



Colocações culturais em corpora escrito e oral formadas por relatos de seringueiros acreanos

Cultural collocations in written and oral corpora formed by reports of rubber tappers from acre

**Raimundo Ibernnon Chaves da Silva¹, Adriane Orenha-Ottaiano²
e Maria Emilia Pereira Chanut³**

Resumo: Com base nos preceitos teóricos da Fraseologia e da Linguística de Corpus, este artigo visa a extração e análise das colocações culturais presentes no Corpus Escrito da Vila Japiim (CEVI) e no Corpus Oral do Acre (COAC), formados por relatos de seringueiros acreanos, homens que se dedicaram ao trabalho nos seringais. Metodologicamente, utilizamos o programa WordSmith Tools. Por meio de suas ferramentas WordList, Keywords e Concord, fizemos o levantamento e análise das palavras-chave de maior frequência com foco em sua significação e carga cultural. Neste procedimento de busca, destacaram-se como base (node) lexicais como 'borracha' e 'estrada'. Com base na observação das concordâncias, foi possível selecionar colocações culturais incomuns e únicas como "barões da borracha", "boom da borracha", "estrada de porta", "espigão da estrada" e "soldado da borracha", que fazem parte das idiosincrasias lexicais do "ciclo da borracha". Os seringueiros que se deslocaram para a Amazônia no período da segunda guerra mundial foram denominados de "soldados da borracha", em razão de participarem do esforço concentrado de guerra através da produção de borracha que se destinaria ao abastecimento dos mercados da Europa beligerante.

Palavras-chave

Amazônia. Corpus. Colocações.

Abstract: Based on principles of Phraseology and Corpus Linguistics, this paper focuses on the extraction and analysis of the cultural collocations present in the Corpus Escrito da Vila Japiim (CEVI) and in the Corpus Oral do Acre (COAC), both corpora were built with the speech of rubber tappers from Acre, workers that acted in the northern rubber plantations of Brazil. Regarding our methodology, we used some tools from the software WordSmith Tools, which were WordList, Keywords and Concord to collate and analyze the keywords with the highest frequency, focusing on their meaning and cultural background. During the search procedure, nodes such as 'borracha' (rubber) and 'estrada' (road) stood among the others for their concordances, from which the observation we could select uncommon and unique cultural collocations as "barões da borracha", "boom da borracha", "estrada de porta", "espigão da estrada", and "soldado da borracha" – all part of the "Amazon Rubber Boom" idiosyncrasies. The rubber tappers who moved into the Amazon during the Second World War were denominated "rubber soldiers" due to the role they played in the concentrated war efforts through the rubber production to supply the belligerent European markets.

Keywords

Amazon. Corpora. Collocations.

¹ Doutor em Linguística pelo Programa de Pós-Graduação em Estudos Linguísticos da Universidade Júlio Mesquita Filho. Professor da Rede Estadual de Ensino do Estado do Acre.

² Docente do Programa de Pós-Graduação em Estudos Linguísticos da Universidade Estadual Paulista "Júlio Mesquita Filho".

³ Docente do Departamento de Letras Modernas da Universidade Estadual Paulista Júlio Mesquita Filho.

Introdução

A expressão “boom da borracha” chama a atenção pela lexia inglesa boom, que significa explosão, crescimento vertiginoso, rapidez, ápice, cume, podendo, portanto, ser traduzida como ‘crescimento vertiginoso da produção de borracha’. O presente artigo tem como objeto de estudo as expressões convencionais e incomuns que compunham o léxico acreano no período do “ciclo da borracha”, expressões estas que, neste estudo, denominaremos “colocações culturais”.

Os relatos que formam os corpora explorados neste artigo atestam a convencionalidade de expressões como “barões da borracha”, “boom da borracha”, “estrada de centro”, “estrada de porta”, “madêra bruta”, “madêra de canto”, “espigão da estrada”, “aí entra o espigão”, “perna da estrada”, “madêra de cinco arriação”, entre outras. Com o intuito de reafirmar os conceitos correspondentes a estas colocações culturais, nos situamos diacronicamente no contexto cultural e lexical da saga dos seringueiros acreanos.

