

A variação na realização do futuro do presente em artigos acadêmicos

The variation in the realization of the future tense in academic articles

Táise Simioni*

Andressa Rodrigues Gomide**

Kerolyn Pereira Sarate***

Resumo

Este trabalho tem por objetivo descrever como ocorre a variação na realização do futuro do presente em artigos acadêmicos, levando em consideração as variantes forma simples e forma perifrástica. Para tal, através de uma metodologia filiada à linguística de corpus, foi organizado um subcorpus referente ao PB a partir do Corpus de Português Escrito em Periódicos. Foram analisadas as variáveis grande área (ciências exatas, ciências humanas e ciências biológicas) e avaliação Qualis-CAPES dos periódicos. Um cruzamento entre essas variáveis permitiu verificar que, por um lado, os periódicos de ciências exatas e ciências humanas preferem o uso da forma perifrástica em estratos mais baixos, ao contrário do que ocorre com os periódicos de ciências biológicas.

Palavras-chave

Varição linguística. Futuro do presente. Linguística de corpus. Artigos acadêmicos.

Abstract

This work aims to describe how variation occurs in the realization of the future tense in academic articles, taking into account the simple and periphrastic forms. With this aim and using Corpus Linguistics methods, we compile a subcorpus from the Corpus of Portuguese Written from Journals containing only texts written in Brazilian Portuguese. For this study, we consider the variables School of Knowledge (School of Exact, Technological and Multidisciplinary Sciences, School of Humanities and School of Life Sciences) and the Qualis-CAPES grades of the journals. A combination among these variables allowed us to verify that, on the one hand, the journals of exact sciences and human sciences prefer the use of the periphrastic form in lower strata, unlike what happens with the journals of biological sciences.

* Universidade Federal do Pampa (Unipampa).

** Lancaster University.

*** Universidade Federal do Pampa (Unipampa).

Keywords

Linguistic variation. Future tense. Corpus linguistics. Academic articles.

Introdução

Este trabalho tem por objetivo descrever a variação na realização do futuro do presente em português brasileiro (PB) a partir de um *corpus* constituído por artigos acadêmicos. Para tal, foram consideradas suas principais variantes: a forma simples (*dormirei*) e a forma perifrástica com *ir* (*vou dormir* ou *irei dormir*). A variante conservadora e padrão¹ é a forma simples, enquanto a variante inovadora e não padrão é a forma perifrástica.

Desde o século XIX, a perífrase com *ir* apresenta crescimento na modalidade falada da língua, o que desencadeou uma mudança linguística que se encontra praticamente concluída (OLIVEIRA, 2006). Na modalidade escrita, entretanto, a forma conservadora não perdeu seu espaço, já que pesquisas referentes ao futuro do presente no PB em textos escritos apontam que há ocorrência de perífrase, porém em menor porcentagem que o futuro simples (OLIVEIRA, 2006; 2011a; OLIVEIRA; OLINDA, 2008; FIGUEIREDO; OLIVEIRA, 2013). Ou seja, o futuro simples ainda é a variante preferida na escrita, mesmo que a perífrase não seja uma variante estigmatizada socialmente: “Mesmo sendo o futuro simples a forma prescrita, não há avaliação negativa (talvez nem a sua percepção) da variação entre as formas existentes para a expressão do futuro, fato que revela ser essa variação isenta de estigma social” (TESCH; YACOVENCO; SCHERRE, 2014, p. 90).

Para a realização desta pesquisa, foi utilizado o CoPEP - Corpus de Português Escrito em Periódicos (*Corpus of Written Portuguese from Journals*) (KUHN; FERREIRA, 2016), a partir do qual foi organizado um subcorpus restrito aos textos representativos do PB. Para a extração dos dados, utilizamos métodos da linguística de corpus e o *software* de código aberto CQPweb (HARDIE, 2012). Neste trabalho, levamos em consideração duas variáveis: a grande área (ciências exatas, ciências humanas e ciências biológicas) e o Qualis-CAPES das revistas (agrupado no estrato A1 até B1 e B2 até B5).

¹ O futuro do presente em PB comumente é apresentado nas gramáticas tradicionais e em manuais escolares sem a inclusão da perífrase com *ir* (OLIVEIRA, 2006), o que exclui uma possibilidade de variante muito frequente no PB atual, principalmente na modalidade oral.

Revisão da literatura

Com o intuito de compreender o processo de gramaticalização vivido pelo verbo *ir*, que resultou na perífrase, Oliveira (2006) fez uma análise em tempo real de longa duração que contemplou dados escritos do português dos séculos XVIII² (cartas oficiais, cartas de comércio e cartas comuns), XIX (cartas oficiais, cartas pessoais e cartas de editores) e XX (editoriais de jornais das décadas de 1970 e 1990).

Oliveira (2006) mostra que o futuro simples sempre foi a variante preferida na modalidade escrita da língua, entretanto o seu principal concorrente foi mudando ao longo do tempo. Resumidamente, em relação ao PB, pode-se dizer que o futuro simples disputou espaço com *haver de* + infinitivo durante o século XVIII. No século XIX, *haver de* + infinitivo e *ir* + infinitivo concorriam por espaço como segunda variante de futuro na preferência dos falantes. Já no século XX, a forma perifrástica com *ir* passa a ser a principal oponente do futuro simples, enquanto a variante com *haver de* decai. O presente, por sua vez, enquanto variante de futuro verbal, sempre se manteve estável ao longo do tempo, como sintetiza Oliveira (2006) no seu esquema apresentado na Figura 1.

Figura 1 – A variação do futuro do presente ao longo dos séculos XVIII, XIX e XX

<p>Século XVIII: futuro simples > <i>haver de</i> + infinitivo > presente / <i>ir</i> + infinitivo</p> <p>Século XIX: futuro simples > <i>haver de</i> + infinitivo / <i>ir</i> + infinitivo > presente</p> <p>Século XX: futuro simples > <i>ir</i> + infinitivo > presente > <i>haver de</i> + infinitivo</p>

Fonte: Oliveira (2006, p. 106).

