoje, a base dos acordos ortográficos da língua portuguesa. Quanto aos objetivos da *Ortografia Nacional*, almejava-se “um equilíbrio entre a funcionalidade imediata do sistema e a sua historicidade, tentava articular o critério histórico com o critério fonético” (GONÇALVES, 2010, p. 12). Assim como no período pseudoetimológico, são prescritas formas ortográficas tidas como “adequadas”. Contudo, na *Ortografia Nacional* uma alta sistematicidade nas produções escritas em língua portuguesa é pretendida, de forma que seja estabelecida uma só ortografia.

Com relação aos padrões ortográficos ⟨#ESC⟩ e ⟨#SC⟩, é pertinente explorar seu estatuto nos dois mais recentes períodos históricos da língua portuguesa, visto que o primeiro deles, o período fonético, não prescreve quaisquer formas ortográficas. Durante o período pseudoetimológico, é prescrito o uso do padrão ⟨#SC⟩ em início de palavra, que se justifica por étimos latinos ou gregos. Sobre o uso do padrão ⟨#SC⟩, Madureira Feijó (1734, p. 502) avalia que “[a]inda, que na nossa lingua todas as palavras, que no Latim principiaõ por s, e consoante, pôdem principiar por e; com tudo, ha huas taõ alatinadas, que seria impropriedade de ãõ se escreverem com a mesma ortografia”². Prezando pela preservação da forma ortográfica ⟨#SC⟩, o autor propõe como exemplos os vocábulos *spirito* e *squelêto* (MADUREIRA FEIJÓ, 1734, p. 503). Contudo, a proposta carece de critérios sistemáticos, o que se

² Esta citação foi transcrita conforme a versão original.

evidencia pela sugestão, em outro trecho da obra, de que as palavras exemplificadas sejam grafadas com uma vogal ⟨E⟩ inicial como *espírito* e *esquelêto* (MADUREIRA FEIJÓ, 1734, p. 305-306). Apesar de a grafia do período pseudoetimológico privilegiar a forma ⟨#SC⟩ em detrimento da forma ⟨#ESC⟩ por identidade com a grafia latina, encontram-se possibilidades concorrentes.

Já no período contemporâneo, a partir dos parâmetros que determinam as regras ortográficas da língua portuguesa, torna-se objetiva a compreensão sobre o estatuto ortográfico das palavras compostas por ⟨S⟩ + consoante em posição inicial no português. Sobre tais sequências, Gonçalves Viana (2010, p. 105) afirma que

[n]ão havendo dição portuguesa começada por s e outra consoante, com qual vogal há de êste s fazer sílaba? Acaso não se pronuncia o st e sp destas e outras palavras exactamente como em estriga, estreme, esparto e esméro? Para quê, pois, se há de dar estranho aspecto àquelas dições, em desacôrdo com a escrita de todas as demais palavras portuguesas de estrutura fonética semelhante?

A partir da proposta ortográfica de Gonçalves Viana, o padrão ⟨#SC⟩ passa a ser considerado em desacordo com a grafia das demais palavras portuguesas. Sendo assim, é avaliado como um caso de preciosismo etimológico inadequado, pois violaria o princípio de que o sistema ortográfico deve priorizar seu funcionamento interno em detrimento de padrões etimológicos. Uma vez que as reformas ortográficas da língua portuguesa são balizadas pelos princípios da *Ortografia Nacional*, o padrão ortográfico ⟨#SC⟩ não é comportado pela ortografia atual da língua portuguesa, sendo considerado um padrão ilícito. Para as palavras da língua portuguesa cujo étimo se relaciona ao padrão ⟨#SC⟩, é prescrita apenas a forma ⟨#ESC⟩.

É crucial salientar, no entanto, que o padrão ortográfico ⟨#SC⟩, ilícito na língua portuguesa, persiste no léxico mesmo no período contemporâneo da ortografia. Línguas como o inglês, o italiano e o russo – das quais muitos estrangeirismos foram incorporados ao léxico da língua portuguesa nos séculos XX ou XXI – licenciam o padrão ortográfico ⟨#SC⟩ em seus sistemas de escrita. O dicionário Houaiss (HOUAISS, 2009) apresenta entradas de alguns desses estrangeirismos juntamente de sua datação, tais como: *stripper* (d. 1930), *strogonoff* (d. 1932), *spaghetti* (d. 1958), *skate* (d. 1964), *stress* (d. 1975), *spa* (d. 1980). Além de empréstimos, o padrão ilícito também pode ser encontrado em nomes-fantasia de empresas e de marcas relacionadas a produtos e serviços em circulação no território brasileiro cuja fundação se deu nos séculos XX ou XXI, tais como: *Skol* (1967), *Sprite* (1984), *SporTV* (1994), *Sky* (1996), *Skype* (2003).

Nota-se que o padrão ortográfico ⟨#SC⟩, excluído pelos acordos ortográficos em vigência desde o início do século XX, foi reintroduzido nas experiências de língua escrita dos falantes do PB em um outro grupo de palavras, como empréstimos, por exemplo. Portanto, os padrões ortográficos ⟨#ESC⟩ e

⟨#SC⟩ co-ocorrem na escrita do PB atual e demarcam dois conjuntos de palavras distintos que têm percursos históricos particulares. A partir do pressuposto de que formas ortográficas podem fornecer informações pertinentes para se entender o conhecimento fonológico (cf. VELOSO, 2019), faz-se relevante considerar que diferentes padrões ortográficos podem corresponder a diferentes realizações sonoras. Na próxima seção, argumenta-se que aspectos da língua escrita podem refletir e se relacionar a propriedades da língua falada, podendo representar uma evidência pertinente para a investigação de fenômenos fonológicos.

2. Ortografia e representação fonológica em interação

Há evidência de que padrões ortográficos em sistemas alfabéticos tenham impacto na representação fonológica, uma vez que fornecem informação acerca dos sistemas sonoros os quais representam. As técnicas de leitura e de escrita exigem habilidades relacionadas a conhecimentos explícitos da estrutura sonora da fala, de modo que o treinamento metalinguístico específico providenciado pelo percurso da alfabetização tem efeitos no conhecimento fonológico do falante (MORAIS et al. 1979; VELOSO, 2007). “Uma vez consolidado, o sistema ortográfico formata a consciência linguística, e mesmo a fonologia, tanto para os que herdaram o sistema quanto para os que o pegam emprestado para transcrever outra língua” (MATTINGLY, 1992, p. 25, tradução nossa³).

Sobre esse assunto, Levitt (1978) descreve um efeito de “pronúncia ortográfica”, em que certas variantes sonoras se justificariam apenas por sua forma ortográfica. O autor fornece, como exemplo do fenômeno, a realização do som [θ], representado pelo grafema ⟨TH⟩, em palavras como *theatre*, *Catholic*, *throne* e *orthography* do inglês que foram importadas do francês. No francês, o dígrafo ⟨TH⟩ corresponde ao som [t], de modo que a realização [θ] teria motivações ortográficas. Sobre o fenômeno de “pronúncia ortográfica”, Chevrot (1999) considera que o uso de novas variantes sonoras pode ser entendido como um efeito de reinterpretações fonográficas.

Devido à sua aproximação, as relações entre o sistema alfabético e o sistema sonoro podem em primeira análise ser resumidas às correspondências fonema-grafema. Contudo, considerar a existência de uma rota de acesso lexical visual, isto é, ortográfica, para além de uma rota fonológica torna-se um complemento à atuação das regras de conversão fonema-grafema (TAFT, 1982; 1985). Para o indivíduo letrado, as regras de conversão fonema-grafema se tornam mais abstratas, consistindo em um acesso de funcionamento imediato em uma representação amalgamada – fonológica e ortográfica

³ But once it has become established, the writing system itself shapes the linguistic awareness, and even the phonology, both of those who inherit the system and of those who borrow it to transcribe some other language.