A superação do ciclo econômico do qual o Acre foi protagonista, a diversidade do contexto social e cultural dos tempos atuais, a evolução do léxico especializado impulsionado pela ciência, pela técnica e por novas formas de organização do trabalho, entre outros fatores, constituem barreiras para que um jovem acreano se interesse pelo significado das supracitadas expressões. É pertinente afirmar que o fim do “ciclo da borracha” provocou a mudança dos fatores culturais e, conseqüentemente, do campo lexical.

Neste artigo, queremos destacar a importância das colocações culturais, revendo expressões que atualmente são pouco usuais, enfatizando, assim, a oportunidade de estudá-las. Tais colocações foram extraídas de dois corpora de pesquisa, a saber: o Corpus Escrito da Vila Japiim⁴ (doravante CEVIJ), e o Corpus Oral do Acre⁵ (doravante COAC), ambos, formados por relatos de seringueiros acreanos.

As colocações da língua geral são entendidas por Orenha-Ottaiano (2017, p. 458), como “combinações recorrentes, arbitrárias e convencionais, lexicalmente e/ou sintaticamente fixas até certo grau e que podem ter um alcance colocacional mais ou menos restrito”.⁶ Na exploração dos referidos corpora, tendo como base a definição de colocações da língua geral, chegamos para a expressão “colocação cultural” (doravante

⁴ Corpus formado pelo capítulo “Depoimentos: o que vi da vida”, da obra “Histórias, causos e memórias da Vila Japiim”, de autoria de Raimundo Ibernnon Chaves da Silva.

⁵ Corpus formado por inquéritos/relatos colhidos junto a seringueiros acreanos do Vale do Acre, do Vale do Puruse do Vale do Juruá, nas décadas de 1980 – 1990, cedidos pela professora Dra. Luísa Galvão Lessa, da Universidade Federal do Acre (UFAC), que coordenou o projeto Atlas Etnolinguístico do Acre.

⁶ Collocations are understood as pervasive, recurrent, arbitrary and conventionalized combinations, which are lexically and/or syntactically fixed to a certain degree and may have a more or less restricted collocational range. (ORENHA-OTTAIANO, 2017, p. 458).

CC) ou colocações culturais (doravante CCs), que passam a ser entendidas como as combinações de palavras que apresentam determinada carga cultural e características únicas, geradas em contextos idiossincráticos e específicos de uma dada comunidade. No caso desta pesquisa, o contexto idiossincrático é o amazônico, do “Ciclo da Borracha”.

Após a superação do modelo extrativista (1º e 2º Ciclos da Borracha), a região acreana se volta para a pecuária e como símbolo do período áureo da borracha temos figuras heróicas como as de Plácido de Castro, Barão do Rio Branco e Chico Mendes. A pecuarização das terras de antigos seringais causou profundo impacto ecológico e não resolveu os problemas sociais que se agigantaram com êxodo rural crescente. Nesse contexto cultural, se inserem os autores dos relatos analisados neste artigo. Vale ressaltar que a carga cultural e a característica incomum própria dos agrupamentos fixos analisados neste artigo é o que justifica a utilização da referida denominação. Mas reiteramos que esse fenômeno pode se dar em contextos culturais variados, especialmente, em comunidades lexicais marcadas culturalmente.

Para o escopo deste artigo, analisamos apenas quatro CCs sob o critério de figuração e pertinência com os conceitos teóricos da Fraseologia e Linguística de Corpus (doravante LC) aqui tratados. Os pontos teóricos básicos pelos quais nos guiamos estão expostos a seguir. A LC busca a extração e análise de palavras em corpora a partir da utilização de programas de computador. Para este estudo, utilizamos o programa WordSmith Tools (SCOTT, 2012), versão 6.0, para auxiliar na extração das CCs que compõem o CEVIJ e o COAC.

A Fraseologia estuda as combinações de palavras na língua, ou seja, as unidades fraseológicas. Presente em todas as línguas naturais, as unidades fraseológicas integram cerca de 50% do léxico, e podem ser conceituadas como “sequências polilexicais que precisam ser memorizadas em blocos fixos, com certo grau de idiomatidade, convencionalizadas pelo uso e que constituem a competência discursiva dos falantes” (MONTEIRO-PLANTIN, 2012, p. 122).