O fato de *haver de* + infinitivo como variante de futuro ter decaído no PB pode estar ligado a outro processo de variação da língua, no qual *haver* e *ter* concorrem no PB. Dessa forma, *haver* passa a ser utilizado em contextos formais (OLIVEIRA, 2006), o que, possivelmente, pode ter afetado o seu uso enquanto perífrase verbal. O verbo *ter* passa a possuir seu sentido amplificado, sendo assim o seu significado não se restringe apenas a posse (***tenho*** *um salão*), mas também se refere a existência (***tem*** *uma menina no salão*): “No português brasileiro, *haver* e *ter* experimentaram um novo embate pelo posto de verbo existencial, com o segundo, pelo observado até aqui, se

² Apresentamos, aqui, um recorte da análise feita por Oliveira (2006), que incluiu, também, dados do português europeu do século XIII ao XVII. Além disso, a autora analisou dados de fala do século XX.

saindo vencedor” (AVELAR, 2006, p. 54). Além disso, segundo Oliveira (2006), a variante *haver de* + infinitivo “possui um forte componente modal superposto ao de tempo, realizado no seu sentido injuntivo de [+ obrigação]. Progressivamente, é possível que esse sentido tenha sido reforçado em detrimento do sentido de tempo” (OLIVEIRA, 2006, p. 126).

O presente é outra variante que aparenta não ter significância na competição principal: futuro simples x futuro perifrástico com *ir* + infinitivo. Segundo Oliveira,

o que se percebeu é que o uso dessa variante se mantém, sistematicamente, em alguns contextos. Segundo as evidências, trata-se de uma variante que não concorre propriamente com o futuro perifrástico, mas sim com a forma de futuro simples. Do ponto de vista da mudança, parece que essa variante não faz parte do ‘tabuleiro de xadrez’ (2006, p. 175).

Com relação à forma perifrástica com *ir*, esta, ao longo do tempo, passou por um processo de gramaticalização no PB. Gonçalves (2013) explica a gramaticalização: “Dentre o universo das mudanças e variações linguísticas, encontra-se um subconjunto de variações que descreve como um item lexical vem a desempenhar funções gramaticais ou um item gramatical vem a assumir funções mais gramaticais ainda. Este fenômeno é denominado Gramaticalização” (GONÇALVES, 2013, p. 332).

O verbo *ir*, em seu sentido pleno, possui significado de deslocamento (**VOU ao restaurante**), entretanto, na realização do futuro do presente, o verbo passa a ser também auxiliar, já que faz parte da perífrase, e, portanto, o seu sentido é estendido também para futuridade (**VOU comer no restaurante amanhã**). Sendo assim, o verbo passou a desempenhar não só suas funções lexicais (verbo pleno), mas também funções gramaticais (verbo auxiliar). O processo de gramaticalização que sofreu o verbo *ir*, constituindo assim a perífrase, não é exclusivo do PB. Esse procedimento é bastante comum e pode ser observado em outras línguas indo-europeias, como, por exemplo, o francês e o espanhol (BORTONE; SIMON, 1994).

A fim de caracterizar o fenômeno variável da realização do futuro do presente em PB, apresentamos, brevemente, algumas pesquisas que se dedicaram à sua análise. Trabalhos como os de Oliveira (2006; 2011a), Oliveira e Olinda (2008) e Figueiredo e Oliveira (2013) adotam diferentes *corpora* e metodologias, o que inclui a análise de dados escritos e/ou falados e a observação da mudança em tempo aparente ou em tempo real, de curta ou de longa duração. Apesar de tais diferenças,

o que impede a observação do que há em comum em termos de variáveis extralinguísticas, algumas variáveis linguísticas se mostram significativas de forma recorrente entre os trabalhos. São elas o paradigma verbal, o papel temático do sujeito e a natureza semântica do verbo.

Quanto ao paradigma verbal, essa variável contempla dois fatores: os verbos regulares (*cantar, beber e partir*³) e os verbos irregulares (*estar, trazer e pedir*). Os estudos apontam que os verbos regulares tendem a favorecer o uso da perífrase, enquanto os irregulares favorecem o uso do futuro simples (OLIVEIRA, 2006; 2011a; OLIVEIRA; OLINDA, 2008; FIGUEIREDO; OLIVEIRA, 2013). Oliveira (2011a) afirma que, em vista de uma possível mudança linguística em curso, a observação de que esse processo avança primeiro nas formas regulares e depois nas irregulares poderia encontrar explicação no fato de as irregularidades verbais serem mais marcadas, portanto seriam estocadas individualmente na mente do falante.

O sujeito, quanto ao seu papel temático, pode ser sujeito agente (*o que direi agora, né?*), sujeito experienciador (*vocês vão perceber como acontece a tuberculose*) e sujeito paciente (*e esse chumbo não vai ser excretado pelo organismo*). Os resultados encontrados são de que o sujeito agente favorece o uso da perífrase, enquanto o sujeito paciente favorece o futuro simples e o sujeito experienciador fica na posição intermediária entre as duas variantes (OLIVEIRA; OLINDA, 2008; OLIVEIRA, 2006; 2011a). Segundo Oliveira (2006), a preferência da perífrase pelo papel temático de agente resultaria de um maior comprometimento do sujeito e um maior grau de certeza de que a ação será realizada, uma vez o próprio sujeito realizará tal ação. Podemos deduzir da explicação de Oliveira (2006) que a perífrase, em comparação à forma simples, manifesta um maior comprometimento com a realização da ação.

