LINGUÍSTICA DE CORPUS E FRASEOLOGIA:UM ESTUDO DAS COLOCAÇÕES COM "FEITO" EM GRANDE SERTÃO: VEREDAS

CORPUS LANGUAGE AND PHRASEOLOGY: A STUDY OF PLACES WITH "DONE" IN GRANDE SERTÃO: VEREDAS

LINGÜÍSTICA DE CORPUS Y FRASEOLOGÍA: UN ESTUDIO DE LAS COLOCACIONES CON "HECHO" EN GRANDE SERTÃO: VEREDAS

Ana Paula Correa PIMENTA[*]
Ariel NOVODVORSKI[**]

Resumo: Com base no pressupostos da Linguística de Corpus, este artigo propõe uma análise léxico-gramatical e semântica das unidades fraseológicas, mais especificamente as colocações com a palavra "feito" que denotam sentido de comparação, na obra **Grande Sertão: Veredas** (1994)[1] de João Guimarães Rosa. Para tanto, buscou-se fundamentar em Corpas Pastor (1996) para a discussão acerca das unidades fraseológicas, em Halliday (1994) sobre as relações entre o léxico, a gramática e o estrato semântico, e em Berber Sardinha (2004, 2009) a respeito da Linguística de *Corpus.*Os resultados deste estudo ressaltam a importância de se considerar os elementos linguísticos como parte preponderante da história e da cultura de um povo, bem como a relevância de se evidenciar a considerável contribuição que a Linguística de *Corpus* tem trazido ao estudo da Fraseologia (especialmente, com *corpus* literário) no Brasil.

Palavras-chave: Fraseologia; Colocações; Linguística de Corpus; Grande Sertão: Veredas.

Abstract: Based on the assumptions of Corpus linguistics, this article proposes a lexical-grammatical and semantic analysis of the phraseological units, more specifically the positions with the word "done" that denote a sense of comparison, in the work **Grande Sertão: Veredas** (1994) de João Guimarães Rosa. In order to do so,

[*] Doutoranda em Estudos Linguísticos pela Universidade Federal de Uberlândia (UFU) e professora substituta de Língua Portuguesa e Linguística na Universidade Federal de Goiás – Regional Catalão. Contacto: anapaula.apcp@hotmail.com.

[**] Orientador da tese de Doutorado e coautor deste trabalho. Professor Adjunto no Curso de Graduação em Letras e do Programa de Pós-Graduação em Estudos Linguísticos (PPGEL) do Instituto de Letras e Linguística (ILEEL) da Universidade Federal de Uberlândia (UFU). Contacto: ariel_novodvorski@yahoo.com.br.

[1] Utilizou-se a edição de 1994 da obra Grande Sertão: Veredas, publicada primeiramente em 1956.

it was sought to base in Corpas Pastor (1996) for the discussion about the phraseological units, in Halliday (1994) on the relations between the lexicon, the grammar and the semantic stratum, and in Berber Sardinha (2004, 2009) a Respect of Corpus Linguistics. The results of this study emphasize the importance of considering linguistic elements as a preponderant part of a people's history and culture, as well as the relevance of highlighting the considerable contribution Corpus Linguistics has brought to the study of Phraseology (especially with Literary corpus) in Brazil.

Keywords: Phraseology; Placements; Linguistics of Corpus; Grande Sertão: Veredas.

Resumen: Con base Con base en los presupuestos de la Lingüística de Corpus, este artículo propone un análisis léxico-gramatical y semántica de las unidades fraseológicas, más específicamente las colocaciones con la palabra "hecho" que denota sentido de comparación, en la obra Grande Sertão: Veredas (1994) de João Guimarães Rosa. Para ello, se buscó fundamentar en Corpas Pastor (1996) para la discusión acerca de las unidades fraseológicas, en Halliday (1994) sobre las relaciones entre el léxico, la gramática y el estrato semántico, y en Berber Sardinha (2004, 2009) a respecto de la Lingüística de Corpus. Los resultados de este estudio resaltan la importancia de considerar los elementos lingüísticos como parte preponderante de la historia y la cultura de un pueblo, así como la relevancia de evidenciar la considerable contribución que la Lingüística de Corpus ha traído al estudio de la Fraseología (especialmente, con corpus literario) en Brasil

Palabras clave: Fraseología; Colocaciones; Lingüística de *Corpus*; Grande Sertão: Veredas.

Introdução

Há de se considerar que o rápido crescimento das pesquisas baseadas em *corpus*, desde a década de 1990, tem influenciado de maneira significativa os estudos linguísticos, trazendo grandes contribuições em diversas áreas, principalmente nas áreas da Linguística Aplicada como: Ensino de Línguas, Tradução, Análise do Discurso e Lexicografia. Neste trabalho, buscou-se apresentar uma pequena demonstração da grande contribuição que a Linguística de *Corpus* tem trazido para análises linguísticas com *corpus* literário no Brasil.