Partindo de palavras-chave com carga cultural, estabelecemos como objetivos: levantar os itens lexicais fixos de maior frequência nos corpora escrito e oral; analisar as CCs, tendo em vista as relações existentes entre o léxico regional, a sociedade e a cultura do “ciclo da borracha”. Para atingir os objetivos estabelecidos, apresentamos na introdução colocamos sua contextualização histórica; no item 1, a teoria sobre a LC, Fraseologia, Colocações Gerais e CCs. No item 2, expomos a metodologia usada para a extração das CCs e analisamos as CCs, tomadas como exemplo para este artigo.

1. Fundamentação teórica

Neste item, discutimos os conceitos fundamentais da Linguística de Corpus, da Fraseologia, das Colocações da língua geral e de Colocações culturais.

1.1 A Linguística de Corpus (LC)

Tomamos, inicialmente, como base para este artigo, por sua amplitude, o conceito expresso por Tagnin (2013, p. 30), que afirma, “para a LC, um corpus é uma coletânea de textos, necessariamente em formato eletrônico, compilados e organizados segundo critérios ditados pelo objetivo de pesquisa a que se destina. O formato eletrônico permite análise automática e o uso de ferramentas computacionais específicas”.

A LC, nas últimas décadas, ganha espaço não só nos centros acadêmicos, mas também no âmbito empresarial, para várias finalidades comerciais: processamento automático de textos, informatização de grandes bases de dados e a montagem de sistemas inteligentes de reconhecimento de voz e gerenciamento de informações. Três corpora da língua inglesa servem como marcos de referência histórica: o Brown Corpus, o British National Corpus – BNC e o Bank of English. Na língua portuguesa destaca-se o Corpus Brasileiro e o Banco do Português (BEBER SARDINHA, 2004).

Dessa maneira, chegamos a um outro ponto de vista sobre corpus. De acordo com o pesquisador brasileiro Berber Sardinha (2004), deve ser entendido como “um conjunto de dados linguísticos (oral ou escrito), sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso linguístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise (BERBER SARDINHA, 2004, p. 3). No entendimento do autor (2004), os quatro pré-requisitos para a formação de um corpus computadorizado são: a) o corpus deve ser composto de textos autênticos, em linguagem natural; b) autenticidade dos textos; c) o conteúdo do corpus deve ser escolhido criteriosamente e apresentar condições de naturalidade e autenticidade; d) representatividade de uma variedade linguística ou mesmo de um idioma.

Para alguns autores, a LC não se comporta como um novo paradigma teórico, nem como um simples procedimento metodológico, seu diferencial está no fato de ser uma abordagem que pode ser aplicada às investigações empíricas em qualquer área de análise linguística (BIBER; CONRAD; REPPEN, 1998), e nos usos que um corpus pode oferecer para a compreensão dos fenômenos da linguagem (KENNEDY, 1998).

Tognini-Bonelli (2001) coloca a LC como um novo empreendimento de pesquisa e uma nova abordagem filosófica da linguagem, e não a considera apenas como uma simples metodologia ou conjunto de regras aplicadas em situações específicas como geralmente acontece em muitas disciplinas. Dessa maneira, no entender da autora (2001), “um corpus é uma coletânea de textos autênticos e computadorizados, passível de análise ou processamento automático, [...] selecionados de acordo com critérios

explícitos a fim de apreender as regularidades de uma língua, de uma variedade de língua, ou de uma sub-língua”.⁷ (TOGNINI-BONELLI, 2001, p. 55).

Um corpus, quando devidamente explorado, de acordo com Biber; Conrad; Reppen (1998) pode fornecer informações sobre o uso da língua, evidenciando padrões de associação ligados a traços linguísticos e não linguísticos, e como tais padrões interagem entre si. Por exemplo, um estudo baseado em corpus pode mostrar as associações lexicais, pautado apenas nas associações entre as palavras em si, ou pode revelar a sua distribuição em diferentes dialetos, registros e períodos históricos.

1.2 Convencionalidade e fraseologia

Desde tempos imemoriais, acredita-se que a sabedoria popular veicula seu legado de conhecimentos empíricos acerca da vida por meio de vários instrumentos (BRAGANÇA JUNIOR, 2012), dentre os quais as expressões paremiológicas possuem um lugar de destaque. Exemplos disso são os ‘sebayts’ (ensinamentos) egípcios, equivalentes aos provérbios atuais, datados do século III A.C. Segundo o autor (2012), entre os gregos, ‘gnôme’ (pensamento) e ‘paroemia’ (instrução) cobrem a noção de provérbio, sentença, máxima, adágio, preceito etc., aparecendo em obras de Platão, Aristóteles e Ésquilo. Tais expressões estão convencionalizadas na língua e fazem parte do repertório de seus falantes.