–, de onde tanto a pronúncia quanto a grafia de uma palavra são geradas. O aprendizado das técnicas de leitura e de escrita demandam, portanto, tanto uma nova relação do indivíduo com o seu conhecimento linguístico explícito quanto a incorporação de um sistema adicional, que é ortográfico, apresentando padrões ortográficos, que são convencionalizados (TAFT, 1982, 1985).

Com relação a padrões ortográficos, Taft (2006) argumenta que a representação de categorias fonológicas pode ser impactada pela ortografia a partir de resultados de testes de percepção de falantes de variedades não-róticas do inglês. Em variedades não-róticas do inglês, a consoante /r/ em posição pós-vocálica não é pronunciada (ex.: *caught* e *court* = [kɔ:t]). Contudo, os falantes categorizam palavras que apresentam <R> ortográfica (ex.: *court*) como distintas de palavras que não apresentam <R> ortográfica (ex.: *caught*), ainda que a realização fonética dos grupos de palavras seja homófona. Os resultados de Taft (2006) fornecem evidência acerca do papel que padrões ortográficos diferentes podem ter para a categorização de formas sonoras.

Sobre a incorporação de efeitos de ortografia em análises de variação sonora, Purse (2019) apresenta dados do francês sobre a apócope da vogal [ə] e sua interação com formas ortográficas. O autor, ao investigar o fenômeno, aponta um efeito favorecedor dos padrões ortográficos relacionados às palavras investigadas. A apócope de [ə] é favorecida em palavras ortograficamente terminadas por consoante (93%) em detrimento de palavras terminadas por <E> (82%).

Com relação a sequências #(i)sC no português europeu (PE), D'Andrade e Rodrigues (1999) buscam explorar a realização dos padrões ortográficos <#ESC> (*escola*), <#SC> (*stress*), <#ISC> (*isqueiro*) e <#EXC> (*experiência*). Para <#ESC>, são atestadas as realizações no PE como [ʃ]C, [iʃ]C, [eʃ]C ou [əʃ]C. Para as sequências <#SC>, presentes em empréstimos, é atestada a realização [s]C, em que não há palatalização da sibilante. Para as sequências <#ISC>, ocorrem as realizações [ʃ]C ou [iʃ]C. Já as sequências <#EXC> são realizadas no PE como [ʃ]C, [iʃ]C, [eʃ]C, [əʃ]C, [ejʃ]C ou [ejʃ]C, podendo ou não licenciar cancelamento da vogal inicial ou ditongação. Apesar de as possibilidades de realização no PE não serem correspondentes às de sequências #(i)sC no PB, a análise de D'Andrade e Rodrigues (1999) evidencia como diferentes padrões ortográficos podem se relacionar a diferentes realizações de uma sequência segmental similar.

Os impactos de padrões ortográficos na fonologia podem ser observados na adaptação de empréstimos, caso que compreende um dos padrões ortográficos investigados neste estudo. Smith (2006) propõe dois diferentes percursos possíveis para a incorporação de empréstimos no sistema fonológico de uma língua importadora: um baseado na percepção auditiva e o outro na forma ortográfica. A autora sustenta sua proposta a partir de exemplos de empréstimos do inglês adaptados ao japonês. Quando os empréstimos são incorporados via percepção auditiva, mantem-se, por exemplo, o número de sílabas e há certa fidelidade na qualidade da vogal na língua exportadora (ex.: *glycerine* /'glɪs.ər.i:n/ > [ri.su.ri:n]). Por outro lado, quando os empréstimos são incorporados via

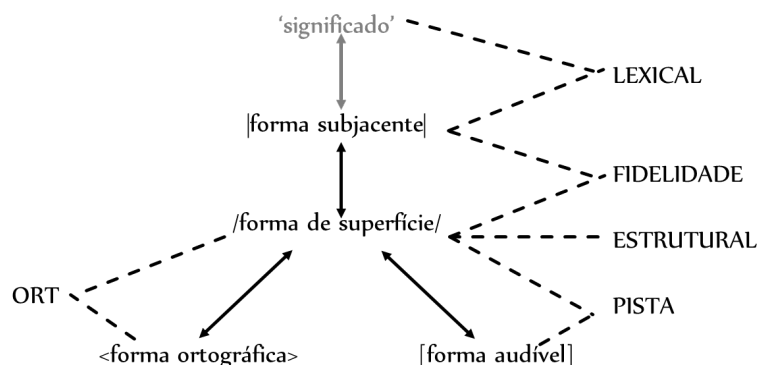
ortografia, são privilegiadas as regras de conversão grafema-fonema do japonês, o que modifica o número de sílabas e a qualidade da vogal em relação à forma na língua exportadora (ex.: *glycerine* > [gu.ri.sɛ.rin]).

Os exemplos fornecidos por Smith (2006) encontram paralelos com empréstimos do inglês adaptados ao PB. Exemplos de incorporação via percepção auditiva são palavras como *cowboy* ([ˈkaʊ.bɔɪ] > [kaʊˈbɔɪ]) e *cupcake* ([ˈkʌp.keɪk] > [kʌ.p(i)ˈkeɪ.k(i)]). Há também exemplos de empréstimos de incorporação via ortografia, em palavras como *ketchup* (ˈkɛtʃ.ʌp/ > [kɛ.tʃ(i)ˈʃu.p(i)]) e *remix* ([ˈɹiː.mɪks] > [hɛˈmi.k(i)s]). É possível, adicionalmente, encontrar exemplos em que as duas rotas de adaptação se aplicam num mesmo vocábulo, como em *mouse* (ˈmaʊs/ > [ˈmaʊ.z(i)]) e *laser* ([ˈleɪ.zə] > ˈleɪ.zeh). À vista dos exemplos fornecidos, a incorporação auditiva não é a única força atuante na adaptação de empréstimos, sendo que a forma ortográfica e sua interação com restrições fonotáticas da língua importadora devem ser consideradas na adaptação. Considerando-se a inserção dos falantes do PB em uma cultura escrita e majoritariamente monolíngue, é imprescindível contemplar os efeitos das formas ortográficas dos empréstimos em sua representação fonológica. Muitas vezes, a propagação de empréstimos envolve exclusivamente seu uso escrito, em que a informação fonológica da língua exportadora não se encontra disponível. Ademais, há evidência adicional de que certos empréstimos possam ser incorporados no léxico em estratos diferentes de palavras nativas (LOVINS, 1975, ITŌ; MESTER, 1995, 1999). A classificação de empréstimos em estratos lexicais diferentes de palavras nativas pode ser refletida em sua estrutura fonológica, que apresentaria restrições excepcionais em relação àquelas que regulam palavras nativas.