Propôs-se realizar, neste estudo, uma análise léxico-gramatical e semântica das unidades fraseológicas, mais especificamente as colocações com a palavra "feito" com sentido de comparação, em **Grande Sertão: Veredas** (1994) de Guimarães Rosa. Para tal, aplicou-se as ferramentas do Programa *WordSmith Tools*®, versão 6,0

(SCOTT, 2012) para a seleção, descrição e análise dos itens em questão, o que propiciou um estudo mais completo do *corpus*.

A escolha dessa obra como *corpus* de análise se deve por sua relevância na literatura brasileira e por sua riqueza linguística, já que Guimarães procurou fixar no plano literário a geografia física e humana do seu estado natal, isto é, a vida rural no interior de Minas Gerais, demonstrando os aspectos históricos, políticos, socioculturais e, especialmente, linguísticos. Portanto, pode-se afirmar que o repertório lexical do *corpus* está assentado nos interesses da pesquisa, pois oferece profuso vocabulário com termos e expressões arraigadas de sentimentos, desejos, e emoções vivenciadas não só pelo autor, ora narrador, como também pelos personagens, sertanejos oprimidos pela seca, pela rusticidade e precariedade do sertão.

Desse modo, com base nos estudos de Corpas Pastor (1996), Halliday (1994) e Berber Sardinha (2004, 2009), os procedimentos da análise envolvem não só um estudo do estrato léxico-gramatical dos itens em questão, como também do estrato semântico, o que propicia uma compreensão mais abrangente dos fenômenos estudados.

Até o momento não se tem conhecimento de que há trabalhos voltados para a análise dos processos mentais em Grande Sertão: Veredas. Portanto, espera-se que este trabalho possa proporcionar uma significativa contribuição para os estudos linguísticos e, ainda, servir como subsídio para pesquisas posteriores.

Fundamentação teórica

Para que se possa dar um tratamento adequado ao exame dos dados e fundamentar as discussões e os resultados alcançados, é necessário lançar mão de teorias que envolvam questões linguísticas de ordem lexicogramatical e semântica. Por essa razão, buscou-se fundamentar teoricamente dentro dos princípios e possíveis aplicações da Linguística de *Corpus* ao estudo da Fraseologia.

Considerando a diversidade conceitual dos termos "Fraseologia" e "unidades fraseológicas", buscou-se apoiar na perspectiva teórica de Corpas Pastor, professora, pesquisadora e linguista espanhola de grande expressividade no campo da Fraseologia, cujos trabalhos (dentre eles se destaca o *Manual de fraseologia española*-1996) têm sido uma referência na área desde 1996.

Conforme Corpas Pastor (1996,) a Fraseologia é um ramo da Linguística que tem como objeto de estudo as unidades fraseológicas (doravante UFs), isto é, aquelas "formadas por mais de duas palavras gráficas em seu limite inferior, cujo limite superior situa-se no nível da oração composta" (CORPAS PASTOR, 1996, p. 20, tradução nossa). São combinações de palavras que ocorrem com frequência na língua, seja pelo princípio da livre escolha (itens linguísticos disponíveis na língua) ou pelo princípio idiomático (padrões ou estruturas fixas na língua).

Nas palavras de Sinclair (1991, p. 110, tradução nossa) "o princípio idiomático é o de que o usuário de uma língua tem à sua disposição um grande número de sintagmas semi-pré-construídos que constituem escolhas únicas, mesmo que pareçam ser analisáveis em segmentos". Em outras palavras, "o significado do todo é diferente da soma dos significados das partes" (WELKER, 2004, p. 165). A expressão "bater as botas", por exemplo, trata-se de um fraseologismo idiomático ou idiomatismo, uma vez que não se refere literalmente ao ato de bater a bota uma na outra, mas possui um sentido específico: significa "morrer". Já o princípio da livre escolha caracteriza-se pela "convencionalidade", ou seja, são as possibilidades de combinações lexicais que o sistema nos oferece e, por isso, pode ser considerado como um ato de "liberdade relativa", motivada pela necessidade do falante em situações reais de comunicação. Diferentemente do princípio idiomático, o sentido do sintagma por ser entendido composicionalmente, pelo significado individual das partes.

São consideradas UFs: as colocações, os binômios, as expressões idiomáticas, os provérbios, as locuções verbais e nominais, as máximas, os aforismos, as fórmulas e as parêmias. De um ponto de vista específico, Corpas Pastor (1996), classifica as UFs em: a) colocações, b) locuções, c) enunciados fraseológicos (parêmias, enunciados de valor específico, citações, fórmulas rotineiras: de saudação, de transição, expressivas, comissivas, diretivas e assertivas).