O linguista Saussure (1973) já havia observado que, na língua, combinações estáveis compunham uma parte significativa do léxico de forma a serem estabelecidas pela tradição e pelo uso coletivo. Ele ressalta isso em seus estudos.

Tagnin (1989, pg. 13) defende que a convencionalidade abrange tudo o que é convencional e acrescenta que, por essa expressão, se entende “aquilo que é tacitamente aceito, por uso ou geral consentimento, como norma de proceder, de agir, no convívio social; costume; convenção social”.

Conforme aponta Orenha-Ottaiano (2004), dentro da convencionalidade, insere-se a Fraseologia, que se ocupa do estudo das combinações recorrentes na língua, das unidades fraseológicas (UFs). Esses estudos ganharam projeção mais recentemente e, segundo a autora, a Fraseologia “tem por objeto de estudo as combinações fixas de palavras, [...]. Essas combinações são recorrentes, mais ou menos estabilizadas/cristalizadas e abrangem as colocações criativas, as coligações, as expressões idiomáticas, os binômios, os provérbios, dentre outros” (ORENHA-OTTAIANO, 2004, p. 7).

⁷ “[...] A corpus is taken to be a computerized collection of authentic texts, amenable to automatic or semi-automatic processing or analysis. The texts are selected according to explicit criteria in order to capture the regularities of a language variety or a sub-language”. (TOGNINI-BONELLI, 2001, p. 55).

Para Xatara (2012) o repertório de unidades fraseológicas (UFs) que engloba provérbios, refrões, expressões idiomáticas, colocações, entre outros, é um verdadeiro tesouro da língua, uma vez que essas expressões e enunciados carregam traços culturais, sociais e históricos de uma determinada comunidade linguística e de espaços geográficos específicos.

De modo geral, os fraseologismos são expressões que apresentam como traço constitutivo a fixação, isto é, seus constituintes caracterizam-se por certa estabilidade sintático-semântica, em geral, não são comutáveis e não se separam, funcionando como uma única unidade de sentido, razão pela qual são também denominadas de unidades fraseológicas (UFs) (ZULUAGA, 1980). Ortiz Alvarez (2012, p. 8), entende que a fraseologia representa “o que de mais característico existe no léxico de qualquer língua. As unidades que a compõem representam a visão de mundo de determinado povo, suas crenças, costumes e modo de compreender a vida e são empregadas de forma espontânea”. Isto é, a compreensão literal ou não de um fraseologismo vai depender deste ou daquele contexto, pois as UFs estão ligadas a aspectos culturais, sociais e históricos. Por exemplo, as expressões idiomáticas apresentam sempre um relativo grau de dificuldade para que se possa fazer uma compreensão literal do que está sendo proferido.

1.3 Colocações da língua geral e colocações culturais

Para este artigo, exploramos as CCs formadas a partir de relatos de seringueiros acreanos que formam nossos dois corpora de estudo, o CEVIJ e o COAC. Estas colocações ganharam estabilidade linguística no contexto enfocado e se cristalizaram pelo uso, tornando-se únicas e incomuns (HORI, 2004), pois dizem respeito à Amazônia da época da borracha. Antes, porém, vejamos o que dizem os autores sobre as colocações da língua geral.

Bevilacqua (2005), em artigo sobre o fenômeno das colocações destaca que estas estão formadas, basicamente, por dois elementos: um considerado a “base” e o outro, o ‘colocado’ ou o ‘elemento co-ocorrente’. A base é um elemento semanticamente autônomo, ao passo que o co-ocorrente é restringido semanticamente pela base. Considera-se, também, que ambos são unidades semi-lexicalizadas e que possuem um valor semi-composicional; ou seja, que seu significado não pode ser deduzido de forma independente por cada um dos elementos que a compõem, mas pelo conjunto desses elementos. (BEVILACQUA, 2005, p. 11)

O raciocínio estabelecido por Bevilacqua (2005) é corroborado por Orenha-Ottaiano (2004), que estabelece como funcionam os dois elementos fundamentais da colocação: a base, que é o elemento autônomo, e o colocado, ou seja, aquilo que estamos buscando.

De modo similar ao que acontece com os demais fraseologismos, também não há, na literatura, consenso acerca da conceituação das colocações.