A variável natureza semântica do verbo se refere ao sentido que o verbo possui. Segundo Oliveira (2006), esse grupo de fatores pode ser distribuído em verbos que indicam processo (*Tal comissão vai fazer uma espécie de radiografia do INPS*), evento (*em cada cidadão, encontrarão um inimigo decidido a defender o Governo de que participa*), estado (*o governo vai manter o controle dos preços*) e verbos cognitivos (*Saberão que terão que enfrentar, muito mais ainda, a antipatia popular*). Figueiredo e Oliveira (2013) encontram como resultado o favorecimento da perífrase nos verbos

³ Os exemplos fornecidos para os fatores das variáveis paradigma verbal, papel temático do sujeito e natureza semântica do verbo foram retirados de Oliveira (2006).

que expressam processo. Em Oliveira (2011a), os verbos de processo obtiveram o segundo maior peso relativo, sendo superados pelos verbos cognitivos. A autora, entretanto, destaca a quantidade reduzida de verbos desta natureza (foram sete em um total de 631 dados). Segundo Oliveira (2011a) e Figueiredo e Oliveira (2013), a preferência da perífrase por verbos que indicam processo seria esperada uma vez que o verbo *ir*, em sua função lexical, expressa natureza semântica de processo.

Destacamos, aqui, que todos os trabalhos mencionados nesta seção analisam dados escritos. Oliveira (2006), além de dados escritos, inclui também dados orais em sua análise, conforme foi mencionado anteriormente. As fontes dos dados escritos incluem gêneros de diferentes níveis de formalidade, que vão de cartas oficiais, como em Oliveira (2006) e Olinda e Oliveira (2008), a quadrinhos, como em Figueiredo e Oliveira (2013).

Dos trabalhos mencionados, dois deles incluem entre as variáveis extralinguísticas tipo de documento (OLIVEIRA; OLINDA, 2008) ou gênero textual (FIGUEIREDO; OLIVEIRA, 2013). Em Oliveira e Olinda (2008), a variável tipo de documento, que tem como fatores cartas oficiais, cartas de comércio e cartas comuns, foi selecionada para os dados do século XVIII. Contrariando as expectativas, a perífrase foi preferida nas cartas oficiais, representativas de um estilo mais formal. Segundo as autoras, “como ainda não havia, no século XVIII, uma normatização da língua, talvez não houvesse tanta diferença no que tange ao tipo de texto (mais formal ou mais informal) e o futuro perifrástico pudesse ser usado mesmo em documentos oficiais” (OLIVEIRA; OLINDA, 2008, p. 110-111). As autoras alertam, entretanto, para a quantidade reduzida de dados (houve cinco perífrases para um total de 109 contextos). Figueiredo e Oliveira (2013), por sua vez, têm como *corpus* textos publicados em jornais. Com relação à variável gênero textual, a hipótese era a de que gêneros menos formais favoreceriam a perífrase, enquanto gêneros mais formais favoreceriam a forma simples. Em conformidade com essa hipótese, o anúncio publicitário foi o gênero em que mais ocorreu a perífrase. Segundo as autoras, isso pode ser explicado pela tentativa de o anúncio se aproximar da língua falada. Já em editais e atas não houve ocorrências de perífrase.

Os resultados descritos mostram uma correlação entre níveis de formalidade e uma maior ou menor ocorrência da forma perifrástica. Diante de tal correlação, optamos por analisar a ocorrência da perífrase em um gênero que pode ser localizado no extremo de formalidade da língua escrita, em conformidade com uma proposta de

contínuo na fala e na escrita (MARCUSCHI, 2010). Segundo Oliveira (2006, p. 103), “a mudança acontece primeiro na fala e só mais tardiamente atinge a língua escrita”. Conforme Scliar-Cabral (2014, p. 56), isso “decorre do caráter de permanência da modalidade escrita, do fato de tal sistema ser o meio normatizado para utilização pelas instituições educacionais e jurídicas, pela mídia e pelas demais formas bibliográficas”. Se a mudança demora mais a atingir a língua escrita, pode-se esperar que ela ocorra ainda mais tardiamente em gêneros escritos mais formais. Assim, cabe uma investigação, como a proposta neste trabalho, da ocorrência da variação no futuro do presente em artigos acadêmicos, que se encontram, como foi mencionado, em um extremo de formalidade na escrita.

Metodologia

Destacamos que, neste trabalho, utilizamos a linguística de corpus como metodologia, em que *corpus* e *software* são apenas ferramentas utilizadas para a pesquisa (McENERY; HARDIE 2011). Esta seção de metodologia está dividida em três partes: na primeira, é feita uma descrição do *corpus*; na segunda, há uma apresentação da ferramenta utilizada para a obtenção dos dados; e, na terceira, encontra-se uma descrição de como foram feitas as buscas dos dados a partir de tal ferramenta.

Descrição do corpus

Para as investigações, utilizamos o *corpus* CoPEP - Corpus de Português Escrito em Periódicos (*Corpus of Written Portuguese from Journals*) (KUHN; FERREIRA, 2016). Este *corpus* é composto por aproximadamente 10.000 artigos científicos, num total de mais de 40 milhões de palavras. Os artigos foram publicados, em sua maioria, entre 2000 e 2016. Alguns textos, que constituem apenas 2% do total, pertencem à década de 1990. Todos os artigos, escritos em português do Brasil ou de Portugal, possuem livre acesso e estão disponíveis na plataforma *Scientific Electronic Library Online* (SciELO).

Escolhemos o CoPEP por ser um *corpus* bem documentado e balanceado. Informações como ano, revista e área de conhecimento foram preservadas, garantindo a fácil busca por textos de acordo com essas variáveis. A distribuição equilibrada de acordo com os metadados fornece uma visão geral do uso da língua em diferentes contextos. Outro motivo para a escolha deve-se ao fato de o *corpus*

cobrir um significativo número de publicações nos últimos 15 anos. Por serem recentes, os textos refletem a produção linguística atual.