A partir de exemplos como os apresentados anteriormente, entende-se que a forma ortográfica e sua interação com restrições fonotáticas da língua importadora devem ser consideradas na adaptação de empréstimos (HAMANN; COLOMBO, 2017). Os efeitos de formas ortográficas na adaptação de empréstimos são formalizados por Hamann e Colombo (2017) a partir de evidências de empréstimos do inglês incorporados ao italiano. No italiano, consoantes geminadas são ortograficamente representadas por grafemas em que uma letra é duplicada (ex.: *fatto* /ˈfat.to/ vs. *fato* /ˈfa.to/). Já no inglês, letras duplicadas correspondentes a consoantes não se referem necessariamente a casos de geminação (ex.: *hobby* /ˈhʌb.i/). O padrão ortográfico de consoantes duplicadas em empréstimos do inglês é pronunciado como consoantes geminadas no italiano (ex.: *hobby* [ˈɔb.bi]). Em casos de empréstimos em que não há duplicação ortográfica de consoantes iguais, não ocorre a pronúncia como consoantes geminadas (ex.: *hacker*: [ˈa.ker] e não [ˈak.ker*]). Assim como os casos previamente apresentados para o japonês e para o PB, padrões emprestados são adaptados fonologicamente pela língua importadora, em que a forma ortográfica atua de modo determinante. Sobre o efeito da ortografia na adaptação, Hamann (2018) afirma que formas ortográficas podem influenciar categorias fonológicas sem que haja correlatos perceptuais. Com relação à interação entre ortografia e percepção

no mapeamento fonológico, Hamann e Colombo (2017) propõem uma formalização da interação entre percepção e ortografia para o reconhecimento lexical e produção de fala.

Figura 1 – Modelo de interação entre ortografia e fonologia



Fonte: adaptado de Hamann e Colombo (2017).

A formalização proposta prevê interação entre restrições fonológicas – pautadas na percepção auditiva – e ortográficas – pautadas no mapeamento do sistema de escrita – na representação fonológica. Hamann (2018) sugere que não apenas a ortografia tenha influência na representação fonológica, como as representações ortográfica e fonológica sejam simultaneamente ativadas durante a produção e a percepção de fala.

Sobre a inter-relação entre representações ortográfica e fonológica, a perspectiva da Integração de Múltiplos Padrões (IMP) (TREIMAN; KESSLER, 2014; TREIMAN, 2017) objetiva explicar como ocorre a aquisição de sistemas de escrita ortográficos. A perspectiva da IMP interpreta que falantes letrados armazenam informações específicas sobre a ortografia das palavras em sua memória a partir de padrões probabilísticos e determinísticos, que são generalizados e se replicam através do léxico. Os falantes associariam palavras que apresentam os mesmos padrões ortográficos.

Conforme a perspectiva da IMP, os padrões ortográficos primeiramente se relacionam às representações visuais possíveis pela ortografia do sistema de escrita. Por exemplo, em relação ao traçado das letras do sistema de escrita, ⟨a⟩ é um padrão licenciado no sistema de escrita da língua portuguesa, enquanto ⟨λ⟩ ou ⟨ä⟩ não o são. Já em relação à combinação e organização de símbolos, padrões de combinação como ⟨Seis⟩, ⟨mãe⟩ e ⟨cou⟩ são possíveis no português, à medida que combinações como ⟨Çeis⟩, ⟨mmm⟩ e ⟨cOu⟩ não o são. Há também relações entre os padrões ortográficos e aspectos linguísticos. As unidades linguísticas relacionadas aos padrões ortográficos podem ter base fonológica ou morfológica. Por exemplo, há uma relação grafofonêmica estabelecida entre ⟨CA⟩ e /ka/ ou entre ⟨CE⟩ e /se/, assim como há uma relação morfológica estabelecida entre ⟨s⟩ e

palavras como *seis*, *sexto*, *sessenta*. Nessa perspectiva, o padrão ortográfico ⟨#SC⟩, ilícito ao sistema de escrita da língua portuguesa, estaria relacionado a um grupo específico de empréstimos, agrupados em um grupo de palavras distinto daquelas que apresentam o padrão ortográfico ⟨#ESC⟩.

Nesta seção, foram exploradas evidências sobre a relação entre a aspectos fonéticos/fonológicos e as possibilidades ortográficas dos sistemas de escrita. Considera-se que incorporar os impactos que padrões ortográficos podem ter na implementação de fenômenos fonéticos pode fornecer contribuições a sua compreensão (cf. VELOSO, 2019). Padrões ortográficos que, em primeira análise, aparentemente correspondem a uma mesma sequência sonora podem se relacionar a diferentes categorias fonológicas. No caso de sequências #*(i)*sC, é possível considerar que os padrões ortográficos correspondentes ⟨#ESC⟩ e ⟨#SC⟩ caracterizam grupos diferentes de palavras, que se realizam foneticamente de modo distinto. A próxima seção explora a representação e a realização de sequências #*(i)*sC no PB.

3. Sequências #*(i)*sC no português brasileiro

A sequência sonora #*(i)*sC analisada neste estudo envolve a alternância zero-vogal, que pode ou não se realizar no PB (BISOL, 1999; COLLISCHONN, 2000; COLLISCHONN; SCHWINDT, 2005; FREITAS, 2019; CRISTÓFARO SILVA; FREITAS, 2020)⁴. Sequências #*(i)*sC no PB apresentam um problema de análise fonológica, em especial no que diz respeito à presença de uma vogal inicial subjacente ou não. Os processos de derivação morfológica no PB (cf. COLLISCHONN; SCHWINDT, 2005) e em outras línguas (cf. HARRIS, 1986; KAYE, 1992; POLGÁRDI, 2017) fornecem evidências para assumir que a vogal, quando manifesta, é resultado derivado de um processo de epêntese. É assumido, para análises sobre sequências #*(i)*sC no PB, que o processo de epêntese estaria ativo, tanto em palavras nativas, que apresentam o padrão ⟨#ESC⟩, como em empréstimos, que apresentam o padrão ⟨#SC⟩ (BISOL, 1999; COLLISCHONN; SCHWINDT, 2005). Dessa forma, palavras que apresentam os padrões ortográficos ⟨#ESC⟩ e ⟨#SC⟩ compreenderiam uma mesma sequência sonora.

O fenômeno de epêntese em sequências #*(i)*sC pode ser caracterizado como uma alternância que envolve uma vogal alta átona em adjacência a uma sibilante, encontrando casos análogos de redução e de queda no PB, que têm implementação gradiente. Sobre a redução vocálica em adjacência a sibilantes, Meneses e Albano (2015) tratam da queda da vogal átona final, em palavras como *passe*

⁴ Nestes trabalhos, o estatuto representacional de sequências #*(i)*sC no PB é explorado de forma detalhada, relacionando o fenômeno a sequências iniciais formadas por sibilante e consoante em outras línguas.

[pasi] ~ [pas]. São identificadas pistas gestuais da vogal no sinal acústico referente à sibilante, sugerindo um efeito de sobreposição de gestos articulatorios da vogal com a sibilante adjacente, em busca de contemplar a gradiência fonética atestada na investigação do fenômeno. A partir das evidências apresentadas pela análise de Meneses e Albano (2015), interpreta-se que fenômenos de cancelamento de vogal operam gradientemente.

Sob essa perspectiva, assume-se que o detalhe fonético (neste caso, caracterizado pela gradiência fonética na alternância zero-vogal em sequências #(i)sC) tenha impacto na organização do conhecimento e da representação fonológicos. Desse modo, o modelo teórico de representação adotado na presente investigação é a Teoria de Exemplos (JOHNSON, 1997, 2007; PIERREHUMBERT, 2001, 2003; BYBEE, 2001, 2006, 2010). Sob os pressupostos da Teoria de Exemplos, o conhecimento linguístico se organiza probabilisticamente a partir da experiência do falante com instâncias de uso – de produção e de percepção. As formas experienciadas seriam continuamente representadas como exemplos à medida que palavras, ou construções mais complexas, são usadas. O armazenamento e a categorização dos exemplos levariam à emergência de uma representação cognitiva: a gramática. As representações mentais, abstratas, não estariam dissociadas das formas generalizadas, que emergem do uso. A gramática é, sob essa perspectiva, fortemente atrelada à experiência de uso da língua pelo falante, de modo que o detalhe fonético é incorporado à representação (BYBEE, 2006).