A partir do que entende por colocação Orenha-Ottaiano (2004) elenca alguns pontos importantes levantados por Tagnin (1989):

(i) recorrência, atestada pela frequência da combinação; (ii) não idiomaticidade, o significado da colocação é composicional; (iii) coesão, há uma forte atração entre os constituintes da colocação; (iv) restrição contextual, há um contexto específico que favorece a ocorrência da combinação; (v) coocorrência arbitrária entre os elementos, pois não há explicação semântica que justifique a colocação. (ORENHA-OTTAIANO, 2004, pg. 32).

A colocação como “a relação habitual de coocorrência entre palavras”⁸ (HORI 2004, p. 23) tem uma história não muito extensa, ao contrário de outros fenômenos fraseológicos, datando de meados do século XX as primeiras discussões, por teóricos primordialmente ingleses. O autor (2004) nos oferece uma classificação a partir do que ele denomina “colocações criativas”, entendendo-as como colocações únicas ou incomuns.

As colocações são entendidas por Orenha-Ottaiano (2017, p. 458) como “combinações recorrentes, arbitrárias e convencionais, lexicalmente e/ou sintaticamente fixas até certo grau e que podem ter um alcance colocacional mais ou menos restrito”.⁹

Como tratamos nesta pesquisa de “colocações culturais”, cabe definir o que compreendemos por cultura, a fim de que possamos chegar a uma definição deste tipo de combinação. Segundo Lyons (1987, p. 223) cultura pode ser entendida como o “conhecimento adquirido socialmente, isto é, aquele que é fruto de um convívio social”. Relacionando este conceito com o objeto deste artigo, as colocações culturais, podemos inferir que, somente a partir de todo o “convívio social” do ‘mundo da borracha’, é que foi possível o surgimento das CCs. Baseado neste entendimento amplo, é importante reafirmar o conceito visto na introdução deste artigo: colocações culturais são, desse modo, combinações de palavras que apresentam determinada carga cultural e características únicas, geradas em contextos idiossincráticos e específicos de uma dada comunidade. No caso desta pesquisa, o contexto idiossincrático é o amazônico, do “Ciclo da Borracha”.

2. Metodologia e análise das colocações culturais

Neste item, destacaremos as questões metodológicas de nossa pesquisa, onde apresentaremos o *Corpus* Escrito da Vila Japiim (CEVIJ), e o *Corpus* Oral do Acre (COAC), bem como as principais ferramentas do programa *WordSmith Tools* (SCOTT, 2012), *WordList*, *KeyWords* e *Concord*. Dessa maneira, a partir dos dados quantitativos proporcionados pelo programa, podemos analisar qualitativamente as CCs extraídas dos

⁸ A relationship of habitual co-occurrence between words. (HORI, 2004, p. 23).

⁹ Collocations are understood as pervasive, recurrent, arbitrary and conventionalized combinations, which are lexically and/or syntactically fixed to a certain degree and may have a more or less restricted collocational range. (ORENHA-OTTAIANO, 2017, p. 458).

referidos *corpora*. Após esta etapa, daremos prosseguimento à análise das colocações culturais levantadas.

O pesquisador brasileiro Beber Sardinha (2009) entende que o programa *WordSmith Tools* é um conjunto de programas integrados destinado à análise linguística, baseado na frequência e na co-ocorrência de palavras em *corpora*. Além disso, ele permite pré-processar os arquivos do *corpus*, antes da análise propriamente dita. O programa serve como uma ferramenta que permite a consecução de tarefas relacionadas à análise de *corpora*.

Para o autor (2009) a LC vem revolucionando a maneira como se investiga a linguagem, nos seus mais diversos níveis, colocando à disposição do analista, quantidades de dados antes inacessíveis. Um dos grandes agentes dessa revolução é a informática; sem ela, a LC contemporânea não poderia existir. Assim, o linguista de *corpus* depende de programas de computador para lidar com *corpora*.

