

FREQUÊNCIA LEXICAL DE SEQUÊNCIAS MEDIAIS DE OBSTRUINTES NO PORTUGUÊS BRASILEIRO

LEXICAL FREQUENCY OF MEDIAL OBSTRUENTS SEQUENCES IN BRAZILIAN PORTUGUESE

Valéria Neto de Oliveira Monareto¹

monareto@ufrgs.br

RESUMO: Este trabalho trata de analisar se há relação no uso de palavras com sequências mediais de obstruintes (*aspecto*, *apto*, *ritmo*, etc.) com a frequência de palavras com esse tipo de estrutura, tendo por base que alguns processos de mudanças linguísticas possam ser motivados pela frequência lexical. Esse tipo de estrutura C1C2 no interior da palavra, herdada do latim, sofreu mudanças ao longo dos tempos, vocalizando ou apagando, principalmente, a primeira consoante (C1). Há, contudo, algumas dessas palavras no português que foram reinseridas na Língua. Apesar de a consoante obstruinte ser proibida na coda do português brasileiro atual, há, curiosamente, palavras, cujas entradas mais frequentes no dicionário são do tipo velar/labial+alveolar. Cabe notar também que essas sequências mediais mais comuns /kt/ e /pt/são as que menos sofrem o processo de epêntese vocálica na fala do Sul do Brasil, segundo os dados do Projeto Variação Lingüística no Sul do Brasil (VARSUL), demonstrando que são estruturas resistentes. Três amostras de natureza e épocas distintas do português brasileiro foram examinadas, levantando-se e analisando-se *types* de sequências mediais de obstruintes e seus *tokens*. Os resultados mostraram que é baixo o número de palavras com esse tipo de contexto na Língua e que a frequência do tipo da sequência (*type*) parece não estar relacionada com o seu número de ocorrências e com determinadas palavras (*tokens*). O português brasileiro atual parece, pois, utilizar diferentes palavras e estruturas antigas.

PALAVRAS-CHAVE: Sequências mediais de obstruintes; frequência de consoantes no interior da palavra no VARSUL; *types* e *tokens* de sequências de obstruintes mediais no português brasileiro.

¹ Doutor em Letras; Departamento de Linguística, Filologia e Teoria Literária do Instituto de Letras da Universidade Federal do Rio Grande do Sul.

* Este trabalho contou com a participação, em sua fase inicial, de Gisela Collischonn, que colaborou com algumas ideias e com o fornecimento de dados de seus estudos sobre a epêntese em contextos de sequências de obstruintes com dados VARSUL (Variação Linguística do Sul do País). Por isso, dedico este artigo a ela, que deixou muitas saudades. Eventuais falhas se devem a mim exclusivamente.

ABSTRACT: This work tries to analyze if there is a relation in the use of words with medial obstruents sequences (aspect, apt, rhythm, etc.) with the frequency of words with this type of structure, based on that some processes of linguistic changes can be motivated by lexical frequency. This type of C1C2 structure within the word, inherited from Latin, underwent changes over time, vocalizing or deleting, mainly, the first consonant (C1). There are, however, some of those words in Portuguese that have been reinserted into the Language. Although the obstruent consonant is prohibited in the coda of the current Brazilian Portuguese, there are, curiously, words, whose most frequent entries in the dictionary are velar / labial + alveolar type. It should also be noted that these most common media sequences /kt/ e pt/ are the ones that suffer the least from the process of vowel epenthesis in southern Brazil, according to data from the Linguistic Variation Project in Southern Brazil (VARSUL), demonstrating that they are resistant structures. Three samples of nature and distinct periods of time of the Brazilian Portuguese were examined, raising and analyzing types of media sequences of obstruents and their tokens. The results showed that the number of words with this type of context in the language is low and that the frequency of the *type* sequence does not seem to be related to its number of occurrences and to certain words (tokens). The current Brazilian Portuguese seems, therefore, to use different words and old structures.

KEYWORDS: medial sequences of obstruents; frequency of consonant within the word in VARSUL; types and tokens of sequences of media obstruents in Brazilian Portuguese.

INTRODUÇÃO

Os encontros consonantais mais frequentes da Língua Portuguesa são formados por consoantes obstruintes e uma líquida, chamados também de grupos consonantais próprios (COUTINHO, 19771, p.118). Há um outro tipo de grupo consonantal chamado de impróprio por aparecer em uma quantidade menor de palavras e por sofrer mudanças ao longo da história, além de ter implicações para a estrutura silábica.

Os grupos consonantais impróprios ocorrem em início e final de sílaba. No interior de palavra, sofrem alterações fonéticas desde o Latim Vulgar até os dias atuais. Este artigo tratará desse tipo de grupo em posição medial, o qual será denominado aqui como sequência medial de obstruinte (CC).

O português brasileiro atual dispõe de um número considerável de contextos com sequências mediais de obstruintes, distribuídos em consoantes de diferentes articulações. Na posição de primeira consoante (C1) da sequência (C1C2), há oclusivas labiais (*optar*); oclusivas alveolares (*atmosfera*); oclusivas velares (*pacto*, *signo*); fricativa labiodental (*afta*); nasal labial (*amnésia*). Em segunda posição (C2), há oclusivas não-nasais (*rpto*); oclusivas nasais (*mogno*, *pigmeu*); fricativas sibilantes (*opção*); fricativas não sibilantes (*advogado*).

Essas consoantes sofreram mudanças ao longo da história no sentido de simplificação em um único som. A primeira consoante da sequência (C1) é apagada em muitas palavras originadas do latim e, em outras, vocaliza-se de forma regular, formando um novo ditongo na Língua Portuguesa. Apesar de muitas transformarem-

se ou desaparecerem, algumas dessas sequências foram reinseridas na Língua, fazendo parte de um vocabulário culto ou erudito.