Para fins do presente estudo, convém esclarecer o que são colocações. Segundo Corpas Pastor (1996), as colocações não constituem uma oração e incluem combinações lexicais fixas somente na norma. São "unidades fraseológicas formadas por, pelo menos, duas unidades lexicais em relação sintática, que não constituem por si mesmas atos de fala ou enunciados" (CORPAS PASTOR, 1996, p. 66,

tradução nossa). "É uma associação entre itens lexicais, ou entre o léxico e campos semânticos" (BERBER SARDINHA, 2004, p. 40). Depreende-se dessas definições que as colocações são combinações de palavras com alta frequência de uso e de co-ocorrência. Dentre uma grande quantidade de combinações possíveis fixadas na norma, o falante escolhe a combinação mais apropriada às suas intenções discursivas. Eis alguns exemplos: fazer uma pergunta; cometer erros; pôr à prova.

Como se trata de constructos léxico-gramaticais e semânticos, não é possível analisá-los a partir de uma Gramática Tradicional ou Normativa. A visão tradicional de que a língua é composicional e que conhecendo o léxico e as regras gramaticais é possível ter absoluto domínio dela, é superficial e incongruente. Para que se possa, de fato, fazer uma análise mais precisa e coerentedas colocações, é preciso buscar subsídios teóricos e gramaticais que não só estabeleçam relações entre o léxico, a gramática e o estrato semântico das unidades, mas, sobretudo, perceba a língua como um sistema probabilístico, moldado e determinado pelo uso. Por essa razão, buscou-se fundamentar na perspectiva teórico-gramatical de Halliday (1994), linguista britânico que utilizou uma abordagem de análise gramatical denominada *Gramática de Estala e Categorias*, que foi sistematizada em 1985 na obra *Na Introductionto Functional Grammar*, revisada pelo autor em 1994 e, posteriormente, ampliada em 2004, com a colaboração de Christian M. I. O. Matthiessen e, em 2014, também nessa mesma parceria.

Para Halliday (1994), a língua é um sistema semiótico moldado e determinado pelo uso; é um potencial de significados que são representados por atos de fala e, por conseguinte, se desenvolve e se modifica para satisfazer as necessidades humanas. Segundo o pesquisador, foram os usos linguísticos que, ao longo de milhares de gerações, deram forma a esse sistema.

Partindo desse pressuposto, Halliday (1994) defende que para se realizar uma análise de qualquer língua, deve-se estabelecer relações entre o léxico, a gramática (traços de forma) e os estrato semântico (traços de significados). Além disso, deve-se também levar em conta que cada item léxico se classifica em função dos traços que possui, ou seja, cada unidade se caracteriza pelas funções que ela pode ocupar na sentença.

Dessa maneira, preocupar-se em realizar análises sob esse enfoque significa considerar os diferentes usos que se faz da linguagem em seus variados níveis ou estratos, bem como compreender os significados que cada estrutura léxicogramatical possui de acordo com o contexto. Em outras palavras, analisar um *corpus* implica um trabalho laborioso de descrição de dados, os quais estão inseridos em sequências formais dotadas de significados de acordo com as associações que realizam dentro de cada unidade linguística. Sendo assim, propor uma análise das colocações com esse aporte teórico implica compreender que qualquer uso linguístico que se constitua em um *corpus* pode representar aspectos léxicogramaticais e semânticos variados de acordo com o contexto em que está inserido.

Em relação à LC, importa mencionar que esta surgiu da necessidade de se realizar pesquisas com um grande número de dados, o que se tornou possível com o advento do computador. Não se trata de uma teoria, mas de uma nova abordagem de "exploração da linguagem por meio de evidências empíricas extraídas por computador" (BERBER SARDINHA, 2004, p. 3). Nas palavras de Berber Sardinha (2004, p. 3): "ocupa-seda coleta e da exploração de corpora, ou conjuntos de dados linguísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade linguística".

Vale lembrar que antes do computador, já se fazia o uso de *corpus,* como por exemplo, o Corpus Helenístico criado na Grécia Antiga e os corpora de citações da Bíblia produzidos na Antiguidade e na Idade Média. Nessa época, os dados eram coletados, armazenados e analisados manualmente. Apesar das dificuldades, os estudos baseados em corpora não pararam. No final dos anos de 1950 com a publicação de *Sintactic Structures* de Chomky, houve uma mudança de "paradigma na linguística: saía de cena o empirismo e a sustentação dos trabalhos baseados em corpora, tomando lugar central as teorias racionalistas da linguagem, notadamente a linguística gerativa" (BERBER SARDINHA, 2004, p. 4). Desse período ganham destaque os estudos de Firth (1957) e o *corpus* SEU (Surveyof English Usage), compilado e etiquetado manualmente em 1959.

Com o advento do computador nos anos de 1960 e a queda de prestígio das pesquisas puramente racionalistas, novos *corpora* são compilados, dentre os principais estão: Brown, Birmengham e BNC.

Segundo Berber Sardinha (2004), o corpus Brown é o pioneiro dos corpora eletrônicospor ter nascido em um período ainda desfavorável para os estudos empiristas e, também, pela dificuldade de compilação em computadores mainframe; "é um ponto de referência inevitável em qualquer retrospectiva sobre a LC em nível mundial" (NOVODVORSKI; FINATTO, 2014, p. 7).