Organizados por similaridade fonética e semântica, os exemplos seriam mapeados em uma rede de relações, agrupando-se em nuvens (PIERREHUMBERT, 2001, 2003). Exemplos similares que são mais frequentemente experienciados se fortalecem, formando representações mais robustas do que exemplos menos usados, que se enfraquecem e são esquecidos (PIERREHUMBERT, 2003; TODD; PIERREHUMBERT; HAY, 2019). Em outras palavras, exemplos que representam experiências recorrentes e recentes são recorrentemente ativados, tendo impacto na categorização de novas instâncias (PIERREHUMBERT, 2001). A ocorrência frequente de vogais reduzidas, em adjacência a sibilantes, no PB teria um impacto em sua representação fonológica. Exemplos incorporam a informação fonética detalhada de formas reduzidas, expressas por menores valores duracionais. Considerando-se o impacto que o detalhe fonético teria nas representações, o uso das variantes reduzidas retroalimentaria a ativação de exemplos na memória.

Os pressupostos da Teoria de Exemplos fornecem subsídios explicativos para a investigação da implementação de fenômenos de variação sonora que se manifestam de forma gradiente, como a redução vocálica em sequências #(i)sC. Embora aparentemente apresentem a mesma sequência segmental, é esperado que diferentes categorias fonológicas, evidenciadas por seu padrão ortográfico, apresentem diferenças no detalhe fonético de seus exemplos. O detalhe fonético, investigado experimentalmente nesta proposta, é caracterizado pela duração da vogal inicial. Como hipótese para

este estudo, foi postulado que diferentes padrões ortográficos podem se associar a valores duracionais diferentes da vogal, caracterizando diferenças no detalhe fonético das sequências #(i)sC. É previsto que sequências correspondentes ao padrão ortográfico ⟨#SC⟩ apresentem taxas de ocorrência da vogal e valores duracionais menores do que sequências correspondentes ao padrão ortográfico ⟨#ESC⟩. O padrão ortográfico ⟨#SC⟩ é descrito como um caso que deve necessariamente ser corrigido por uma epêntese de uma vogal inicial no PB (cf. COLLISCHONN, 2000). Ademais, há evidência de que vogais epentéticas apresentam menor taxa de ocorrência e menor duração do que vogais não epentéticas no PB (CRISTÓFARO SILVA; ALMEIDA, 2008; CANTONI, 2015).

Para que se teste a hipótese postulada, este estudo se vale de uma investigação experimental para motivar a análise e discussão dos resultados em consonância com os pressupostos teórico-metodológicos da Fonologia de Laboratório (PIERREHUMBERT; BECKMAN; LADD, 2011). O experimento desenvolvido consistiu em uma tarefa de nomeação de figuras com objetivo de examinar acusticamente a sequência segmental em sequências #(i)sC. Na próxima seção, é apresentado o procedimento metodológico para coleta e análise dos dados.

4. Metodologia

Os estímulos utilizados na tarefa de nomeação de figuras adotada correspondem a 30 palavras. Os estímulos foram igualmente divididos entre cada padrão ortográfico: ⟨#ESC⟩ ou ⟨#SC⟩, consistindo em 15 estímulos para cada padrão. Para cada padrão ortográfico, os estímulos foram igualmente distribuídos entre cada uma das oclusivas referentes à consoante ⟨C⟩: [p], [t] ou [k]⁵. Durante a seleção dos estímulos, as 15 palavras que apresentam o padrão ortográfico ⟨#SC⟩ foram primeiramente selecionadas, tendo sua aceção registrada no Dicionário Houaiss (HOUAISS, 2009) ou correspondendo a um produto ou nome-fantasia de circulação no território brasileiro. As 15 palavras que apresentam o padrão ortográfico ⟨#ESC⟩ foram selecionadas por similaridade fonológica ao conjunto de palavras ⟨#SC⟩, no que diz respeito ao número de sílabas, configuração silábica e incidência de acento. No Quadro 1, estão apresentados os estímulos utilizados para a coleta dos dados.

⁵ Outras consoantes não foram consideradas para o desenho experimental em razão de sua escassez no léxico da língua portuguesa, impossibilitando a elaboração de uma amostra tanto nos casos de empréstimos como de palavras nativas.

Quadro 1 – Estímulos utilizados para a coleta dos dados

| (<#ESC>) | | | (<#SC>) | | |
|-------------------|------------------|------------------|------------------|-------------------|----------------|
| [p] | [t] | [k] | [p] | [t] | [k] |
| <i>espaço</i> | <i>estádio</i> | <i>escada</i> | <i>spa</i> | <i>status</i> | <i>skate</i> |
| <i>espada</i> | <i>estojo</i> | <i>escama</i> | <i>spaghetti</i> | <i>stop</i> | <i>sky</i> |
| <i>espetáculo</i> | <i>estrada</i> | <i>escola</i> | <i>SporTV</i> | <i>stress</i> | <i>Skype</i> |
| <i>espiga</i> | <i>estrela</i> | <i>escravo</i> | <i>spray</i> | <i>stripper</i> | <i>Skol</i> |
| <i>esponja</i> | <i>estudante</i> | <i>esqueleto</i> | <i>Sprite</i> | <i>strogonoff</i> | <i>squeeze</i> |

Os 30 estímulos experimentais foram apresentados em conjunto com 35 estímulos distratores. Com relação à sequência de apresentação dos estímulos, 5 estímulos distratores fixos foram apresentados ao início da tarefa com objetivo de familiarização dos participantes com o experimento. Os 60 estímulos restantes (30 experimentais e 30 distratores) foram aleatorizados para cada gravação com auxílio o macro `sort_rand` do software *Microsoft PowerPoint 2016*.

O experimento contou com a contribuição de 24 participantes nascidos e residentes na região metropolitana de Belo Horizonte, MG, sem desvios ou queixas de audição e/ou de fala relatados. Com relação à escolaridade, todos os participantes possuíam curso superior completo ou em andamento durante a realização do experimento. Com relação à idade, a faixa etária dos participantes compreende o intervalo entre 18 e 28 anos. Com relação ao sexo, o conjunto de 24 participantes se divide igualmente por sexo, sendo 12 mulheres e 12 homens.

Os participantes foram instruídos a dizer em voz alta qual a ilustração com que se deparassem na tela de um computador portátil. As gravações de áudio das produções dos falantes foram obtidas com auxílio de um gravador Zoom H4N (com taxa de amostragem de 44,1 KHz, quantização de 16 bits, em arquivos de formato `.wav`). Cada ilustração foi apresentada em slides do software *Microsoft PowerPoint* que se mantinham automaticamente em tela por 2,5 segundos. Os estímulos escritos foram apresentados em conjunto com a figura em fonte Arial, tamanho 88 pt. Cada item experimental foi apresentado em duas modalidades de apresentação durante a tarefa, uma após a outra. Na primeira modalidade, os slides exibiram a ilustração em conjunto com a forma escrita. Na segunda modalidade, as ilustrações foram exibidas em conjunto com a informação: “Diga novamente, em voz alta, a palavra que você acabou de dizer.”⁶.

⁶ Os resultados para os efeitos da modalidade de produção dos dados para este experimento são explorados na dissertação de mestrado de Freitas (2019).