A instabilidade da presença ou não da primeira consoante da sequência, tanto na língua escrita como na língua falada, ficou registrada em gramáticas históricas e nas reformas ortográficas, oscilando-se interpretações sobre seu o valor, sua pronúncia e seu registro gráfico entre os mais destacados estudiosos da língua. Na estrutura silábica, causou também uma discussão se a C1 deveria ficar na coda da sílaba anterior, como em *ap.to*, por exemplo, ou no ataque da sílaba seguinte, como em *a.pto*.

Uma vogal epentética é usada no português brasileiro atual para licenciar a C1, formando-se uma nova sílaba, como em a.p[i].to. Em alguns casos, essa nova sílaba desloca o acento: 'op. ção ~ o.'p[i]. ção. Em muitas variedades, essa vogal epentética é usada quase que categoricamente. No sul do País, apresenta-se a como uma regra variável de aplicação em torno dos 40%.

Os dados apresentados em estudos sobre a regra variável da epêntese vocálica de Collischonn (2002, 2003, 2004) permitem-nos tecer algumas considerações em busca de evidências empíricas para a preservação de codas com obstruintes no português brasileiro. A velar desvozeada /k/ é a C1 mais frequente em termos de ocorrências, assim como as sequências /kt/, /ks/, /pt/ e /ps/. Essas sequências inibem a aplicação da epêntese e parecem violar a condição de coda no estágio atual da língua².

Numa perspectiva em que a frequência de *type*, bem como de *tokens*, determina a realização fonética das palavras, a distribuição de contextos de sequências mediais de obstruintes parece acompanhar o tipo de consoante na combinação C1C2. Resta saber, no entanto, se o tipo de palavra e sua frequência de uso também teriam algum papel.

Este artigo procurará responder se há ou não relação da frequência de *types* e *tokens* de sequências mediais de obstruintes (CCs) no português brasileiro, por meio do exame de três amostras de natureza e épocas distintas. Para tanto, esse texto está organizado como segue. Na primeira seção, serão apresentadas algumas considerações de ordem histórica, no sentido de situar o leitor sobre a evolução de CCs desde o latim. Na segunda, apresentam-se a metodologia para a aferição de *types* e de *tokens*, bem

² Apesar da formação de africadas com oclusivas seguidas de sibilantes, como em an[ts] (antes) e medicina (me[ds]ina), por exemplo, evidências estatísticas de distribuição de dados e de taxa de aplicação de epêntese mostram que diferentes combinações de oclusivas + /s/ apresentam valores de peso relativo variantes de 0.26 (*oxida*) a 0.52 (*observar*), contrariando a sustentação de hipótese das africadas. Todavia, em sequências com C1 desvozadas /ps, ts, ks/, verificam-se taxas de aplicação de epêntese próximas entre as diferentes sequências (Collischonn, 2002, p. 225).

como a definição da estrutura examinada, os objetivos específicos desta investigação e a explicitação das amostras utilizadas. Por último, na terceira seção, os resultados serão apresentados, seguidos pela conclusão.

1. PERCURSO DAS CCS MEDIAIS NA LÍNGUA PORTUGUESA

Do Latim para o português, houve um processo regular de simplificação dessas sequências mediais. Algumas mudanças foram implementadas: um ditongo pela vocalização de C1, como em *noite* (*nocte*), *conceito* (*conceptu*), dentre outros casos; o apagamento de C1 em *fato* (*facto*), *sinal* (*signal*), *aluno* (*alumno*), *aumento* (*augmento*), *sete* (*septe*), *pronto* (*prompto*), por exemplo, e o surgimento de uma palatal em *punho* (*pugnu*), dentre outros exemplos.

Algumas palavras com sequências adjacentes de obstruintes reapareceram no português por empréstimo do latim literário a partir do século XVI (CAMARA JR, 1985) e por questões de substrato (MATTOS e SILVA, 2001). O apreço à erudição greco-latina e à estética da palavra escrita contribuiu para o ressurgimento de palavras com grafias antigas (GONÇALVES, 1992), assim como o contato com a literatura clássica serviu como inspiração cultural e linguística (CASTRO, 2011).

Contudo, já no século XVI, a realização da consoante, que seria “muda” apresentava variação entre os escritores e estudiosos. No dicionário de Rafael Bluteau³ do século XVIII, é possível atestar variações de grafia, como, por exemplo, em *aspecto~aspecto~aspeito*; *adaptar~adatar*; *advogado~avogado*, *técnica~ténica*, dentre outros casos.

Apesar de a mudança fônica ter se efetivado em muitas palavras no português brasileiro na direção de se apagar a C1 principalmente (*acto>ato*; *damno>dano*, *concepto>conceito*, *augmento>aumento*, *prompto>pronto*), há diferentes CCs no léxico do português brasileiro atual que parecem se manter (*pacto*, *amnésia*, *conceptivo*, *pigmeu*, *peremptório*, etc.). A sequência /kt/ é a que possui maior número de itens (1.954 verbetes, segundo versão 3.0 digital do dicionário Aurélio), seguida de /pt/ com 988 ocorrências, de /ps/ com 629 e de /gn/ com 579, conforme pode ser visto em (1).

³ Este dicionário foi oferecido ao rei de Portugal D. João V, reunindo exemplos de palavras usadas pelos melhores escritores portugueses e latinos. São oito volumes com 43.664 verbetes. Os exemplos citados nesse trabalho foram retirados de sua versão eletrônica.