Para melhor atender aos objetivos deste estudo, preparamos um subcorpus do CoPEP com as seguintes modificações. Primeiro, extraímos apenas os textos produzidos na variedade brasileira. Em seguida, adicionamos a cada artigo a avaliação Qualis-CAPES da revista de sua publicação. Para isso, foi considerada a classificação do quadriênio 2013-2016, a mais recente que está disponível na Plataforma Sucupira. Por fim, anotamos e lematizamos o *corpus* utilizando a ferramenta de anotação morfossintática *TreeTagger* (SCHMID, 1994). Desta forma, o *corpus* utilizado (CoPEP – Brasil) possui 24.261.106 palavras distribuídas em 3.811 textos. Embora não haja um consenso sobre qual seja o tamanho ideal de um *corpus* (McENERY; HARDIE, 2011), podemos dizer que este *corpus* possui um bom número de palavras e textos. Isso porque o número de textos corresponde a uma porção próxima ao número total de artigos de acesso livre publicados no Brasil no mesmo intervalo de tempo. Além disso, o tamanho é superior a outros *corpora* e/ou subcorpora frequentemente utilizados como referência, como é o caso da seção acadêmica do British National Corpus (BNC XML Edition 2007), que possui menos de 18 mil palavras distribuídas em 497 textos.

A ferramenta

Para realizarmos as análises, utilizamos o *software* de código aberto CQPweb (HARDIE, 2012). CQPweb é um conjunto de ferramentas para análise de *corpora*. Dentre as inúmeras vantagens que o CQPweb possui, destacam-se (i) a possibilidade de compartilhar e acessar os *corpora on-line*; (ii) a presença de uma poderosa ferramenta de busca (*Corpus Query Processor* - CQP); e (iii) a facilidade em restringir as investigações no *corpus* de acordo com os metadados.

Ser uma ferramenta *on-line* garante que, independentemente da máquina utilizada para exploração do *corpus*, os dados e a versão do programa serão sempre os mesmos. Desta forma, não há conflito em resultados, como pode ocorrer quando utilizamos programas de instalação local, como é o caso do AntConc (ANTHONY, 2019). Além disso, uma vez que o *corpus* é preparado e disponibilizado *on-line*, outras pesquisas podem ser realizadas através do mesmo servidor, sem que haja a necessidade de preparar os dados novamente.

As buscas

A busca CQP nos permite criar buscas elaboradas através de combinações entre palavras, anotação e posição no texto. Para encontrar dados no *corpus* da forma simples de futuro do presente, utilizamos a seguinte fórmula:

```
[word=".*((a|e|i)(rei|rás|remos|reis|rão))" & pos="VERB.Fin.*" & word!="queremos"  
%c]
```

Com essa busca, obtemos apenas ocorrências de verbos na forma finita e com a terminação a(e,i)rei, a(e,i)rás, a(e,i)remos, a(e,i)reis, a(e,i)rão, exceto o verbo “queremos”, um resultado falso positivo.

Para o futuro perifrástico, buscamos pelo verbo “ir”, em todas as suas formas (exceto “iria”), seguido por verbos no infinitivo, utilizando a fórmula abaixo:

```
[lemma="ir" & word!="iria" %c][pos="VERB.Inf.*"]
```

Essas duas fórmulas foram utilizadas para buscas por conteúdos nos textos. Para restringir a investigação apenas aos textos selecionados, utilizamos a função “consulta restrita” do CQPweb. Esta função permite que as buscas sejam feitas em apenas partes do *corpus*, restringidas de acordo com os metadados. As restrições foram feitas de acordo com a grande área (ciências exatas, ciências humanas e ciências biológicas); com a avaliação Qualis-CAPES; e com uma combinação destas duas variáveis. Para a avaliação Qualis-CAPES, consideramos as revistas com Qualis-CAPES A1 a B1 em um grupo e as revistas B2 a B5 em outro. Esses agrupamentos tiveram como critério a noção tácita de que revistas com Qualis-CAPES A1 a B1 devem ter preferência no momento da seleção de um periódico para a submissão de um artigo. Embora isso possa variar a depender da área, tal critério se justifica, por exemplo, no fato de que há programas de pós-graduação que adotam como critério para credenciamento de docentes a publicação de artigos que incluem o estrato B1 e excluem estratos inferiores a este. Além disso, entre os critérios adotados para a classificação dos periódicos, há um que agrupa os estratos A1, A2 e B1, de maneira a não permitir que estes estratos ultrapassem 50% do total de periódicos.

Para verificarmos se as diferenças em frequência não são ao acaso, utilizamos o teste de significância não paramétrico qui-quadrado. Alguns autores (como, por exemplo, Brezina e Meyerhoff (2014) e Bestgen (2017)) argumentam que, em linguística de corpus, o teste do chi-quadrado tende a gerar resultados significantes em demasia. Por isso, optamos por também gerar gráficos *boxplot* para melhor visualizar a distribuição das frequências no *corpus*.

Apresentação e discussão dos resultados

O Quadro 1 apresenta uma visão geral dos resultados obtidos. Como é possível observar, há uma frequência relativa maior da variante simples, em conformidade com outros trabalhos que analisam a variação na realização de futuro do presente em *corpora* escritos, conforme mencionamos anteriormente.