A LC, como frisamos, faz uso de programas computacionais para processamento de arquivos em *corpora*. O *WordSmith Tools*, usual pela forma didática com que podemos operacionalizá-lo, se destaca pelas seguintes ferramentas:

a) A ferramenta *WordList* produz listas de palavras em ordem de frequência ou alfabética, contendo todas as palavras dos arquivos selecionados. Tal lista fornece informações que podem ser usadas na análise vocabular contrastiva, uma vez que fornece dados estatísticos que possibilitam a comparação da ocorrência de palavras em arquivos diferentes.

b) Já, a ferramenta *KeyWords* gera a lista de palavras-chave de um dado arquivo por meio da comparação da lista de palavras (*WordList*) com as listas de palavras de um *corpus* de referência. Ou seja, extrai palavras de uma lista cujas frequências são estatisticamente diferentes (maiores ou menores) do que as frequências das mesmas palavras num outro *corpus* (de referência). Calcula também palavras-chave, que são chave em vários textos. Neste trabalho, o *corpus* de referência utilizado para gerar as listas de palavras-chave é o *Corpus* da Folha.

Em relação ao *Concord*, esta ferramenta fornece as linhas de concordância, ou listagens de uma palavra específica (nódulo) em seu contexto. Oferece também listas de colocados, isto é, a lista de palavras que ocorrem próximas ao nódulo em questão.

Figura 1 - A lista de palavras-chave no CEVIJ.

N	Key word	Freq.	%	Texts	RC Freq.	RC %	Keyness
1	JAPIIM	79	0,18	1	1		1.218,00 0,0
2	MÃNCIO	65	0,15	1	0		1.010,98 0,0
3	BORRACHA	93	0,22	1	968		816,95 0,0
4	JURUÁ	54	0,12	1	60		682,20 0,0
5	ERA	245	0,57	1	72.800	0,07	591,43 0,0
6	EU	239	0,55	1	68.610	0,07	591,34 0,0
7	CASOU	67	0,15	1	750		579,24 0,0
8	LÃ	146	0,34	1	20.500	0,02	551,29 0,0
9	E	1.529	3,54	1	1.911	1,86	530,30 0,0
10	MÔA	34	0,08	1	0		528,79 0,0

Fonte: PrintScreen da tela do WordSmith Tools.

Na figura 1, partir da ferramenta *KeyWords*, vemos o item lexical de busca “borracha”, palavra-chave neste trabalho, com 93 frequências.

Dessa forma, temos na primeira coluna (N): a sequência das palavras extraídas; na segunda coluna (*KeyWord*): as palavras-chave; na terceira (Freq.): a frequência dessas palavras no *corpus* de estudo; na quarta coluna (%): a porcentagem que a palavra representa em relação ao *corpus* de estudo; na quinta coluna (R.C Freq.): frequência dessas palavras no *corpus* de referência; na sexta coluna (RC.%): a porcentagem que a palavra-chave representa em relação ao *corpus* de referência; e na última coluna (*Keyness*): o índice de chavicidade. Esta lista fornece as palavras mais relevantes para as análises, pois, quanto maior o índice de chavicidade, mais representativa ela será dentro do *corpus* de estudo.

Figura 2 - As linhas de concordância no CEVIJ.