(1) Sequências mediais de obstruintes no PB segundo o Dicionário Aurélio eletrônico (versão. 3.0) em ordem crescente de número de registros

C1C2	Exemplo	Registros		C1C2	Exemplo	Registros
1º kt	pacto	1.954		12º b3	objeto	100
2º pt	apto	988		13º d3	adjetivo	74
3º ps	opção	629		14º dm	admirar	73
4º gn	signo	579		15º bt	obtusos	64
5º bs	abcesso	315		16º bk	subclasse	56
6º ks	facção	287		17º dv	advogado	50
7º kn	técnica	219		18º bm	submundo	38
8º gm	pigmeu	190		19º bn	abnegação	31
9º bd	abdômen	143		20º bv	óbvio	28
10º ft	afta	141		21º bp	subproduto	18
11º tm	atmosfera	105		22º kp	ecpiesma	12

Fonte: com base em Monaretto (2015)

Boa parte dos CCs é formada principalmente pelos prefixos sub-, sob- (subtítulo, sobnegar), ad- (adjunto) e ab- (abdicar), por palavras compostas (bem-nascido) e na forma de siglas ou estrangeirismos (CPMF, Km, *walkman*, etc.). Muitas dessas palavras são de origem grega, eruditas e ligadas a áreas da química, biologia e física.

Palavras com sequências de consoantes parecem ser pouco utilizadas. Segundo Silveira (2007, p.22), um exame em um *corpus* de extratos de textos eletrônicos da Folha de São Paulo (CETENFolha) revelou um número reduzido de palavras com CCs, como *afta*, *amnésia* e *compacto* (2%). Acima de 10% de frequência sobre um total de 18.824 dados, a autora encontrou as palavras *advogado* (32,14%), *técnica* (24,43%), *objeto* (13,92%), *obter* (11,80%).

Em época oitocentista no Brasil, Barbosa (2005, p.36), encontra 3,69% de palavras com grafias latinizadas, com CCs e geminadas, em folhetins cariocas. Proporção similar de ocorrência desse tipo de contexto pode ser vista em jornais gaúchos da mesma época também é encontrada em Monaretto (2015).

Na língua falada, contextos de CCs são pouco usados também. Collischonn (2003) examina a fala de 144 indivíduos do Banco VARSUL (Variação linguística do Sul do Brasil) e observa a inexistência de palavras com /ft/ (afta) e /mn/ (amnésia), além de um baixo número de dados com outras sequências. O maior número de ocorrências corresponde à combinação alveolar+nasal (**admin**istrar, **ritmo**), com 66% do total de dados.

Apesar de as obstruintes em posição de coda serem ruins ao padrão silábico do português, novos contextos são criados por realizações com síncope de vogal postônica não final, como em palavras como *tráfego*>tra[**fk**]o; *monótono*>mono[**tn**]o; *lúgubre*>lu[**gb**]re (BRANDÃO; DE PAULA, 2015) e em *cócegas* > co[**sk**]a, *atlântico*>atlân[**tk**]o, *rápida*>ra[**pd**]a (AMARAL, 2002), dentre outras. Segundo Vennemann (1975), a regra da síncope parece ser uma *regra espelho* à de epêntese vocálica.

A instabilidade da C1 na sequência continua operante nos dias atuais. A presença/ausência de C1 no português brasileiro moderno é variável em algumas palavras, como pode ser visto em registros de dicionários (contacto/contato; espectro/espetro, aspecto/aspeto, expectativa/expetativa, conectivo/conetivo, por exemplo).

Para Camara Jr. (1985, p.61), a inserção de uma vogal epentética, formando uma nova sílaba, é o recurso usado pelo falante para não se violar o padrão silábico do português brasileiro. Para Camara Jr. (1985), a reintrodução de consoantes mediais ocorreu pela língua escrita, pois na língua oral desenvolveu-se uma nova sílaba com a vogal /i/ ou /e/ no Brasil.

Contrariamente a outras regiões do País, a variedade do sul do Brasil parece não corresponder a essa tendência apontada por Camara Jr. Bisol (1999) acredita que a existência de obstruinte na coda seja um resquício de uma regra antiga, como um afrouxamento na condição de padrão silábico.

Collischonn (2003) encontrou na variedade do Sul do Brasil (VARSUL) um total de 41% de aplicação de epêntese, sendo que as alveolares/oclusivas (adquirir) e labiais/oclusivas (optar, corrupto) são as sequências que mais favorecem a epêntese. A posição postônica (óbvio) e as velares desvozeadas na posição de C1 (aspecto) são os fatores menos favoráveis para a aplicação da epêntese, somando-se a combinação velar/oclusiva desvozeada (pacto, aspecto) e velar/sibilante (opção, ficção).

A epêntese parece estar condicionada ao tipo de C1 (velar), ao tipo de sequência e à posição de C1 como pretônica (objeto). É possível que as palavras e contextos de CCs existentes estejam relacionados às condições contrárias à aplicação da epêntese? Essa questão, apesar de relevante, não será explorada nesse trabalho.

2. LEVANTAMENTO DE *TOKENS* E DE *TYPES* DE CCs MEDIAIS: CONFRONTO SÉCULO XIX E XX

2.1 METODOLOGIA

A investigação de *types* e de *tokens* de CCs tem por base a ideia de que palavras mais frequentes estejam mais suscetíveis a processos de mudança (Phillips, 1984; Bybee, 2002). A frequência de ocorrência (*tokens*) indica quantas vezes uma determinada palavra ocorre em determinado *corpus*, e o tipo (*type*) indica o uso de certa estrutura em relação à quantidade de itens.