Quadro 1 – Visão geral dos resultados

	Forma perifrástica	Forma simples
Busca	[lemma="ir" & word!="iria" %c][pos="VERB.Inf.*"]	[word=".*(a e i)(rei rás remos reis rã o)" & pos="VERB.Fin.*" & word!="queremos" %c]
Frequência Absoluta	3,785 (em 24,261,106 palavras)	5,431 (em 24,261,106 palavras)
Frequência Relativa	156.01 ocorrências por milhão de palavras	223.86 ocorrências por milhão de palavra
Textos	1,321 (em 3,811 textos)	1,712 (em 3,811 textos)
Dispersão	DPnorm: 0.61; Juilland's D: 0.96	DPnorm: 0.49; Juilland's D: 0.97
Outliers	textos 1467, 3227, 1213	textos 92, 55, 984

A Tabela 1 mostra os resultados obtidos com relação à grande área dos artigos acadêmicos sob análise. Considerando as frequências relativas, é possível observar que, enquanto a área de ciências exatas (CE) e a de ciências humanas (CH) têm preferência pela forma simples, a área de ciências biológicas (CB) apresenta uma maior ocorrência da forma perifrástica, em comparação com a forma simples. Esse comportamento diferenciado de CB será reforçado adiante quando será apresentado um cruzamento entre as grandes áreas e o Qualis-CAPES das revistas. Destacamos,

também, que a perífrase é mais frequente em CB e menos frequente em CE, ficando CH em uma posição intermediária.

Tabela 1 – Grande área e tipo de construção

	Palavras no subcorpus	Ocorrências	Em quantos textos	Frequência relativa
Futuro Simples				
Exatas	835785.00	183.00	60 de 187	218.96
Humanas	15436402.00	4400.00	1,149 de 1,538	285.04
Biológicas	7988919.00	848.00	503 de 2,086	106.15
Forma Perifrástica				
Exatas	835785.00	76.00	32 de 187	90.93
Humanas	15436402.00	2046.00	733 de 1,538	132.54
Biológicas	7988919.00	1663.00	556 de 2,086	208.16

Quanto ao Qualis-CAPES das revistas, conforme mostram as Tabelas 2 e 3, há uma frequência maior da forma simples no grupo de estratos A1-B1 em comparação ao grupo de estratos B2-B5. O mesmo ocorre em relação à forma perifrástica, ou seja, ao contrário do que era esperado, a forma perifrástica é preferida no grupo de estratos mais altos (A1-B1).

Tabela 2 – Qualis-CAPES e tipo de construção (forma simples)

Forma simples	Palavras na categoria	Ocorrências	Número de textos	Frequência normalizada
A1	11221604.00	3299.00	853 de 1,127	293.99
A2	9103869.00	1646.00	694 de 1,635	180.8
B1	1966478.00	182.00	71 de 576	92.55
A1 - B1	22291951.00	5127.00	1547 de 3338	229.993328
B2	459271.00	76.00	24 de 124	165.48
B3	971991.00	44.00	32 de 299	45.27
B5	530650.00	184.00	38 de 49	346.74
B2 - B5	1961912.00	304.00	94 de 472	154.950885

Tabela 3 – Qualis-CAPES e tipo de construção (forma perifrástica)

Forma perifrástica	Palavras na categoria	Ocorrências	Número de textos	Frequência normalizada
A1	11221604.00	1381.00	507 de 1,127	123.07

A2	9103869.00	2128.00	694 de 1,635	233.75
B1	1966478.00	100.00	45 de 576	50.85
A1-B1	22291951.00	3609.00	1246 de 3338	161.8970004
B2	459271.00	33.00	13 de 124	71.85
B3	971991.00	34.00	26 de 299	34.98
B5	530650.00	109.00	36 de 49	205.41
B2 - B5	1961912.00	176.00	75 de 472	89.7084069

A fim de detalhar e compreender melhor os resultados apresentados até aqui, optamos por fazer um cruzamento entre a grande área e o Qualis-CAPES das revistas, levando em consideração apenas a forma perifrástica. Os resultados são apresentados na Tabela 4. No grupo de estratos mais altos (A1-B1), observa-se o mesmo comportamento obtido no resultado geral, como mostra a Tabela 1: CB tem a maior frequência de perífrase, CE, a menor frequência, e CH encontra-se em uma posição intermediária (CB > CH > CE). No grupo de estratos mais baixos, entretanto, a frequência da perífrase sofre modificações entre as áreas: CH > CE > CB.

Tabela 4 – Cruzamento entre grande área e Qualis-CAPES

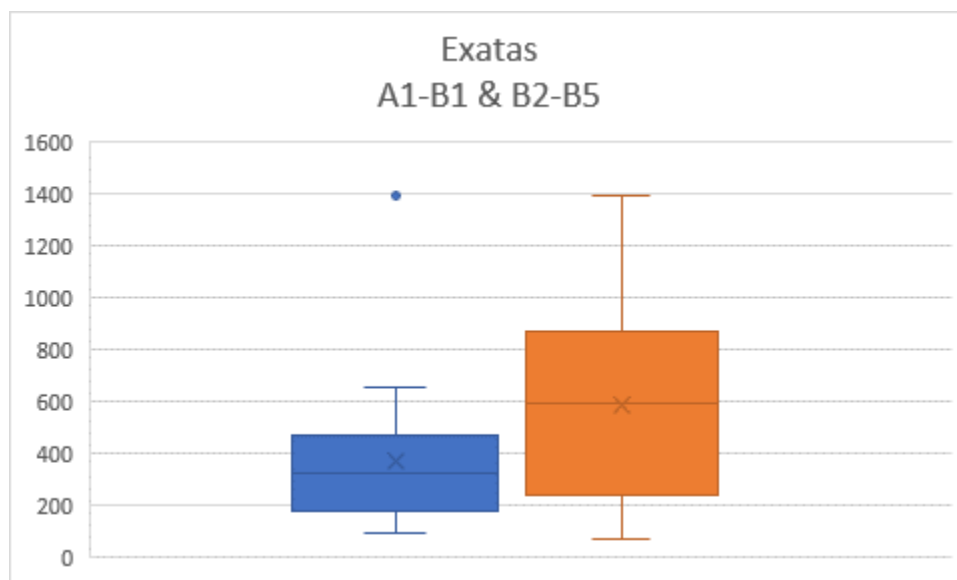
A1 - B1	Distribuição de ocorrências por busca “[lemma="ir" & word!="iria" %c][[pos="VERB.Inf.*"]”, restrita a textos que se adequam ao critério “Qualis-CAPES da área: A1 ou A2 ou B1”, retornou 3,609 correspondências em 1,246 textos diferentes			
	Palavras na categoria	Ocorrências	Número de textos	Frequência normalizada
Exatas	432,462	44	20 de 69	101.74
Humanas	14,905,752	1,937	697 de 1,489	129.95
Biológicas	6,953,737	1,628	529 de 1,780	234.12
Total:	22,291,951	3,609	1,246 de 3,338	161.9
B2 - B5	Distribuição de ocorrências por busca “[lemma="ir" & word!="iria" %c][[pos="VERB.Inf.*"]”, restrita a textos que se adequam ao critério “Qualis-CAPES da área: B2 ou B3 ou B5”, retornou 176 correspondências em 75 textos diferentes			
	Palavras na categoria	Ocorrências	Número de textos	Frequência normalizada
Exatas	396,080	32	12 de 117	80.79
Humanas	530,650	109	36 de 49	205.41
Biológicas	1,035,182	35	27 de 306	33.81
Total:	1,961,912	176	75 de 472	89.71