Figura 2 – Exemplo de slides utilizados para a coleta dos dados



Os dados foram coletados em uma cabine de gravação com isolamento acústico no CEFALA (Centro de Estudos da Fala, Acústica, Linguagem e música da Universidade Federal de Minas Gerais) em julho de 2018. Cada gravação teve duração aproximada de 15 minutos para cada participante. Das produções realizadas pelos participantes, 1.440 amostras compõem a base de dados deste estudo. As amostras foram editadas e anotadas com auxílio do recurso TexGrid do software *Praat* (BOERSMA; WEENICK, 2015).

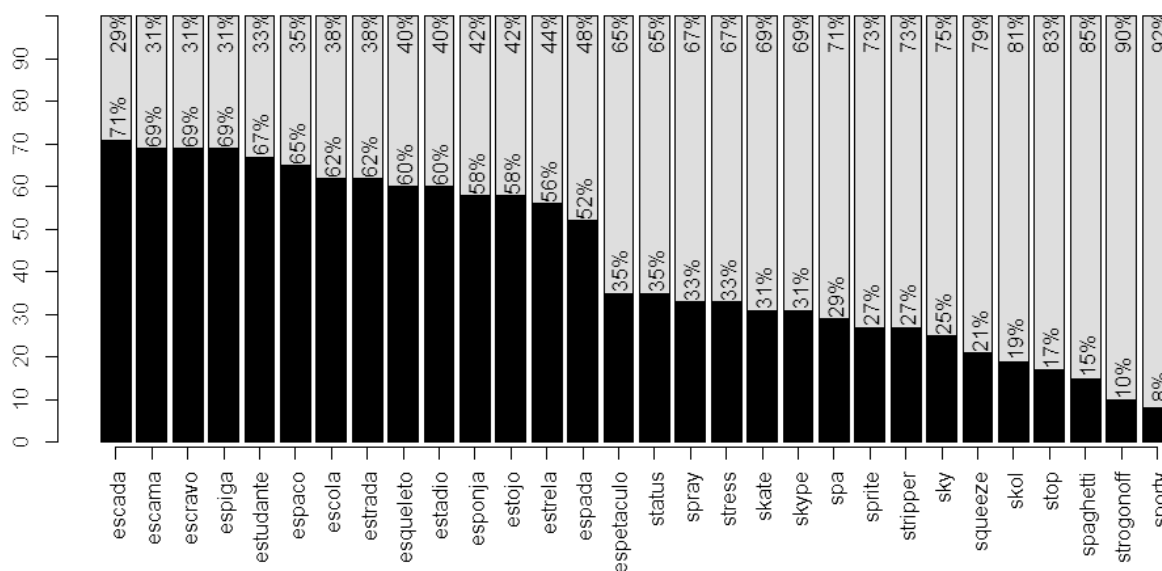
Os resultados sobre a ocorrência a duração de uma vogal inicial nas sequências #(i)sC foram extraídos por um script adaptado de Lennes (2002). Na análise acústica, como parâmetro para determinar a existência e a duração de uma vogal nas sequências #(i)sC foi considerada identificação de qualquer intervalo periódico da onda sonora no início da palavra. Quando atestada alguma periodicidade na onda sonora no início da palavra, a duração do intervalo foi mensurada. Em casos em que não se atestou a periodicidade da onda sonora, foi considerado que não houve realização de uma vogal.

A análise e modelagem gráfica e estatística foi realizada por meio da plataforma R (R Core Team, 2017), utilizando-se modelos de efeitos mistos (BAAYEN; DAVIDSON; BATES, 2008) – funcionalidade disponível no pacote lme4 (BATES et al., 2015). Para a análise de taxas de ocorrência da vogal, modelos de regressão logística foram empregados, em que os efeitos fixos e aleatórios dos seguintes fatores foram considerados: o padrão ortográfico, o item lexical, o falante e o sexo do falante. Para a análise da duração da vogal, modelos de regressão linear foram empregados, em que os efeitos fixos e aleatórios dos seguintes fatores foram considerados: o padrão ortográfico, o item lexical, o falante, o sexo do falante e a duração da palavra. Os modelos foram comparados em um teste de razão de verossimilhança que respeita a distribuição qui-quadrado. Para todos os testes estatísticos, a tolerância de confiança adotada foi de 95%, ou de $p < 0.05$.

5. Resultados e discussão

A primeira análise feita consistiu na observação sobre a presença ou ausência da vogal inicial em seqüências #(i)sC. Como resultado, foi obtido de que a vogal foi produzida em 42,6% dos dados. Os resultados indicam que há alternância entre as formas #isC e #sC no PB atual falado em Belo Horizonte, caracterizando o fenômeno como uma alternância variável. Em seguida, as taxas de ocorrência da vogal inicial foram consideradas para cada padrão ortográfico investigado: <#ESC> ou <#SC>. O Gráfico 1 ilustra a ocorrência da vogal inicial para cada palavra analisada.

Gráfico 1 – Ocorrência da vogal inicial em seqüências #(i)sC por palavra



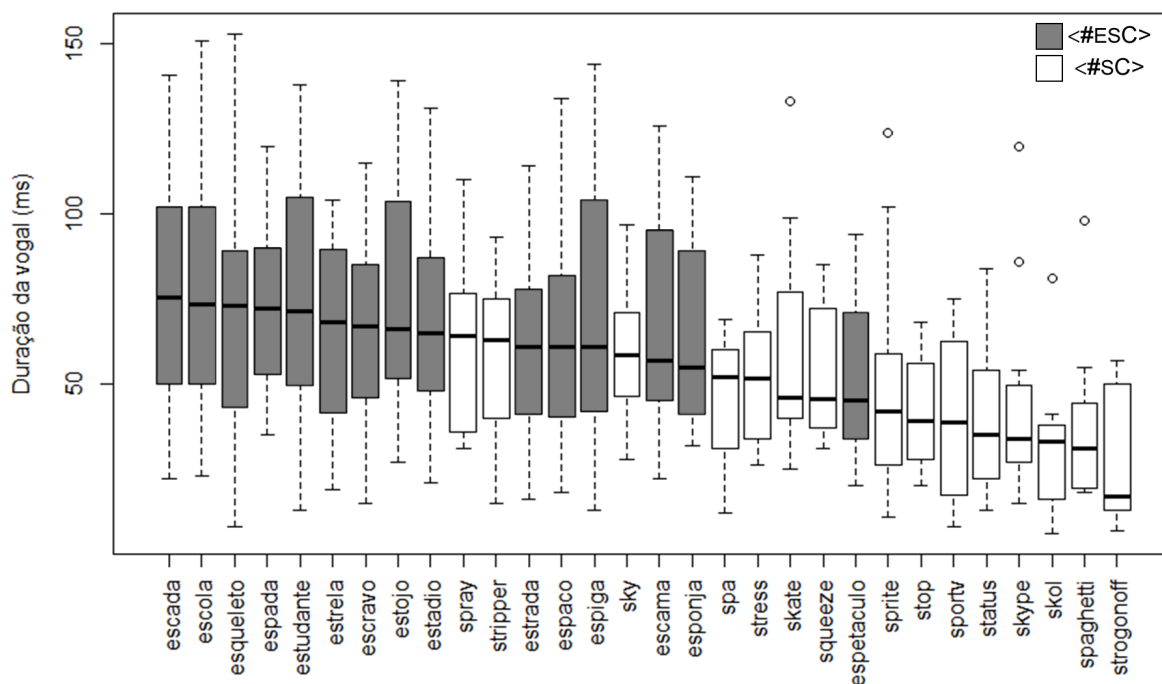
No Gráfico 1, estão distribuídas as taxas de ocorrência da vogal inicial em seqüências #(i)sC para cada palavra analisada em ordem decrescente. Os resultados indicam tanto que há alta variabilidade na amostra investigada quanto à ocorrência da vogal inicial em seqüências #(i)sC (71%~8%) como que há uma tendência de que a vogal inicial seja produzida em palavras que apresentam o padrão ortográfico <#ESC> em comparação ao padrão ortográfico <#SC>. Os resultados para cada padrão ortográfico apontam para uma ocorrência da vogal inicial em palavras com o padrão <#ESC> de 61%, enquanto para o padrão <#SC> a taxa de ocorrência da vogal inicial é de 24,2%. Conforme a análise de variância entre os modelos de regressão adotados na análise estatística, o padrão ortográfico é um fator significativo na ocorrência da vogal inicial em seqüências #(i)sC ($\chi^2 = 56,835$, $p < 0,001$) na amostra analisada.