A popularização dos estudos com corpora ocorreu nos anos de 1980 com o aparecimento dos computadores pessoais. Desde então, a LC tem trazido grandes contribuições para os estudos linguísticos em diferentes áreas. No Brasil, o seu aparecimento se deu com a publicação do livro *Linguística de Corpus* em 2004, de Tony Berber Sardinha, maior expoente da LC no Brasil, hoje professor associado do Departamento de Linguística e do Programa de Estudos Pós-Graduados em Linguística Aplicada e Estudos da Linguagem da Pontifícia Universidade Católica de São Paulo (PUC-SP).

Fato a ser considerado é que a LC contemporânea, segundo Berber Sardinha (2004),caracteriza-se como uma abordagem empirista de exploração da linguagem, que é vista como sistema probabilístico por não se apresentar de forma aleatória. Isto significa dizer que os traços linguísticos não ocorrem com a mesma frequência, uma vez que as regularidades (padrões) estão associadas às características linguísticas e contextuais. Por isso é comum dizer que a linguagem é padronizada, ou seja, há uma regularidade expressa na recorrência sistemática de unidades coocorrentes de várias ordens: lexical, gramatical, semântica, dentre outras. Nas palavras de Berber Sardinha (2004, p. 31): "A padronização se evidencia pela recorrência, isto é, uma colocação, coligação ou estrutura que se repete significativamente e mostra sinais de ser, na verdade, um padrão lexical ou léxico-gramatical".

Por fim, a LC é uma área que vem ganhando cada vez mais destaque nas análises linguísticas nos seus mais diversos ramos, principalmente nas áreas da Linguística Aplicada como: Ensino de Línguas, Tradução, Análise do Discurso e Lexicografia. Dentre os vários software existentes que auxiliam o linguista de *corpus*, ganha destaque o Word Smith Tools, software de referência nos estudos linguísticos, particularmente os lexicais, cujas principais ferramentas são: "Word List, Concord e KeyWords" (BERBER SARDINHA,

2009). Tais ferramentas serão utilizadas nesse estudo, o que propicia uma análise mais vigorosa e precisa dos processos.

Diante do exposto, pode-se afirmar que há uma relação ideal de simbiose entre os pressupostos teóricos de Corpas Pastor (1996), Halliday (1994) e a LC. Apesar de Halliday (1994) não se denominar linguista de *corpus*, as suas visões de linguagem se encaixam perfeitamente nos preceitos da LC e servem como arcabouço teórico para qualquer trabalho que se dedica à análise de corpora. Portanto, propor uma análise das colocações com "feito" em **Grande Sertão: Veredas** com esse aporte teórico significa dar um tratamento adequado ao trabalho de descrição e teorização dos dados. Afinal, apresentar os traços lexicogramaticais e semânticos dos itens em questão de nada vale se divorciados de uma adequação empírica.

Procedimentos metodológicos

Para a seleção, descrição e análise dos dados foram adotados os seguintes procedimentos: 1) inicialmente, realizou-se a busca do *corpus* na Internet por meio do Google e, após encontrado, converteu-se o arquivo para txt, formato adequado para a leitura com o programa *WordSmith Tools*® (WST) em sua versão 6.0; 2) em seguida, fez-se a limpeza do *corpus*, retirando os elementos indesejáveis; 3) utilizando a ferramenta *WordList,*conseguiu-se (através da função *Statistics*) um *corpus* com as seguintes características: tokens (itens)- 191.990, types (formas)- 19.171e type/tokenratio (razão forma/item) – 10,03, como mostra a figura a seguir:

Figura 1: Levantamento dos dados estatísticos por meio da ferramenta *WordList– Statistics*



4) logo após, realizou-se a seleção dos dadospor meio da lista de frequência (*WordList- frequency*), verificando quais unidades lexicais eram mais recorrentes e significativas no *corpus:* observou-se que as palavras"cara", "boa" "bicho" "pé e "feito" apareciam com um grande número de ocorrências, sendo assim, optou-se por realizar a análise somente das colocações com a palavra "feito", uma vez que melhor evidenciam as particularidades linguísticas da obra, demonstrando as experiências vivenciadas pelo autor, ora narrador, no sertão mineiro. Neste ponto é importante destacar que foram consideradas somente as acepções em que o "feito" exercia a função de conjunção ou de conectivo comparativo; 5) das 270 ocorrências encontradas, conforme mostra a Figura 2, buscou-se, por meio da ferramenta *Concord* (concordance), todas as colocações com um sentido de comparação, como se pode observar na Figura 3. Aquelas em que a palavra "feito" não exercia a função de conjunção ou conectivo comparativo foram excluídas, restando, assim, 232 colocações para análise.