Analisando as áreas separadamente, observamos que CE e CB têm uma frequência maior de perífrase no grupo de estratos mais altos em comparação ao grupo de estratos mais baixos, enquanto CH mostra um resultado diferente (e

compatível com nossa hipótese inicial): maior frequência de perífrase no grupo de estratos mais baixos. O teste do qui-quadrado foi aplicado, e os resultados apontam que a diferença de frequência em CH e CB é extremamente significativa (valor de p de 0.000003645626 e 1.898653×10^{-39} , respectivamente), já em CE a diferença não é significativa (valor de p de 0.3789793)⁴. O teste, então, confirma que, em CH, a perífrase é mais usada no grupo de estratos mais baixos, e, em CB, esta forma é mais usada no grupo de estratos mais altos, enquanto em CE não parece haver diferença entre os grupos.

A fim de visualizarmos melhor a distribuição dos dados por grande área, foram gerados os *boxplots* apresentados nas Figuras 2 a 4. A Figura 2 contém o *boxplot* referente a CE. Como pode ser observado, há um *outlier* (texto 1). Apesar de o teste do qui-quadrado não ter sido significativo, avaliando a distribuição por textos, pode-se verificar que o uso do futuro perifrástico é mais consistente em revistas com Qualis-CAPES do grupo de estratos mais baixos.

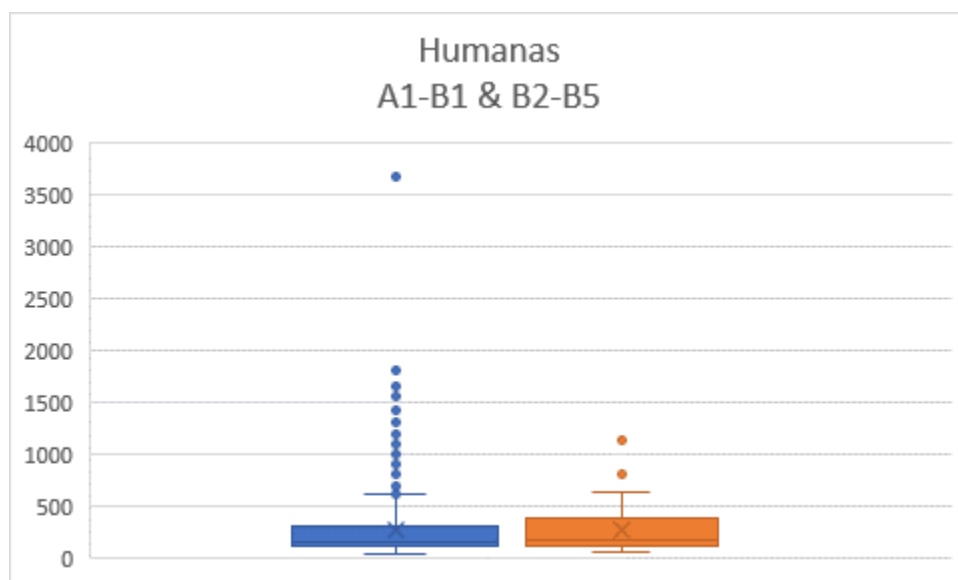
Figura 2 – *Boxplot* de CE



Na Figura 3, é apresentado o *boxplot* de CH. Como podemos observar, há vários *outliers* (como, por exemplo, os textos 3227, 3192, 3384, 2874, 3389, 3186, 2321) com alto uso da forma perifrástica em revistas de Qualis-CAPES alto.