A medida de *type* diz respeito, neste trabalho, à sequência de obstruintes no interior de palavras representadas por dois grafemas (kt, gn, pt, etc.) ou por um, no caso da letra “x” (ks), em palavras como tóxico, sexo, etc. Foram considerados como *tokens* diferentes, contabilizados, pois, como unidades, as palavras flexionadas ou derivadas, como, por exemplo, advogado, advogada, advogamos, advogados, etc, como 4 tokens.

Não serão consideradas nesse levantamento as palavras derivadas do prefixo sub-, siglas e palavras compostas. No caso do prefixo sub-, acredita-se que possa ter um comportamento fonológico independente (cf. COLLISCHONN, 2002).

Para fins de análise, as medidas de frequência são diferentes nas amostras, partindo-se do pressuposto de que não há certa medida de frequência em uma mesma comunidade de fala, já que os indivíduos têm experiências linguísticas diferentes (HUBACK 2013, p.81).

A investigação de CCs mediais tem como objetivos específicos:

1. levantar os *types* de CCs mais frequentes e verificar a proporção de *tokens/types* da amostra;
2. verificar que palavras com CCs mediais são mais utilizadas em cada amostra;

3. examinar se há relação na proporção *type/tokens* de CCs utilizada em dois períodos do português brasileiro;
4. comparar se os contextos mais frequentes de CCs são os mesmos em um *corpus* de referência.

Como pode ser observado em (1), sequências com velares e labiais /kt/, /ks/, /pt/ e /ps/ são as que têm mais número de entradas em dicionário atual. Por isso, procura-se descobrir também se essas CCs são utilizadas em um mesmo item lexical, ou seja, em palavras que se repetem por certa familiaridade, no sentido de serem mais conhecidas pelo indivíduo. Há necessidade de se saber se essas sequências são produtivas em razão de serem usadas em diferentes itens lexicais ou não.

2.2 AMOSTRAS

Serão utilizadas amostras de dois períodos do português brasileiro e um *corpus* de referência⁴.

2.2.1 SÉCULO XIX: amostra de língua escrita de cartas em jornais e cartas manuscritas pessoais entre membros de uma família. Foram examinadas 114 cartas aproximadamente.

As cartas impressas são de redatores e de leitores de jornais brasileiros pertencentes à amostra do Projeto “Para a História do Português Brasileiro” (PHPB). São analisados seis estados brasileiros (Rio de Janeiro, São Paulo, Minas Gerais, Pernambuco, Ceará, Bahia), cujos dados foram obtidos pelo endereço <https://sites.google.com/site/corporaphpb/>. O total de *tokens* nas cartas examinadas é de 46.290 ocorrências.

As cartas manuscritas são 21 correspondências entre familiares de Júlio Prates de Castilhos, personagem da história gaúcha de reconhecido papel na formação social e política do Rio Grande do Sul na época oitocentista. O total de *tokens* dessa amostra é de 16.810 ocorrências.

⁴ Segundo Berber Sardinha (2000), *corpus* de referência pode ser entendido como aquele usado para fins de contraste com o *corpus* de estudo. Deve ser, pelo menos, cinco vezes maior do que a amostra utilizada na investigação.

2.2.2 SÉCULO XX: amostra de língua falada do Banco de Dados VARSUL (COLLISCHONN, 2003) formada por 973 *tokens* distribuídos entre 144 indivíduos de seis cidades da região sul do Brasil - Porto Alegre, Florianópolis, Curitiba, Flores da Cunha, Panambi e Blumenau⁵.

2.2.3 CORPUS DE REFERÊNCIA: amostra de português brasileiro popular escrito do jornal “Diário Gaúcho” de Porto Alegre-RS nos anos de 2008, pertencente ao “Projeto PorPopular”. São 1.897.360 *tokens* de palavras lexicais e 126.032 *types* disponíveis nessa amostra.

3. RESULTADOS

3.1 SEQUÊNCIAS MEDIAIS DE OBSTRUENTES NO BRASIL NO SÉCULO XIX

O exame em cartas de redatores e de leitores do século XIX do *corpus* do PHPB mostrou um certo equilíbrio de ocorrências de CCs nas duas modalidades (redator/leitor) e um percentual bem baixo de palavras com CCs em relação ao número total de palavras (cerca de 1,38% em média). Os diferentes CCs (*types*) nessa amostra representam 59,72% de ocorrências em média sobre o total de *tokens*, ou seja, há uma certa diversidade de CCs entre as palavras utilizadas, o que é esperado, de certa forma, para o período examinado.

A distribuição da relação do número de palavras com CCs e os diferentes tipos de CCs utilizados nessa amostra (*tokens/types*), ou seja, a divisão do número de *tokens* pelo número de *types* em cada região, pode ser observada no Gráfico 1 a seguir.

⁵ Collischonn (2003) compôs esta amostra de dados para fins de investigação do processo de epêntese vocálica em contextos de sequências de obstruintes iniciais, mediais e finais. Outras publicações dessa autora derivaram de parte dessa amostra utilizada aqui neste trabalho (COLLISCHONN, 2002, 2004; COLLISCHONN; SCHWINDT, 2005). Essa amostra de 2003 é a mais abrangente em termos de localidades examinadas. Por isso é examinada neste trabalho. Foram retirados os contextos iniciais e finais, processando-se os dados novamente pelo pacote de programas estatísticos de análise de regra variável Goldvarb X.