Figura 2: lista de frequência (*WordList- frequency*)



N	Word	Freq.	%	Texts	%	Lemmas	Set
76	HERMÓGENES	283	0,15	1	100,0(		
77	TODO	283	0,15	1	100,0(		
78	DAS	282	0,14	1	100,0(		
79	OUTROS	281	0,14	1	100,0(		
80	ALI	277	0,14	1	100,0(		
81	TÃO	274	0,14	1	100,0(		
82	FEITO	270	0,14	1	100,0(		
83	MEIO	270	0,14	1	100,0(		
84	PODIA	270	0,14	1	100,0(		
85	TEMPO	264	0,14	1	100,0(		
86	ELES	263	0,14	1	100,0(		
87	SEI	262	0,13	1	100,0(		
88	VER	262	0,13	1	100,0(		
89	MÃO	259	0,13	1	100,0(		

Figura 3: Lista das colocações geradas por meio das linhas de concordância (*Concord-collocates*)



Utilizando a função *collocates* foi possível visualizar todas as colocações (com a palavra "feito") presentes no *corpus* e, observando as linhas de concordâncias, pôde-se selecionar e excluir aquelas em que a palavra "feito" exercia a função de verbo e não de comparação. 6) por fim, procedeu-se à descrição e análise dos dados. Com o propósito de apresentar um estudo mais preciso e coerente dessas unidades buscou-se analisar além do estrato léxico-gramatical, o estrato semântico. Acredita-se que, estabelecendo as relações entre os aspectos lexicogramaticais e semânticos das unidades fraseológicas, será possível evidenciar, com mais facilidade, as características marcantes da obra, corroborando para uma compreensão mais contundente do texto.

Análise das colocações com a palavra "feito" no *Corpus*

Nesta seção serão apresentados os dados hauridos do *corpus* em uma tabela para fins de sistematização e análise. Importa

mencionar que esta análise foi realizada numa perspectiva guiada pelo *corpus.*

Das diversas construções presentes no *corpus*, muitas são repetidas, o que torna imprescindível, dado o escopo da pesquisa, fazer um recorte, disponibilizando apenas as ocorrências que correspondam aos diferentes usos com "feito". Sendo assim, pararealizar a amostragem dos dados (as diferentes ocorrências, a recorrência de cada uma e a sua coocorrência) e, ainda facilitar a visualização dos aspectos lexicogramaticais e semânticos das colocações,buscou-se apresentar, a seguir, uma tabela com os seguintes resultados:

Tabela 1: resultados da análise

Tipos (Estrutura)	Exemplo de ocorrência	Frequência
Feito + substantivo	Ex. 1: Mas nós passávamos, **feito flecha, feito faca, feito fogo.**	**71**
Feito + artigo + subst.	Ex. 2: Ele terrivelmente todo pensava **– feito o carro e os bois** se desarrancando num atoleiro.	**63**
Feito + verbo	Ex. 3: E, todos, tinha vez eu achava que queria-bem o meu pessoal, **feito fossem** irmãos meus, da semente dum pai na madre de uma mãe gerados num tempo.	**22**
Feito + oração	Ex.4: De não pitar, me vinham uns rangidos repentes, **feito eu tivesseira de todo o mundo**.	**20**
Feito + prep. + subst.	Ex. 5: **Feito num traslo**copiado de sonho**,** eu preparava os distritos daquilo, que, no começo achei que era fantasia.	**15**
Feito + conjunção	Ex.6: Saiba o senhor: eu estava ali, assim em padastro de todos, de do ar, de rechego, **feito que** em jirau-de-espera, para castigar onça assassinã.	**12**
Feito + pronome	Ex. 7: Assim que, então, os de lá – os judas – não deviam de ser somente os cachorros endoidecidos; mas, em tanto, pessoas, **feito nós**, jagunços em situação.	**10**
Feito + locução adjetiva	Ex. 8: [...] para a beira do córrego ia indo, nu, nu, **feito perna de jaburu.**	**5**
Feito + adj + subst.	Ex. 9: Arrasar que vem de para onde não se olha: **feito forte sol**; e vem como sol nascendo!	**3**
Feito + locução verbal	Ex.10: Eu leve, leve, **feito de poder correr** o mundo ao redor.	**2**
Feito + prep. + pron.	Ex. 11: Eu tinha receio de que me achassem de coração mole, soubessem que eu não era **feito para aquela** influição, que tinha pena de toda cria de Jesus.	**2**

Feito + pron. + subst.	Ex. 12: **Feito aquele Luzié**, que cantava sem mágoas, cigarra de entre-chuvas.	**2**
Feito + pron. + adj.	Ex. 13: Otacília sendo forte como a paz, **feito aqueles largos** remansos do Urucuia, mas que é rio de braveza.	**1**
Feito + loc. adverbial	Ex. 14: E chamou minha atenção para o mato da beira, em pé, paredão, **feito à régua** regulado.	**1**
Feito+ art.+ adj. + subst.	Ex: 15[...] e concebia por ele a vexável afeição que me estragava, **feito um mau amor** oculto – por mesmo isso,	**1**
Feito +art. + pron. + sub.	Ex. 16: Atirava e tanto com qualquer quilate de arma, sempre certeira a pontaria, laçava e campeava **feito um todo vaqueiro**, amansava animal de maior brabeza	**1**
Feito + prep. + verbo	Ex. 17: A lembrança dela me fantasiou [...], **feito entre madrugar e manhecer.**	**1**
TOTAL: 17 estruturas diferentes		**232**