Em consonância com a Teoria de Exemplos, é possível compreender as motivações para as diferenças encontradas entre os dois grupos de palavras. É sugerido que exemplos com a vogal inicial

são mais robustos para o padrão <#ESC> do que para o padrão <#SC>. Sendo assim, o padrão sonoro inovador, em que há ausência da vogal inicial em sequências #(i)sC, implementa-se de modo mais avançado em palavras <#SC>. Assume-se, a partir de tais resultados, que os dois conjuntos de palavras se agrupem separadamente, o que encontra respaldo na interpretação de que padrões ortográficos se relacionam a diferentes propriedades linguísticas, como sugerido pela IMP. Cada padrão ortográfico, <#ESC> ou <#SC>, caracteriza um grupo distinto de palavras, nativas e empréstimos, respectivamente, que apresentam como possibilidades as realizações #isC ou #sC inicial. Contudo, as realizações se realizam em diferentes taxas de ocorrência da vogal, o que sugere que cada grupo de palavras seja mapeado de modo distinto.

Ademais, uma vez que este trabalho assume a incorporação do detalhe fonético às representações, conforme os pressupostos da Teoria de Exemplos, torna-se relevante a investigação de propriedades detalhadas na investigação de sequências #(i)sC. Há evidências para que a queda da vogal inicial em sequências #(i)sC seja entendida como uma reconfiguração gradiente da vogal e da sibilante, estabelecendo um contraponto a uma interpretação de cancelamento categórico da vogal. Para tanto, a duração da vogal foi avaliada experimentalmente, de modo a testar a hipótese de que os valores duracionais se distinguem em sequências #(i)sC a depender do seu padrão ortográfico. Tal hipótese é fundamentada a partir de evidências de que vogais epentéticas apresentam duração menor do que vogais não epentéticas (cf. Seção 3). O Gráfico 2 ilustra a duração da vogal inicial para cada palavra analisada.

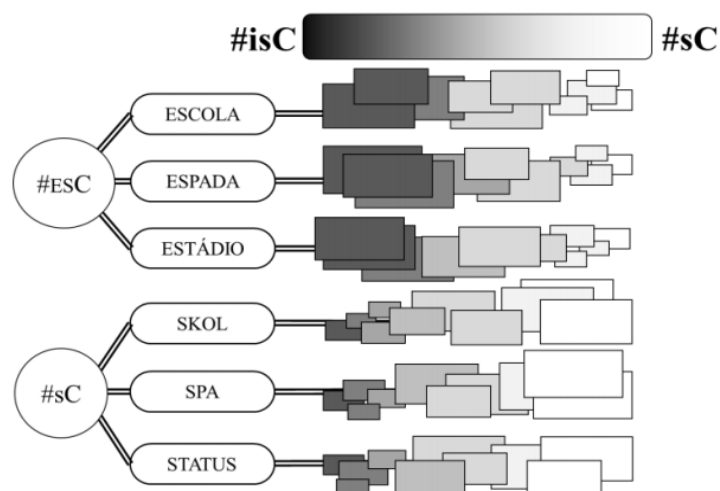
Gráfico 2 – Duração da vogal inicial em sequências #isC por palavra



Em 42,6% dos dados analisados, uma vogal inicial foi produzida. Para essa parcela dos dados, foi mensurada a duração da vogal inicial. No Gráfico 2, estão distribuídos os valores duracionais da vogal [i] inicial em sequências #isC para cada palavra ordenados pela mediana de cada amostra. Palavras que apresentam o padrão ortográfico (#ESC) estão representadas em cor cinza, e palavras que apresentam o padrão ortográfico (#sC) estão representadas em cor branca. Assim como nos resultados sobre a ocorrência da vogal inicial, é notável a distribuição específica das palavras conforme seu padrão ortográfico, (#ESC) ou (#sC). Os resultados para cada padrão ortográfico apontam para uma correspondência de palavras com o padrão (#ESC) a valores maiores de duração da vogal inicial (mediana = 63 ms, média = 68,5 ms (d = 31,1 ms)), enquanto palavras com o padrão (#sC) correspondem a valores menores de duração da vogal inicial (mediana = 46 ms, média = 49 ms, (d = 25 ms)). Conforme a análise de variância entre os modelos de regressão adotados na análise estatística, o padrão ortográfico é um fator significativo para a duração da vogal inicial em sequências #isC ($\chi^2 = 38,341$, $p < 0,001$) na amostra analisada.

De acordo com a Teoria de Exemplos, o detalhe fonético tem impacto nas representações. No presente caso, o detalhe fonético caracterizado e avaliado pela duração da vogal inicial. Sugere-se que as diferenças nos valores duracionais da vogal inicial correspondam à sua categorização em conjuntos de palavras distintos, de modo correspondente à análise da taxa de ocorrência da vogal inicial (cf. Gráfico 1). O padrão ortográfico (#ESC), relacionado a maiores taxas de ocorrência da vogal inicial, também apresenta maior duração da vogal inicial quando ela é produzida, em comparação ao padrão (#sC). Os resultados obtidos demonstram que há diferenças no detalhe fonético para palavras de cada padrão ortográfico na produção de sequências #(i)sC. Frente aos resultados obtidos, levanta-se uma proposta de mapeamento fonológico de sequências #(i)sC considerando-se os efeitos dos padrões ortográficos ilustrada na Figura 3.

Figura 3 – Mapeamento fonológico de sequências #(i)sC



Conforme os pressupostos da IMP, é prevista generalização de padrões ortográficos a partir da experiência do falante com formas ortográficas. Dessa maneira, as palavras analisadas neste trabalho seriam categorizadas em dois estratos lexicais representados pelos padrões ortográficos <#ESC> e <#sC>. Para cada grupo, são encontradas diferenças no detalhe fonético das formas produzidas de sequências #(i)sC. A escala de cinza na imagem representa a ocorrência da vogal inicial. Os exemplares são representados por retângulos que têm seu tamanho relacionado à sua robustez de sua representação. A alternância da ocorrência da vogal inicial em sequências #(i)sC ocorre em ambos os grupos de palavras. Contudo, para palavras com o padrão <#ESC>, exemplares mais robustos se relacionam ao favorecimento da ocorrência e à maior duração da vogal inicial, enquanto, para palavras com o padrão <#sC>, exemplares mais robustos se relacionam ao desfavorecimento da ocorrência e à menor duração da vogal inicial. A interação entre as informações sobre os padrões ortográficos e o mapeamento fonológico das palavras é prevista pelo pressuposto teórico de que exemplares consistem em instâncias de experiência detalhada que agrupam e relacionam informações diversas, em que podem se incluir a forma ortográfica das palavras (JOHNSON, 2007; BYBEE, 2006).