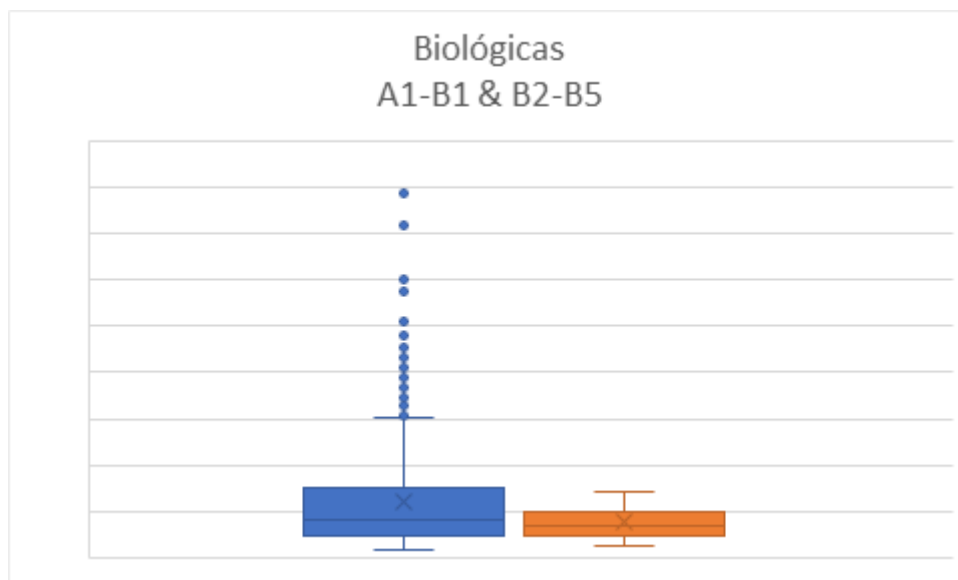
⁴ O nível de significância considerou o seguinte: >0,05 = não significativa; 0,01 a 0,05 = significativa; 0,001 a 0,01 = muito significativa; <0,001 = extremamente significativa.

Figura 3 – *Boxplot* de CH



A Figura 4 mostra o *boxplot* de CB. Como em CH, há vários *outliers*. Mesmo assim, o *boxplot* permite observar que CB usa mais a forma perifrástica em revistas com Qualis-CAPES do grupo de estratos mais altos.

Figura 4 – *Boxplot* de CB



Os *boxplots*, então, permitem verificar, mais uma vez, o comportamento diferenciado de CB: esta é a única grande área em que a forma perifrástica é preferida no grupo de estratos mais altos. CH, apesar da grande quantidade de *outliers*, e CE,

apesar de o teste do qui-quadrado não confirmar tal resultado, preferem fazer uso da perífrase no grupo de estratos mais baixos.

Tais resultados parecem mostrar que, entre as duas variáveis consideradas neste trabalho, o que realmente influencia a seleção da perífrase é a grande área, uma vez que não há resultados uniformes quanto à atuação do Qualis-CAPES da revista. Cumpre-se, desta forma, nosso objetivo de descrever a variação do futuro do presente em artigos acadêmicos.

Considerações finais

Os resultados obtidos neste trabalho evidenciam que as grandes áreas de ciências exatas, ciências humanas e ciências biológicas exercem influência na realização do futuro simples em PB em artigos acadêmicos. Como foi mostrado, enquanto CH e CE favorecem a forma perifrástica em estratos mais baixos, CB a favorece em estratos mais altos. Tal evidência aponta para a possibilidade de um aprofundamento futuro da análise a partir da observação das diferentes áreas contempladas em cada grande área.

Para além do objetivo descritivo de nosso trabalho, apontamos, na sequência, outras questões que merecem ser investigadas futuramente a fim de buscar possíveis explicações para a atuação das grandes áreas na variação do futuro do presente. Quanto a CE, a maior frequência de perífrase no grupo de estratos mais baixos pode ser resultado de um processo de revisão mais atento e cuidadoso que leva a uma mais frequente opção pela forma conservadora e padrão em revistas de estratos mais altos. Isso evidenciaria, assim, um maior conservadorismo linguístico das revistas de estratos mais altos. Uma possibilidade para verificar tal hipótese, que extrapola os limites deste trabalho, seria verificar se há uma diferença nas etapas de revisão pelas quais os artigos passam a depender do estrato em que a revista se enquadra.

Quanto a CH, pode-se aventar a possibilidade de um conflito entre um conservadorismo linguístico, por um lado, e uma relação menos prescritiva com a língua, por outro. Evidências de tal conflito podem ser encontradas na posição intermediária ocupada por CH quanto ao uso da perífrase no grupo de estratos mais altos, conforme a Tabela 4, e pela presença de vários *outliers*, como mostra a Figura 3. O polo do conservadorismo pode ser motivado por uma maior atenção às formas linguísticas decorrente do fato de que, em várias áreas dentro de CH, o texto, escrito e falado, é um importante objeto de análise e fonte de dados. O polo menos prescritivo,

por sua vez, pode ser fruto de uma possível maior sensibilidade linguística, derivada também de um trabalho sobre o texto como objeto de análise e fonte de dados, o que permitiria perceber que a forma perifrástica, apesar de inovadora e não padrão, não é estigmatizada.

Uma das formas possíveis de verificar tal hipótese seria a realização de testes de percepção linguística com pesquisadores de áreas diferentes. Conforme Oliveira (2011b, p. 3), “os testes de percepção/atitude permitem captar generalizações que revelam possíveis relações entre avaliação subjetiva e motivação para a seleção e o uso consciente de formas linguísticas alternativas”. Como mostra Eckert (2005), o estilo pode ser interpretado como um recurso para a construção de identidades. Nesse sentido, os testes de percepção poderiam trazer indícios para a opção pela forma perifrástica como uma maneira de o sujeito mostrar uma identificação com uma concepção menos prescritiva de língua. Severo (2014, p. 39-40) afirma que “os gêneros das esferas acadêmicas tendem a impor fortes restrições sobre as escolhas linguísticas a partir das valorações que os sujeitos nessas esferas compartilham sobre os usos linguísticos”. Assim, os testes de percepção poderiam revelar se áreas diferentes atribuem valores diferentes às variantes de futuro do presente.

Com relação a CB, não temos, por hora, hipóteses que possam justificar sua preferência pela forma perifrástica nos estratos mais altos.