Fontes: os autores

Observando os excertos, percebe-se que a palavra "feito" relaciona-se com diferentes unidades lexicais, formando unidades maiores de sentido. Ora aparece associada a um substantivo, ora a um adjetivo, a um verbo, a uma locução, enfim. São diferentes colocados que ocorrem, maiormente, à direita da palavra-alvo e que se repetem com frequência no *corpus,* compondo sintagmas factualmente possíveis na língua.

Analisando a frequência de cada estrutura, verifica-se que as duas primeiras (FEITO + SUBSTANTIVO/ FEITO + ARTIGO + SUBSTANTIVO) são mais recorrentes que as demais. Isto mostra não só a preferência do autor por esses padrões, como também a própria sistematicidade da língua: as combinações possíveis que o sistema (língua) nos oferece e que são regidas pela gramática. Daí a explicação de algumas palavras ocorrerem sempre com determinada palavras e outras não, por dada forma não ser aceita. No exemplo 2: "Ele terrivelmente todo pensava **– feito o carro e os bois** se desarrancando num atoleiro", o "feito" se combina facilmente com "o carro", mas se fosse "um carro" ou carro um/carro o", soaria estranho, ou seja, não formaria uma unidade de sentido.

Apesar das escolhas lexicais recaírem, em sua maioria, sobre os substantivos, muitas das colocações são construídas com verbos, como no exemplo 3: "E, todos, tinha vez eu achava que queria-bem o

meu pessoal, **feito fossem** irmãos meus, da semente dum pai na madre de uma mãe gerados num tempo". Do ponto de vista sintático, esta construção não é comum e, tampouco, aceita pela gramática normativa, mas isso não significa que é menos importante. Embora pareça destoante, anormal, é preciso ser levada em conta, pois é uma forma típica da linguagem coloquial que possui um potencial sintático e semântico tão amplo e significativo como outras recorrentes na linguagem culta. "**Feito fossem irmãos**" possui um sentido de "como se fossem irmãos meus", trata-se de uma estrutura reduzida que deve ser analisada semanticamente, recuperando o seu sentido a partir do contexto.

Pode-se notar, também, que muitas colocações representam orações completas, como no exemplo 4: "De não pitar, me vinham uns rangidos repentes, **feito eu tivesseira de todo o mundo**." Dependendo da intencionalidade do discurso, a palavra "feito" é associada a um grande número de itens lexicais, formando orações carregadas de significado. Essa é uma característica do nosso sistema, ou seja, para cada situação vivenciada pelo falante, há uma gama de escolhas lexicais que este realiza na tentativa de ajustar-se aos propósitos comunicativos.

Existem, ainda, construções em que as conjunções se relacionam com a palavra-alvo, constituindo uma forma linguística aceitável, como se pode observar no exemplo 6: "Saiba o senhor: eu estava ali, assim em padastro de todos, de do ar, de rechego, **feito que** em jirau-de-espera, para castigar onça assassinã." Outras se colocam com pronomes, como: Ex. 7 "Assim que, então, os de lá – os judas – não deviam de ser somente os cachorros endoidecidos; mas, em tanto, pessoas, **feito nós**, jagunços em situação". Em ambas, de menor frequência, a relação entre cada unidade lexical faz com que a sua representação semântica seja coerente com os eventos reais do discurso. É certo que se não fossem formas aceitáveis tudo o mais ficaria comprometido.

As últimas estruturas contêm apenas duas ou uma ocorrência, mas são tão significativas quantas as demais. Isto revela que há uma preferência por certos padrões, e que as variabilidades de combinações se configuram como estratégias inerentes ao uso da língua. Considerando as estruturas (FEITO + LOCUÇÃO VERBAL): "Eu leve, leve, **feito de poder correr** o mundo ao redor"; e (FEITO +

LOC. ADVERBIAL): Ex. 14: E chamou minha atenção para o mato da beira, em pé, paredão, **feito à régua** regulado; tem-se que a língua é composta por uma grande quantidade de signos linguísticos que se combinam, se colocam uns com outros, e que se constituem como regularidades disponíveis aos falantes. Estes exemplos nos deixa claro que é possível representar os atos de fala de forma criativa e contundente.