Com relação ao estatuto representacional de sequências #(i)sC no PB, os resultados indicam que palavras nativas, grafadas com a sequência <#ESC>, e empréstimos, grafados com a sequência <#sC>, se realizam de modo distinto, confirmando as hipóteses levantadas. Embora sejam interpretados como sequências segmentais idênticas, cada grupo de palavras apresenta diferenças fonéticas caracterizadas pelas taxas de ocorrência da vogal e por sua duração. Os resultados indicam que diferentes padrões ortográficos, que têm trajetórias históricas específicas no léxico (cf. Seção 1), podem ser evidência crucial para caracterizar diferentes grupos de palavras, inclusive por meio de propriedades detalhadas (cf. Seção 3). A presente interpretação para os resultados está alinhada com interpretações e pressupostos teóricos que assumem que aspectos visíveis na língua escrita podem estar diretamente relacionados a propriedades da língua falada (cf. Seção 2).

Considerações finais

A investigação abordada neste artigo tratou de sequências #(i)sC no PB, isto é, sequências formadas por uma sibilante e uma consoante, que podem ou não ser precedidas por uma vogal em início de palavra. Buscou-se avaliar o fenômeno tendo em vista como ele se implementa em dois grupos de palavras, grafadas com os padrões ortográficos <#ESC> e <#sC>. Foi atestado que a ocorrência da vogal é um fenômeno variável em sequências #(i)sC no PB, o que foi evidenciado pela taxa de ocorrência da vogal inicial em 42,6% dos dados analisados.

Tendo em consideração os efeitos dos padrões ortográficos <#ESC> e <#SC> para a implementação do fenômeno, pretendeu-se evidenciar que formas ortográficas podem fornecer contribuições para a compreensão de fenômenos fonológicos. Os resultados indicam que há diferenças nas realizações de sequências #(i)sC para cada padrão ortográfico na amostra analisada. A taxa de ocorrência da vogal inicial em sequências #(i)sC é favorecida em palavras que apresentam o padrão <#ESC> (61%) em detrimento daquelas que apresentam o padrão <#SC> (24,2%).

Com relação à duração da vogal inicial, quando produzida, foi atestado também um efeito significativo do padrão ortográfico, indicando diferenças no detalhe fonético na produção de cada grupo de palavras. Palavras que contêm o padrão <#ESC> apresentaram vogais com maior duração do que palavras que apresentam o padrão <#SC>. Foram encontradas evidências para considerar o papel do detalhe fonético na caracterização de categorias fonológicas, que se relacionam a diferentes padrões ortográficos.

Em suma, este trabalho explorou os aspectos a seguir. Em primeiro lugar, foi feita uma revisão do percurso histórico dos padrões ortográficos <#ESC> e <#SC> na língua portuguesa. Foi apresentado que cada padrão se relaciona historicamente a diferentes grupos de palavras no léxico do PB. Assim sendo, embora coocorram e correspondam a uma mesma sequência segmental, é possível associar cada padrão ortográfico atualmente a grupos específicos de palavras: palavras nativas ou empréstimos.

Em seguida, foram apresentadas evidências de que os padrões convencionalizados pela ortografia podem ser uma evidência contributiva para compreender fenômenos fonológicos. Esse posicionamento é respaldado pelo modelo teórico da IMP, que assume que os padrões ortográficos são adquiridos durante a alfabetização, sendo associados a informações linguísticas.

Por conseguinte, o estatuto de #(i)sC no PB quanto à sua representação e, em especial, quanto à sua realização foi explorado. Foram apresentadas evidências de que fenômenos de redução vocálica se implementam de modo gradiente, podendo ser caracterizados por propriedades fonéticas detalhadas, como considerado no modelo de representação da Teoria de Exemplares. Embora sequências #(i)sC sejam analisadas como a mesma sequência segmental para os padrões ortográficos <#ESC> e <#SC>, buscou-se avaliar como diferenças no detalhe fonético podem caracterizar os dois grupos de palavras de modo distinto.

Os resultados alcançados se alinham a premissas de modelos como a IMP e a Teoria de Exemplares, que assumem representações detalhadas e integradas a diversos tipos de informações, emergindo a partir das experiências dos falantes com eventos linguísticos de escrita e de fala, respectivamente. Buscou-se contribuir para a análise de sequências #(i)sC no PB considerando-se o impacto que padrões ortográficos com diferentes trajetórias históricas pode ter na caracterização de sequências #(i)sC no PB, relacionando-se a diferentes conjuntos de palavras.

Referências bibliográficas

- BAAYEN, R.; DAVIDSON, D.; BATES, M. Mixed-effects modeling with crossed random effects for subjects and items. **Journal of Memory and Language**, n. 59, p. 390–412, 2008.
- BATES, D.; MÄCHLER, M.; BOLKER, B.; WALKER, S. Fitting Linear Mixed-Effects Models Using lme4. **Journal of Statistical Software**, v. 67, n. 1, p. 1–48, 2015.
- BISOL, Leda. A sílaba e seus constituintes. In: NEVES, M. H. M. (org). **Gramática do Português Falado**, v. VII. Campinas: Editora da Unicamp, 1999. p. 701–739.
- BOERSMA, P.; WEENINK, D. **Praat: doing phonetics by computer** [Programa de computador]. Versão 5.4.08, de 24 de março de 2015. Disponível em: (<http://www.praat.org/>)
- BYBEE, J. **Phonology and language use**. Cambridge: Cambridge University Press, 2001.
- BYBEE, J. From usage to grammar: the mind's response to repetition. **Language**, v. 82, n. 4, p. 711–733, 2006.
- BYBEE, J. **Language, Usage and Cognition**. Cambridge: Cambridge University Press, 2010.
- CANTONI, M. M. A epêntese no português brasileiro em uma perspectiva multirrepresentacional. **Gragoatá**, n. 38, p. 231–246, 2015.
- CHEVROT, J. L'effet Buben: de la linguistique diachronique à l'approche cognitive (et retour). **Langue française**, n. 129, p. 104–125, 1999.
- COLLISCHONN, G. A epêntese vocálica no português do sul do Brasil: análise variacionista e tratamento pela Teoria da Otimalidade. **Letras de Hoje**, v. 35, n. 1, p. 285–318, 2000.
- COLLISCHONN, G.; SCHWINDT, L. C. Considerações sobre a sequência /sC/ inicial em Português Brasileiro. **Revista Lingua(gem)**, Macapá: ILAPEC, v. 3, p. 249–266, 2005.
- CRISTÓFARO-SILVA, T.; ALMEIDA, L. S. On the nature of epenthetic vowels. In: BISOL, L.; BRESCANCINI, C. (org.). **Contemporary Phonology in Brazil**, 1a. ed. Cambridge: Cambridge University Press – Cambridge Scholars Publishing Series, 2008. p. 193–212.
- CRISTÓFARO SILVA, T.; FREITAS, M. sC-clusters in Brazilian Portuguese. **Journal of Portuguese Linguistics**, v. 19, n. 1, p. 1–14, 2020. DOI: <http://doi.org/10.5334/jpl.231>
- D'ANDRADE, E.; RODRIGUES, M. C. Das Escolas e das Culturas: História de uma sequência consonântica. In: **Actas do XIV Encontro Nacional da Associação Portuguesa de Linguística**. Lisboa: APL/Colibri, 1999. p. 117–133.
- FREITAS, M. **A redução segmental em sequências #(i)sC no português brasileiro**. 2019. Dissertação (Mestrado em Estudos Linguísticos) – Universidade Federal de Minas Gerais, Belo Horizonte.
- GONÇALVES, M. F. A Ortografia Nacional (1904) de Gonçalves Viana e as ideias ortográficas dos reformistas sul-americanos. **Eutomia**, v. 1, n. 6, p. 1–17, 2010.