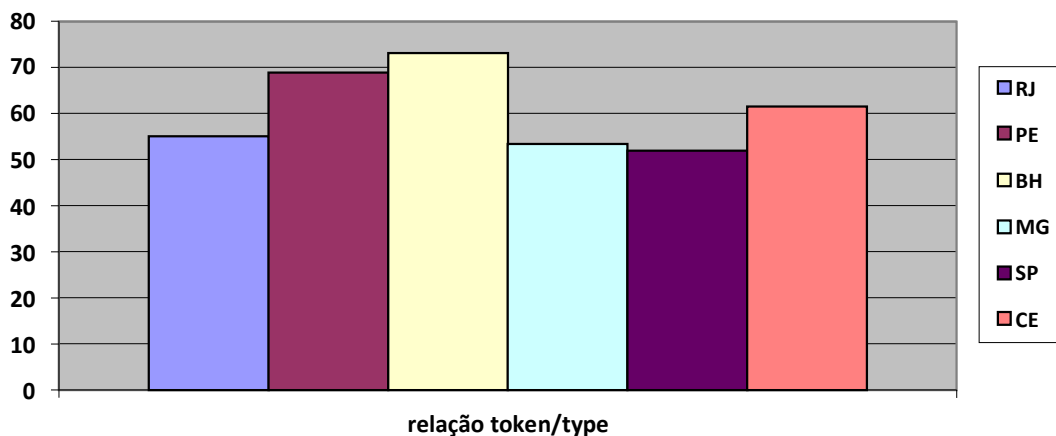


Gráfico 1: Distribuição de *tokens* e *types* de CCs mediais em cartas de leitores e redatores

O Gráfico 1 mostra que há diferentes proporções de *tokens* utilizados em cartas de redatores e leitores de jornais brasileiros do século XIX . Desse modo, é possível presumir que há certa riqueza lexical nesse material na medida em que o uso de sequências de consoantes diferentes ultrapassa mais da metade (59%) de ocorrências de *tokens* com este tipo de contexto. Isso demonstra que há palavras como *redactor* com /kt/, mas também que há outras palavras com o mesmo CCs, como *facto*, *actual*, *projecto*, por exemplo.

Em contrapartida, ao se examinar se o mesmo *token* é usado repetidamente na amostra, como a palavra *redactor*, por exemplo, podemos verificar que são poucos os *tokens* utilizados com a mesma CC, ou seja, o número de *tokens* refere-se às mesmas formas usadas repetidamente, como é o caso das formas mais frequentes como *redactor* (99 ocorrências), *redactores* (57 ocorrências.), *conducta* (12 ocorrências), *instrucção*, (12 ocorrências), *factos* (12 ocorrências), *prompto* (6 ocorrências).

No Quadro 1, é possível ver exemplos de *tokens* com CCs e os *tokens* e *types* mais recorrentes na amostra de cartas de redatores e leitores. Em virtude do número baixo de *tokens* repetidos, foram selecionados como os mais frequentes aqueles que tiveram acima de três ocorrências.

Sequências mediais de obstruintes mais frequentes (CCs)	Exemplos de <i>tokens</i> na amostra	<i>Tokens</i> mais frequentes (acima de três ocorrências)
1º kt	redactor, caracter	redactor, instrucção, actual, actos, circunspecto, distincto, dicto, facto, objecto, respectivo, projecto, victima, vitória
2º gn	assignatura, benigno	assignado, assignante, assignou, dignou, indigno, indignação
3º pt	assumpto, transcripto	escripto, escriptores, prompts
4º mn	calumnia, solemne	-----
5º ks	instrucção, convicção	acção, instrucção, direcção
6º dm	admiráveis, administração	admiração, admiráveis, administração, admitir
7º dv	adversários, advertência	-----
8º ps	lapso, utópia	-----
9º d3	adjacente, coadjuvado	-----
10º bt	obter, obtiveram	-----
11º gm	augmento, aumentar	dogma
Total de palavras da amostra: 46.290 Percentual médio de <i>tokens</i> com CCs: 59,72% Percentual de <i>types</i> de CCs: de 1,18% a 2,39%		

Quadro 1: Relação de CCs e de *tokens* mais frequentes em cartas de redatores e leitores século XIX

É possível também observar, por meio do Quadro 1, que o número de contextos com CCs é expressivo, mas que não há os 22 tipos relacionados no dicionário em (1), o que é bem compreensível devido à natureza da amostra examinada. Há também palavras “solitárias”, abaixo de três ocorrências, como é o caso dos CCs /mn/ (calumnia, solemne, /dv/ (adversários, advertência), /ps/ (lapso, autópsia) e / d3/ (adjacente, coadjuvado).

Comparando-se os dados no Quadro 1 com os CCs mais numerosos em termos de entradas lexicais no Dicionário atual em (1), podemos verificar que os CCs /kt/ e /gn/ estão entre os três mais frequentes.

Analisando-se o século XIX ainda, temos como uma segunda amostra, cartas pessoais entre familiares. No Quadro 2, é possível ver que a relação dos CCs encontrados é bem inferior ao dos jornais no Quadro 1, em virtude da natureza da amostra e da quantidade de *tokens*.

Sequências mediais de obstruintes (CCs)	<i>Tokens</i> da amostra
kt	directamente, facto, victoria, afecto, fructa, necto, dicto, dactada, ficta, projecto, projectar, contracto, exacto, successo, direcção, dacta, refractaria, efectivando, expectativa, actividade, junctar, directamente, contracto, actual, actualmente
pt	promptidão, prompto (3x), inenterrupto, assumpto, adaptado, escriptorio (3x), saptisfeito, isempta,
gn	augmentar, repugnância
mn	calumniosos
bz	obséquio
Total de <i>tokens</i> da amostra: 16.810 Total de <i>tokens</i> com CCs: 42 Total de <i>Types</i> de CCs: 5 Percentual de <i>tokens</i> sobre número de palavras: 0,23%	

Quadro 2 – Relação de *types* e *tokens* de CCs em Cartas entre familiares no século XIX

Do mesmo modo que em cartas de jornais, as cartas manuscritas pessoais evidenciam que os CCs com mais *tokens* são os mesmos: /kt/, /pt/ e /gn/, embora com um número de ocorrências bem mais baixo.