Além disso, esse arranjo lexical, tão peculiar de Guimarães, ocorre de forma distinta: depois de ponto: "Ou parar curto quieto. **Feito** os bichos fazem." (linha 19); depois de vírgula: "Sério, quieto, **feito** ele mesmo, só igual a ele mesmo nesta vida." (linha 21); ponto e vírgula: "Ali, por onde eu estava, eu marcava muito suave a mão da morte; **feito** um boiadeiro." (linha 67); dois pontos: "Arrasar que vem de para onde não se olha: **feito** forte sol; e vem como sol nascendo!" (linha 9); travessão: "Do Hermógenes, mesmo, existido, eu mero me lembrava – **feito** ele fosse para mim uma criancinha moliçosa e mijona, em seus despropósitos..." (linha 50). Vê-se, portanto, que a criatividade linguística do autor não se resume apenas às incríveis combinações lexicais, mas também às estruturas sintáticas e semânticas, que na maioria das vezes, fogem dos princípios e explicações da Gramática Tradicional ou Normativa. A palavra feito, por exemplo, não aparece como conjunção ou conectivo comparativo e muitos menos é, segundo as regras gramaticais, utilizada depois de ponto ou associada a um verbo, como no excerto: "E, todos, tinha vez eu achava que queria-bem o meu pessoal, **feito fossem** irmãos meus, da semente dum pai na madre de uma mãe gerados num tempo.

Enfim, das 270 ocorrências encontradas ao todo, se observa que 87% possuem sentido de comparação, o que mostra que as UFs em questão são proeminentes no *corpus.* Contudo, para que se possa, de fato, compreender as colocações em seus aspectos léxico-gramatical e semântico, importa mencionar o contexto em que estão inseridos.

No que tange ao *corpus* da pesquisa vale salientar que **Grande Sertão: Veredas** é um romance regionalista pertencente a terceira geração modernista, publicado em 1956 por João Guimarães Rosa, mineiro, nascido em Cordisburgo em 27 de junho de 1908, considerado um dos escritores de maior expressividade na literatura brasileira. Único livro brasileiro a integrar a lista dos cem melhores de todos os tempos do Clube do Livro de Noruega, dedicado à sua

segunda esposa Aracy de Carvalho Guimarães Rosa, **Grande Sertão: Veredas**, alavancou definitivamente a trajetória artística de Guimarães, fazendo-o alcançar um *status* privilegiado não sóna literatura regional, mas na literatura nacional e, até mesmo, internacional.

Fato a ser considerado é que a produção de **Grande Sertão: Veredas** está profundamente arraigada à situação econômica, política e social da época em que Guimarães viveu em Minas. Na década de 1920, a região era dominada pelos poderosos coronéis, que seguia uma política de proteção aos seus interesses, menosprezando, de forma agressiva e violenta, o humilde trabalhador rural. A implantação do sistema coronelista se deu por várias razões: a precariedade da região imposta por um espaço sertanejo, a baixa produtividade agrária e o isolamento frente ao progresso dos estados que já haviam conquistado uma condição mais cosmopolita e centralizadora.

A supremacia dos coronéis era mantida pelo seu poder econômico, obtido geralmente por meio da posse da terra, e pelo seu poder social, que dava a estes a posse sobre o ser humano. Esse grupo "dominante", além de possuir a propriedade da terra que era o principal meio de sobrevivência, mantinha em seu poder os trabalhadores e os instrumentos de trabalho. É esse quadro de interesses e conflitos que o escritor deixa transparecer em sua narrativa: a vida miserável do sertanejo e a sua luta pela sobrevivência, a exploração do trabalhador rural pelos coronéis e a impotência do ser humano diante da força opressora da natureza, consequência do primitivismo agreste do sertão rústico e isolado.

Percebe-se que a representação do espaço e das personagens não é puramente ficcional, antes o autor faz uso da ficção para revelar a realidade do sertão mineiro, a rusticidade e a precariedade da região, bem como a subjugação do pobre, do camponês. Portanto, pode-se afirmar que esta obra é fruto das experiências pessoais e sociais de Guimarães. A sua convivência com o homem do campo instituiu-lhe uma preocupação singular, que o levou a produzir uma literatura de valorização (das peculiaridades regionais) e de protesto com merecida consagração. Suas histórias, narradas por Riobaldo, apresentam enredos memoráveis e personagens surpreendentes que fazem parte não apenas de um universo ficcional, mas de uma realidade histórica, política e social de um povo.

Os recortes apresentados revelam as emoções e os sentimentos vivenciados pelo autor na figura de Riobaldo, narrador em primeira pessoa e também personagem protagonista da história. Riobaldo narra a sua vida de jagunço a um doutor, demonstrando seus medos, anseios, desejos, reflexões e angústias sobre o seu passado difícil e perigoso no sertão mineiro. Percebe-se que o autor, ao tecer sua narrativa considerada por ele como "autobiografia irracional", utiliza uma linguagem inovadora, por vezes simples e rústica, por vezes culta e poética, perfazendo escolhas lexicais que denotam intenções pessoais. Quer isto dizer que a presença das colocações com "feito" no romance não ocorreu por acaso. Guimarães criou o personagem Riobaldo para, através dele, expor suas experiências vivenciadas no sertão mineiro e suas reflexões como ser humano, como mostra os excertos acima.