- GONÇALVES VIANA, A. R. **Ortografia Nacional**. Simplificação e uniformização das ortografias portuguesas. Lisboa: Livraria Editora Viúva Tavares Cardoso, 1904.
- HAMANN, S. Ghost phonemes in second languages: How orthography can create contrasts without perceptual correlates. In: **15th Old World Conference on Phonology**. Londres: University College London, 2018. (Comunicação oral)
- HAMANN, S.; COLOMBO, I. A formal of the interaction of orthography and perception: English intervocalic consonants borrowed into Italian. **Natural Language & Linguistic Theory**, n. 35, p. 683–714, 2017.
- HARRIS, J. Epenthesis Processes in Spanish. In: NEIDLE, C.; CEDEÑO, R. A. N. (eds.). **Studies in Romance Languages**, v. 25. Foris: Dordrecht, 1986. p. 107–122.
- HOUAISS, A.; VILLAR, M. S. **Dicionário Houaiss de Língua Portuguesa**. Elaborado pelo Instituto Antônio Houaiss de Lexicografia e Banco de Dados da Língua Portuguesa S/C Ltda. Rio de Janeiro: Objetiva, 2009.
- ITŌ, J.; MESTER, A. Japanese Phonology. In: GOLDSMITH, J. (ed). **The Handbook of Phonological Theory**. Blackwell, 1995. p. 817–838.
- ITŌ, J.; MESTER, A. The structure of the phonological lexicon. In.: TSUJIMURA, N. (ed). **The Handbook of Japanese Linguistics**. Malden, MA, e Oxford, U.K.: Blackwell Publishers, 1999. p. 62–100.
- JOHNSON, K. Speech perception without speaker normalization: An exemplar model. In: JOHNSON, K.; MULLENIX, J. W. (eds.). **Talker Variability in Speech Processing**. San Diego: Academic Press, 1997. p. 145–165.
- JOHNSON, K. Decisions and mechanisms in exemplar-based phonology. In: SOLE, M.; BEDDOR, P.; OHALA, M. (eds.). **Experimental Approaches to Phonology**. Oxford: Oxford University Press, 2007. p. 25–40.
- KAYE, J. Do you believe in magic? The story of s+C sequences. **SOAS Working Paper in Linguistics and Phonetics**, v. 2, p. 293–313, 1992.
- LENNES, M. **Calculate segment durations**. Script para Praat. 2002. Disponível em: <http://www.helsinki.fi/~lennes/praat-scripts/public/calculate_segment_durations.praat>. Acesso em 8 de setembro de 2018.
- LEVITT, J. The influence of orthography on phonology: a comparative study (English, French, Spanish, Italian, German). **Linguistics**, 208, p. 43–67, 1978.
- LOVINS, J. Loanwords and the phonological structure of Japanese. **Linguistics Club**, v. 163, n. 1, 1975.
- MADUREIRA FEIJÓ, J. de M. **Orthographia, ou arte de escrever, e pronunciar com acerto a lingua portugueza para uso do excellentissimo duque de Lafoens**. Lisboa Occidental: Na Oficina de Miguel Rodrigues, 1734.

- MATTINGLY, I. G. Linguistic awareness and orthographic form. **Advances in Psychology**, v. 94, p. 11–26, 1992.
- MENESES, F.; ALBANO, E. From Reduction to Apocope: Final Poststressed Vowel Devoicing in Brazilian Portuguese. **Phonetica** (Basel), n. 72, p. 121–137, 2015.
- MORAIS, J.; CARY, L.; ALEGRIA, J.; BERTELSON, P. Does awareness of speech as a sequence of phones arise spontaneously?. **Cognition**, v. 7, p. 323–331, 1979.
- PIERREHUMBERT, J. Exemplar dynamics: word frequency, lenition and contrast. In BYBEE, J.; HOOPER, P. (eds.). **Frequency and the emergency of linguistic structure**. Amsterdam: J. Benjamins, 2001. p. 137–157.
- PIERREHUMBERT, J. Probabilistic Phonology: discrimination and robustness. In: BOD, R.; HAY, J.; JANNEDY, S. (eds.). **Probability theory in linguistics**. Cambridge, MA: The MIT Press, 2003. p. 177–228.
- PIERREHUMBERT J.; BECKMAN, M.; LADD, R. Conceptual foundations of phonology as a laboratory science (reprint from 2000). In: COHN, A.; FOUGERON, C; KUFFMAN, M. (eds.). **The Oxford Handbook of Laboratory Phonology**. Oxford: Oxford University Press, 2011 [2000]. p. 1–23.
- POLGÁRDI, K. The representation of sC-clusters: A CV analysis. In: BERCÉS, K. (ed.). **Proceedings of the Government Phonology Roundtable 2017**, Budapest, Hungria, 18 de novembro de 2017. Budapest: Pázmány Péter Catholic University.
- PURSE, R. Variable Word-Final Schwa in French: an OT Analysis. **U. Penn Working Papers in Linguistics**, v. 25, n. 1, p. 199–208, 2019.
- R CORE TEAM. **R: a Language and Environment for Statistical Computing**. R Foundation for Statistical Computing. Vienna, Austria. 2013. Disponível em: <<http://www.R-project.org/>>. Acesso em: 5 de dezembro de 2017.
- SCLIAR-CABRAL, L. **Princípios do Sistema Alfabético do Português do Brasil**. São Paulo: Contexto, 2003.
- SMITH, J. L. Loan phonology is not all perception: evidence from Japanese loan doublets. **Japanese/Korean Linguistics**, n. 14, p. 63–74, 2006.
- SOARES, M. **Alfabetização: a questão dos métodos**. São Paulo: Contexto, 2018.
- TAFT, M. An alternative to grapheme-phoneme conversion rules?. **Memory & Cognition**, n. 10, p. 465–474, 1982.
- TAFT, M. The Influence of Orthography on Phonological Representations in the Lexicon. **Journal of Memory and Language**, v. 24, n. 3, p. 320–335, 1985.
- TAFT, M. Orthographically influenced abstract phonological representation: Evidence from non-rhotic speakers. **Journal of Psycholinguistic Research**, v. 35, n. 1, p. 67–78, 2006.

- TODD, S.; PIERREHUMBERT, J.; & HAY, J. Word frequency effects in sound change as a consequence of perceptual asymmetries: an exemplar-based model. **Cognition**, n. 185, p. 1–20, 2019.
- TREIMAN, R. Learning to spell: phonology and beyond. **Cognitive Neuropsychology**, 2017.
- TREIMAN, R.; KESSLER, B. **How children learn to write words**. New York: Oxford University Press, 2014.
- VELOSO, J. **Da Influência do Conhecimento Ortográfico sobre o Conhecimento Fonológico: Estudo Longitudinal de um Grupo de Crianças Falantes Nativas do Português Europeu**. München: Lincom Europa, 2007.
- VELOSO, J. Phonology and Writing: Can we look at written productions to “see the unseeable” in phonology?. **Loquens**, v. 6, n. 1, p. 1–12, 2019.