Podemos concluir, a partir do exame de amostras de impressos e de manuscritos do século XIX, que *tokens* com CCs são utilizados de modo reduzido em relação às demais palavras e que os *types* mais frequentes são os mesmos nas duas amostras. A baixa aplicação de contextos com sequências mediais de obstruintes vai ao encontro da análise de Barbosa (2005) em folhetins cariocas oitocentistas.

3.2 SEQUÊNCIAS MEDIAIS DE OBSTRUENTES EM AMOSTRA DO SÉCULO XX

3.2.1 AMOSTRA VARSUL

A amostra examinada a seguir é diferente das anteriores não só em relação ao estágio da língua, mas também por que é de língua falada. A amostra é composta por entrevistas gravadas por volta de 1992 pelo Projeto VARSUL. Contextos de sequências de obstruintes mediais foram examinados na fala de 144 indivíduos por Collischonn (2003), para se descrever o comportamento variável da epêntese vocálica, que se aplica em 41% no total de 1025 dados.

Essa epêntese é usada de maneira diferenciada conforme a comunidade linguística. Dentre as seis cidades examinadas, Blumenau (SC) e Panambi (RS) apresentam o percentual mais baixo de realização de epêntese: 23% e 24%, respectivamente. Já as capitais, Porto Alegre (RS) e Curitiba (PR), lideram na aplicação da epêntese: 58% e 52%, respectivamente. Segundo Collischonn e Schwindt (2005, p. 260), a epêntese medial é bastante variável entre as comunidades do Sul do País, diferentemente da epêntese inicial, que atinge 90% das palavras, o que evidencia fenômenos diferentes.

Os resultados da análise da regra variável da epêntese em contextos mediais mostram que a velar /k/ é a consoante que menos favorece a inserção da vogal, conforme pode ser visto na Tabela 1. Do mesmo modo, coincidência ou não, esta consoante é a que se mostra mais frequente nos dados do século XIX e nos registros do dicionário atual.

Tabela 1: Aplicação de Epêntese cf. a C1 em Sequências Mediais de Obstruintes (Dados Collischonn 2003)

C1	Apl./Tot	Percentual	Peso Relativo
Alveolar (ritmo, advogado)	185/291	64%	0,67
Labial (optar, opção)	134/347	39%	0,54
Velar (aspecto, signo)	77/335	23%	0,31

Fonte: a autora

Input: 0,37

A consoante velar em coda violaria a condição de sílaba no português. A presença de obstruente nessa posição parece ser um afrouxamento na condição de coda ou um resquício de uma regra antiga, nos termos de Bisol (1999). Collischonn (2003) acredita que a velar desvozeada /k/ possa ser uma consoante não-marcada.

A fim de se analisar os contextos de sequências dessa amostra, examinamos a distribuição desses, excluindo-se siglas, bem como a combinação de C1 com C2 da sequência C1C2 na aplicação da regra de epêntese.

Tabela 2: Aplicação de Epêntese pelo cruzamento de C1 com C2

C1 →	Alveolar		Labial		Velar		Total
C2 ↓	aplic./total		aplic./total		aplic./total		
Nasal	117/176 = 66% <i>ritmo</i>		4/8 = 50% <i>apneia</i>		46/89 = 52% <i>técnica</i>		167/273=61%
Fricativa Não-sib.	36/63 = 57% <i>advogado</i>		16/28 = 57% <i>óbvio</i>		—————		52/84=62%
Oclusiva	27/37 = 73% <i>adquirir</i>		28/103 = 27% <i>helicóptero</i>		11/75 = 15% <i>pacto</i>		66/215=31%
Sibilante	5/22 = 23% <i>adstrato</i>		86/208 = 42% <i>opção</i>		20/171 = 12% <i>sexo</i>		11/401=28%
Total	185/291 = 45%		134/347=39%		77/335=23%		

Fonte: a autora

A Tabela 2 mostra, dentre outros aspectos, uma distribuição equilibrada de dados da C1 (alveolar: 291 ocorrências; labial: 347 ocorrências; velar: 335 ocorrências). No entanto, podemos também observar que há bem mais dados de combinações de sequências com alveolar e nasal (*ritmo*, *admitia etnia*, *administrativo*, etc.), com 176 ocorrências; de labial com sibilante (*opção*, *opcional*, *concepção*, *egípcio*, etc.), com 208 ocorrências e de velar com sibilante (*facção*, *ficção*, *confeccionar*, *sexo*, etc.), com 171 ocorrências.

Sintetizando, os contextos de CCs mais frequentes por ordem decrescente de ocorrências nos dados são: **labial+sibilante** /ps/ ou /bs/; **alveolar+nasal** /tm/, /dm/; **velar+sibilante** (/ks/), sendo que a diferença entre essa quantidade de ocorrências não é grande. Contudo, a regra de aplicação da epêntese não está na mesma

proporção de número de contextos de ocorrências, sendo a velar+sibilante com menos epêntese, seguida pela labial+sibilante.

Apesar de existir mais dados com /ps/ e menos com /ks/, é interessante notar que a C1 com labial /p/ e a velar /k/ mantêm-se como os mais frequentes também em dados do século XIX, mas é diferente quanto à C2. Em síntese, podemos resumir os resultados de CCs mais frequentes no século XIX e XX:

- a) Século XIX – CCs mais frequentes em termos de número de ocorrências são /kt/, /pt/ e /gn/;
- b) Século XX (dados VARSUL): CCs mais frequentes em termos de número de ocorrências é /ps/, /tm/ e /ks/;

Nos dois períodos examinados, as velares e as labiais são as C1 mais frequentes em termos de tokens. Mas seriam as mesmas palavras (types)? Vejamos a seguir uma análise mais detalhada dos dados VARSUL.