Além do que foi apontado até aqui, outras questões merecem uma investigação, como o gênero/sexo do pesquisador. Segundo um postulado clássico da sociolinguística laboviana, mulheres tendem a liderar a inovação no caso de mudanças abaixo do nível da consciência social⁵ (LABOV, 1990). Os resultados de Oliveira (2006) apontam para uma preferência da perífrase entre as mulheres (peso relativo de 0,70) em comparação aos homens (peso relativo de 0,37). No caso das grandes áreas, seria possível verificar se há uma concentração de determinado gênero/sexo em uma área, e isso poderia trazer contribuições para a discussão sobre as diferenças encontradas entre as áreas.

Para finalizarmos o apontamento de possibilidades de continuidade da pesquisa, destacamos que a observação das variáveis linguísticas apresentadas acima (paradigma verbal, papel temático do sujeito e natureza semântica do verbo) pode apontar para certos padrões em uma área, como, por exemplo, uma maior

⁵ Para uma problematização da variável gênero/sexo nos estudos sociolinguísticos, cf. Freitag e Severo (2015).

frequência de verbos que indicam processo, o que pode contribuir para um maior uso da perífrase.

Referências

ANTHONY, L. AntConc (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University, 2019. Disponível em: <https://www.laurenceanthony.net/software>.

AVELAR, J. De verbo funcional a verbo substantivo: uma hipótese para a supressão de *haver* no português brasileiro. *Letras de Hoje*, v. 41, n. 1, p. 49-74, 2006.

BESTGEN, Y. Getting rid of the Chi-square and Log-likelihood tests for analysing vocabulary differences between corpora. *Quaderns de Filologia: Estudis Lingüístics*, n. 22, p. 33-56, 2017.

BORTONE, E. M.; SIMON, M. L. M. Do sintético para o analítico: uma tendência em três línguas neolatinas. *Revista Signótica*, v. 6, n. 1, p. 73-90, 1994.

BREZINA, V.; MEYERHOFF, M. A critical review of sociolinguistic generalisations based on large corpora. *International Journal of Corpus Linguistics*, v. 19, n. 1, p. 1-28, 2014.

The British National Corpus, version 3 (BNC XML Edition). Distributed by Bodleian Libraries, University of Oxford, on behalf of the BNC Consortium, 2007. URL: <http://www.natcorp.ox.ac.uk/>

ECKERT, P. *Variation, convention, and social meaning*. Paper presented at the Annual Meeting of the Linguistic Society of America. Oakland CA. Jan 7, 2005.

FIGUEIREDO, J. G. dos S.; OLIVEIRA, J. M. A expressão do futuro verbal em Irará-BA. *Revista (Con)Textos Linguísticos*, v. 7, n. 8, p. 36-50, 2013.

FREITAG, R. M. Ko.; SEVERO, C. G. (orgs.) *Mulheres, linguagem e poder: estudos de gênero na sociolinguística brasileira*. São Paulo: Blucher, 2015.

GONÇALVES, A. Estudo do processo de gramaticalização do verbo *ir*: uma análise diacrônica. *Revista Práticas de Linguagem*, v. 3, n. 2, p. 323-343, 2013.

HARDIE, A. CQPweb - combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics*, v. 17, n. 3, p. 380-409, 2012.

KUHN, T. Z.; FERREIRA, J. P. Building a corpus of written academic texts in Portuguese. Teaching and Language Corpora Conference (TaLC12), Giessen, Germany. Book of Abstracts, 103. Giessen, 2016.

LABOV, William. The intersection of sex and social class in the course of linguistic change. *Language variation and change*, n. 2, p. 205-254, 1990.

MARCUSCHI, L. A. *Da fala para a escrita*. São Paulo: Cortez Editora, 2010.

McENERY, T.; HARDIE, A. *Corpus linguistics: Method, theory and practice*. Cambridge: Cambridge University Press, 2011.

OLIVEIRA, J. M. de. *O futuro da língua portuguesa ontem e hoje: variação e mudança*. 2006. Tese (Doutorado em Língua Portuguesa) - UFRJ, Faculdade de Letras. Rio de Janeiro, 2006. Disponível em: <https://docplayer.com.br/5144665-O-futuro-da-lingua-portuguesa-ontem-e-hoje-variacao-e-mudanca-josane-moreira-de-oliveira.html>. Acesso em: 13 mar. 2019.

OLIVEIRA, J. M. de. A expressão variável do futuro verbal na escrita: Brasil e Portugal em confronto. *Revista da ABRALIN*, v. 10, n. 3, p. 368-383, 2011a.

OLIVEIRA, J. M. de. A variação do futuro verbal em português: teste de percepção/atitude na cidade de Feira de Santana-BA. *Tabuleiro de Letras*, n. 3, p. 1-22, 2011b.

OLIVEIRA, J. M. de; OLINDA, S. R. M. de. A trajetória do futuro perifrástico na língua portuguesa: séculos XVIII, XIX e XX. *Revista da ABRALIN*, v. 7, n. 2, p. 93-117, 2008.

SCHMID, H. Probabilistic Part-of-Speech Tagging Using Decision Trees. *Proceedings of the International Conference on New Methods in Language Processing*, 1994. p. 44-49.

SCLIAR-CABRAL, L. Registros e variedades sociolinguísticas. In: GÖRSKI, E. M. et al. (orgs) *Variação estilística: reflexões teórico-metodológicas e propostas de análise*. Florianópolis: Insular, 2014. p. 51-63.

SEVERO, C. G. Estilo, variação e discurso. In: GÖRSKI, E. M. et al. (orgs) *Variação estilística: reflexões teórico-metodológicas e propostas de análise*. Florianópolis: Insular, 2014. p. 31-49.

TESCH, L. M.; YACOVENCO, L. C.; SCHERRE, M. M. P. Variação e mudança na fala e escrita: caminhos e fronteiras. *Revista (Con)Textos Linguísticos*, v. 8, n. 10.1, p. 87-106, 2014.

Recebido em: 02 mar. 2020

Aceito em: 18 abr. 2020