Através desses recortes é possível notar que Riobaldo sofre com situações que ocorreram no passado e que ainda fazem parte de seu presente, causando-lhe inquietações e tormento à sua vida. Tais sentimentos expressos por esses itens lexicais revelam o seu perfil psicológico: homem atormentado com várias dúvidas e incertezas sobre a existência ou não do Demônio; a relação entre o Bem e o Mal; o sentimento que experimentou por Diadorim; o sentido de sua vida como jagunço; a busca de uma explicação para a condição humana.

Não obstante, para bem configurar o seu enfoque regionalista, o escritor não só elege o sertão mineiro como cenário de suas narrativas, como também o coloca em uma posição de destaque: como título do livro, caracterizando-o como "Grande Sertão: Veredas", que significa "campos extensos e desabitados". Disso decorre que a intenção em escrever um romance regionalista não se reduz simplesmente em ressaltar as peculiaridades do sertão mineiro ou de preservar os costumes, as tradições e as manifestações linguísticas de um povo, mas tambémde mostrar os problemas políticos e sociais da região sob a forma de denúncia social como meio de ação para a transformação.

Enfim, toda essa descrição e análise objetiva mostrar que para que se possa compreender, de fato, qualquer texto, é preciso estabelecer relações entre os aspectos lexicogramaticais e semânticos dos itens, levando sempre em consideração o contexto em que estão inseridos. Foi exatamente observando tais fatores que se conseguiu uma análise coerente, capaz de revelar não só características

(linguísticas e literárias) marcantes em Grande Sertão: Veredas, como também traços peculiares desse admirável escritor, João Guimarães Rosa, magnífico como ser humano e, sobretudo, como profissional.

Considerações finais

Este trabalho procurou apresentar uma pequena demonstração da grande contribuição que a LC tem trazido ao estudo da Fraseologia. Através desse poderoso instrumento de pesquisa foi possível verificar os padrões mais recorrentes no *corpus*, o número de ocorrências e co-ocorrências das colocações com "feito", bem como as relações entre os aspectos lexicogramaticais e semânticos dos itens em questão. É indubitável que sem esse auxílio seria impossível visualizar todos esses aspectos a olho nu. Daí percebe-se qual útil é a LC para o estudo das UFs.

A partir da análise das colocações, numa perspectiva guiada pelo *corpus*, pôde-se constatar não só a exuberante criatividade linguística de Guimarães, como também a riqueza lexical de **Grande Sertão: Veredas**, uma bela obra de arte, de extremo valor literário e, particularmente, linguístico. Trata-se de um trabalho intencional criado a partir de dados do mundo real, matizado por meio dos recursos próprios da literatura e modulado por meio dos artifícios da linguagem. Razão pela qual possui até os dias de hoje um status singular na literatura brasileira.

Os resultados desta análise convergem para quatro pontos: um de caráter relevante e indissolúvel, que diz respeito à íntima relação entre Fraseologia, a LC e os estudos de Halliday (1994); outro de caráter apenas elucidativo, que diz respeito à intrínseca relação existente entre história, política e literatura; outro de natureza fundamental e eminente, que se refere aos aspectos lexicogramaticais e semânticos dos itens analisados, os quais revelam características marcantes do *corpus* antes não observados; o último de natureza substancial – a relevância da LC para o estudo da Fraseologia e, sobretudo, os grandes avanços que têm trazido e vêm trazendo para os estudos linguísticos em diferentes áreas.

Referências

BERBER SARDINHA, Tony. **Linguística de Corpus**. São Paulo – Barueri: Manole, 2004.

_____. **Pesquisa em Linguística de Corpus com WordSmith Tools**. São Paulo – Campinas: Mercado de Letras, 2009.

CORPAS PASTOR, Gloria. **Manual de fraseología española**. Madrid: Gredos, 1996.

HALLIDAY, Michael Alexander Kirkwood. **Introduction to functional Grammar**. 2. ed. London: Edward Arnold, 1994.

NOVODVORKI, Ariel; FINATTO, Maria José Bocorny. **Linguística de Corpus no Brasil**: uma aventura mais do que adequada. *Letras & Letras*, Uberlândia: v. 30, n. 2, p. 7-16, jul/dez. 2014.

ROSA, João Guimarães. **Grande Sertão: Veredas**. Rio de Janeiro: Editora Nova Aguilar, 1994. Disponível em: < http://stoa.usp.br/carloshgn/files/-1/20292/GrandeSertoVeredasGuimaresRosa.pdf>.Acesso em: 10 out. 2016.

SCOTT, Mike. **WordSmith Tools, version 6.0.** Liverpool: Lexical Analysis Software, 2012.

SINCLAIR, John McHardy. **Corpus, concordance, collocation**. Oxford: Oxford University Press, 1991.

WELKER, Herbert Andréas. **Dicionários** - uma pequena introdução à lexicografia. 2. ed. Brasília: Thesaurus, 2004.

Recebido em: 10/08/2017
Aceito em: 28/09/2017