3.2.2 COMPARAÇÃO DO VARSUL COM CORPUS DE REFERÊNCIA

Como se pode observar, tanto em jornais e manuscritos do século XIX como na fala de informantes do banco VARSUL, palavras com sequências mediais de consoantes obstruintes são de número bem reduzido se comparadas às demais do vocabulário da Língua Portuguesa. Palavras com essa estrutura foram reinseridas na Língua pela via erudita (escrita), na época renascentista, consideradas, pois, da modalidade culta.

Uma hipótese para o baixo número de palavras com essa estrutura poderia ser a de que as duas amostras até agora analisadas são formadas por narrativas pessoais e coloquiais por serem do gênero “carta” do século XIX (de leitores e de redatores, no caso de jornais, e de familiares, no caso dos manuscritos) e por entrevistas sociolinguísticas do século XX, conduzidas com o objetivo de se captar uma fala vernacular (no caso do Banco VARSUL). Essa hipótese é difícil de ser defendida, pois muitas palavras com CCs são de uso bem popular e frequente nos dias de hoje, como *advogado*, *ritmo*, *objeto*, *corrupto*, *obter*, etc.

A pergunta agora que se levanta é sobre a existência dessas palavras no estágio atual da Língua: as palavras mais frequentes com essa estrutura CC medial na amostra do VARSUL são também as mais sujeitas ao processo de epêntese vocálica e em algum outro *corpus* de maior amplitude? Para responder essa questão, examinamos

detalhadamente a amostra VARSUL e uma amostra bem maior, em termos de número de palavras⁶ em exemplares do ano de 2008 de um jornal de ampla circulação e popular, chamado “Diário Gaúcho”, do Rio Grande do Sul, amostra esta pertencente ao Projeto PorPopular. São 1.897.360 *tokens* de palavras lexicais e 126.032 *types* disponíveis nesse *corpus*.

Os 15 *tokens* mais utilizados na amostra VARSUL, bem como os respectivos *types* e o número de aplicações da epêntese são demonstrados no Quadro 3 a seguir. Também nesse Quadro é possível observar o número de ocorrência de cada palavra no Projeto PorPopular e um *ranking* de posição em relação às palavras mais frequentes no VARSUL.

Tokens	CC	Nº ocorr. VARSUL	Ranking VARSUL	Nº aplic. epêntese/%	Nº ocorr. PorPop.	Ranking PorPop.
administração	dm	36	1 ^o	20/55 %	64	7 ^o
táxi	ks	33	2 ^o	2/7%	87	5 ^o
opção	ps	30	3 ^o	14/47%	85	6 ^o
absurdo	bs	28	4 ^o	18/64 %	23	9 ^o
aspecto	ks	26	5 ^o	2/8%	8	12 ^o
ritmo	tm	21	6 ^o	6/29%	97	3 ^o
advogado	dv	20	7 ^o	13/65%	90	4 ^o
fixo	ks	14	8 ^o	2/14%	14	11 ^o
técnica	kt	14	8 ^o	5/36%	133	1 ^o
hectares	kn	13	9 ^o	4/31%	9	13 ^o
objetivo	b3	13	9 ^o	9/69%	130	2 ^o
opções	ps	12	10 ^o	4/33%	56	8 ^o
observar	bs	11	11 ^o	9/82%	21	10 ^o
apocalipse	ps	10	12 ^o	0	1	15 ^o
excepcional	ps	9	13 ^o	2/22%	4	14 ^o

Quadro 3: Comparação dos *tokens* mais frequentes no VARSUL x *Corpus* de Referência (PorPopular)

⁶ Neste Projeto levanta-se o número de palavras que há em um texto (*tokens*) e o número de palavras diferentes que é repetido nele (*types*). A noção de palavra é a de *palavra gráfica*. Maiores detalhes podem ser vistos em <http://www.ufrgs.br/textecc/porlexbras/porpopular/objetivos.php>.

É possível verificar pelo Quadro 3 vários aspectos:

- a) as estruturas CCs mais utilizadas são /ps/, em quatro *tokens*, seguido por /ks/, em dois *tokens*. Os demais CCs aparecem em um único *token*;
- b) a C1 mais frequente é a velar, realizada como /ks, kt/. Esse tipo de CC aparece mais recorrente na amostra, seguida pelas labiais /ps, bs/. Esses resultados nos indicam que a C1 velar é a que mais ocorre, o que vai ao encontro dos encontrados no século XIX, em que velar e a labial são as C1 mais utilizadas na estrutura CC;
- c) a aplicação da epêntese parece ser maior em alguns *tokens* mais frequentes, como em *administração* (55%), *absurdo* (64%), *ritmo* (29%) e *advogado* (65%). Dentre essas quatro palavras, apenas *ritmo* não está no contexto pretônico, o qual favoreceria a aplicação da epêntese, segundo Collischonn (2002). Por outro lado, em *táxi* (7%), e em *aspecto* (8%), classificadas em segundo e quinto lugar no *ranking*, a aplicação é bem inferior em relação a outras palavras mais frequentes;
- d) o *ranking* das palavras mais frequentes no VARSUL é dessemelhante ao da amostra do PorPopular.

Diante do exposto, é possível concluir, a partir dos resultados do Quadro 3, que não há relação de frequência de tipo de palavra com o processo de epêntese vocálica e que também a frequência de uso de palavras com CCs mediais não é a mesma em um *corpus* de referência.

Um aspecto a se considerar em pesquisa futura é se essas palavras com sequências mediais de obstruintes são utilizadas por certos indivíduos da amostra, que mostrariam um léxico mais especializado do que os demais.

CONCLUSÃO

O exame de *types* e de *tokens* de sequências mediais de obstruintes em três tipos de amostras revelou que não há relação da C1 com a forma da palavra (*token*). Apesar do baixo número de *tokens* nas amostras, há diferentes combinações de sequências, distribuídas em diferentes palavras. Palavras com CCs são usadas numa proporção baixa em relação ao léxico em geral, mas são diversificadas em termos de *types*.

As sequências C1C2 mais frequentes são formadas por velares /k/ e labiais /p/ na posição de C1 em ambos estágios examinados. Já a C2 mais frequente é distinta. A sequência CC mais frequente depende, pois, do *corpus* examinado, não se confirmando que o tipo de C1 na coda seja determinado pela frequência lexical.

REFERÊNCIAS BIBLIOGRÁFICAS

AMARAL, Marisa Porto do Amaral. A síncope em proparoxítonas: uma regra variável. In: BISOL, Leda.; BRESCANCINI, Cláudia. *Fonologia e Variação: recortes do português brasileiro*. Porto Alegre: EDIPUCRS, 2002. p. 99-127.

AURÉLIO século XXI: o dicionário da língua portuguesa. 3. ed. rev. ampl. Rio de Janeiro: Nova Fronteira, 1999. Versão eletrônica 3.0.

BARBOSA, Afrânio Gonçalves. Tratamento dos corpora de sincronias passadas da língua portuguesa no Brasil: recortes grafológicos e linguísticos. In: LOPES, C. R. dos S. *A Norma Brasileira em Construção: fatos linguísticos em cartas pessoais do século 19*. Rio de Janeiro: FAPERJ, 2005. p. 25-42.

BERBER SARDINHA Tony. *Linguística de Corpus: Histórico e Problemática*. DELTA, v.16 n.2 : São Paulo, 2000.

BISOL, Leda. A Sílabas e seus Constituintes. In: NEVES, M. H. M. (org.). *Gramática do Português Falado V: novos estudos*. Campinas: Editora da UNICAMP, 1999.

BYBEE, J. Mechanisms of change in grammaticization: the role of frequency. In: Brian D. Joseph & Richard D. J. (Eds).. *Handbook of Historical Linguistics*. Oxford: Blackwell. 2002.

CAMARA JUNIOR, José Mattoso. *História e Estrutura da Língua Portuguesa*. Rio de Janeiro: Padrão, 1985.

CASTRO, Ivo. *Introdução à História do Português*. 2.ed. rev. e muito ampl. Lisboa: Edições Colibri, 2011.

COLLISCHONN, Gisela. A epêntese Vocálica no Sul do Brasil. In: BISOL, Leda; BRESCANCINI, Cláudia. *Fonologia e Variação: recortes do português brasileiro*. Porto Alegre: EDIPUCRS, 2002. p. 205-230.

COLLISCHONN, Gisela. Epêntese Vocálica e Restrições de Acento no Português do Sul do Brasil. *SIGNUM: Estudos Linguísticos*, Londrina, n. 7/1, p. 61-78, 2004.

COLLISCHONN, Gisela. Epêntese vocálica no português do Sul do Brasil: variáveis extralinguísticas. *Revista Letras*. Curitiba, n. 61, Especial, p. 285-297, 2003.

COLLISCHONN, Gisela; SCHWINDT, Luiz Carlos. Considerações sobre a sequência /sc/ inicial em português brasileiro. *Lingua(gem)*, v.2, n.2, Macapá, 2005, p. 249-266.

COUTINHO, Ismael de Lima. *Gramática Histórica*. 7. ed. Rio de Janeiro: Ao Livro Técnico, 1976.

CRISTÓFARO-SILVA, Thaí.; ALMEIDA, L. On the nature of epenthetic vowels. In: BISOL, L.; BRESCANCINI, C. (ed.). *Contemporary Phonology in Brazil*. Cambridge: Cambridge Scholars Publish, 2006.

- DE PAULA, Alessandra; BRANDÃO, Sílvia Figueiredo. Vogais em Contexto Postônico Não Final em Variedades do Português. *Revista da ABRALIN*, v.14, n. 1, 2015, p.47-84
- GONÇALVES, Maria Filomena. *Madureira Feijó/ortografista do século XVIII. Para uma História da Língua Portuguesa*. Lisboa: Ministério da Educação: Instituto de Cultura e Língua Portuguesa, 1992.
- HUBACK, Ana Paula. A interferência da frequência em fenômenos linguísticos. *Revista D.E.L.T.A.*, 29:1, 2013. p.79-94
- MATTOS E SILVA, Rosa Maria. *O Português Arcaico: fonologia*. São Paulo: Contexto, 2001.
- MONARETTO, Valéria Neto de Oliveira. Sequência de Obstruintes no Interior das Palavras: retrato em duas modalidades. *Revista Estudos Linguísticos e Literários*. Programas de Pós- Graduação em Língua e Cultura e em Literatura e Cultura da Universidade Federal da Bahia , n. 52, ago-dez|2015, Salvador: p. 205-230.
- PHILLIPS, Betty. Word frequency and the actuation of sound change. *Language* 60. n. 2. 1984.
- SILVEIRA, Francine. *Vogal Epentética no Português Brasileiro: um estudo acústico em encontros consonantais*. 2007. 92 f. Dissertação (Mestrado em Linguística) – Centro de Documentação e Expressão, Universidade Federal de Santa Catarina, Florianópolis, 2007.
- VENNEMANN, Theo. Rule inversions. *Lingua*, n. 29, 2015, p. 209-242.