Line	Text	Count	Percentage
16	de aproximadamente 30 toneladas de borracha por ano produzidas por mais	43,82	2,224 10 83
17	carregado de gente, os soldados da borracha. Morou muitos anos no	42,84	2,2 10 10 80
18	Não, não tive sorte pra soldado da borracha. Conheço gente que	41,55	2,2 10 10 70
19	valor algum, como se sabe, o qual da borracha na época tinha um valor	43,97	2,3 71 10 84
20	e Peró. Com as lutas ardidas da borracha e dos movimentos	49,18	2,7 10 10 99
21	o fabrico deste artefato é necessário a borracha. Dai então surgu os	49,14	2,7 10 10 99
22	garantencia propriedade, mesmo assim a borracha em crise. Data do inicio dos	44,72	2,4 95 10 80
23	ate dez toneladas. Compre muita borracha a preço de açucar, também	40,52	2,1 44 10 73
24	plantar cana, além do trabalho com a borracha. Amanhei os bois, o sítio já	39,72	2,0 10 10 71
25	seringa. Os bois é quem carregava a borracha. Eri na voga braba... tinha	39,08	2,0 10 10 69
26	dessa, dona Margarida a Sra. recebe a borracha, e em três anos eu he pago	40,88	2,1 50 10 74
27	ele disse: "Seu Francisco, soldado da borracha é pra toda a vida e voreador	41,23	2,1 30 10 75
28	quando apareceu o soldado da borracha. Sou, mas eu perdi dois	41,17	2,1 10 10 75
29	Josué se aposentou como soldado da borracha, quando apareceu o soldado	41,16	2,1 00 10 75
30	temporada a média de 300 quilos de borracha. Toda essa indústria foi	25,17	1,2 10 10 26
31	e navegado por força e imposição da borracha nas últimas décadas do	9,740	44 82 0 64
32	de 1950.00, uma vez que a crise da borracha já em fato consumado e	9,697	44 50 0 64
33	que a partir de 1912 o monopólio da borracha amazônica é quebrado pelos	1,761	57 24 0 12
34	Destruídos, homens da borracha, lendários seringueiros que	9,775	44 45 0 64
35	a agropicultura. Com o fim do ciclo da borracha e a posterior transformação	9,967	44 20 0 60
36	Moa e Azul conheceu vários ciclos a borracha, a madeira, a agropicultura	9,956	44 76 0 65
37	muitos outros. Com a "queda" da borracha, as famílias mais	9,887	44 10 0 65
38	da Cláudio Costa, A Batalha da Borracha, e Pedro Martello, A	459	13 88 0 36
39	procurado. Durante o auge do ciclo da borracha embarcaram para a	144	3 24 0 15
40	a sítio, além disso, exportava-se a borracha, que depois se destacou	128	2 30 0 15
41	seringas. Com a queda do boom da borracha em 1912 e a consequente	1,217	40 20 0 85
42	criação de sítios e bovinos A borracha, nosso motor propulsor, o	1,670	55 55 0 11

Fonte: PrintScreen da tela do *WordSmith Tools*.

Visando à localização das CCs no *corpus*, a ferramenta *Concord* apresenta uma amostra das linhas de concordância a partir do nóculo selecionado. A lista de concordância, exibida acima, apresenta o item lexical de busca, “borracha”. Por meio da observação das linhas de concordância, podemos selecionar as CCs.

Os *corpora* utilizados neste artigo, em linguagem natural, autêntica, são formados por relatos de seringueiros acreanos, que falam sobre o dia-a-dia do seu trabalho nos seringais. As CCs que analisamos são empregadas a partir de padrões reais de uso. O nóculo “borracha”, neste trabalho, é o que gera o maior número de colocações culturais. Tanto no CEVIJ quanto no COAC, essa lexia se destaca por sua frequência (em destaque no próximo item, nos quadros 1 e 2).

Os seringais dão sustentação à autenticidade destes *corpora*, tornando-os representativos de um léxico regional que brota do isolamento do homem na floresta (“*mateiro*”, “*toqueiro*”, “*seringal*”, “*madêra*”), da busca por praticidade nos trabalhos diários (“*porta*”, “*manga*”, “*tigela*”, “*rodo*”), da referência cultural ‘além seringal’ (“*estrada*”, “*espigão*”), da necessidade de alfabetização e leitura (“*noteiro*”), do reconhecimento da utilidade concreta e prática da matemática (“*oito*”, “*quinze*”), da identificação com os símbolos nacionais (“*bandeira*”), da aceção à bravura e despendimento (“*batalha*”, “*soldado*”), da prática do poder e do controle (“*barão*”, “*coronel*”).

De acordo com Isquierdo (1998), ao se estabelecer o estudo de uma língua (ou um léxico regional, por exemplo), têm-se que levar em conta a cultura, “*considerando-se que, o sistema linguístico, nomeadamente o nível lexical, armazena e acumula as aquisições culturais representativas de uma sociedade [...], elementos significativos relacionados à história, ao sistema de vida, à visão de mundo*”. (ISQUERDO, 1998, p. 89).

A compreensão expressada no destaque acima, nos leva ao entendimento de que, sem a realidade cultural do “mundo da borracha”, a história, o modo de vida, sem a visão de mundo do seringueiro, o léxico amazônico não poderia ter registrado as CCs que se tornam, aqui, objeto de nossa análise.

Nas tabelas 1 e 2 apresentamos as palavras-chave (nódulos) geradores das colocações culturais, bem como, sua frequência nos *corpora*:

Tabela 1 – Palavras-chave extraídas do CEVIJ.

Palavras-chave	Frequência	Palavras-chave	Frequência
<i>Borracha</i>	59	<i>Juruá</i>	61
<i>Engenho</i>	56	<i>Pé</i>	50

Tabela 2 – Palavras-chave extraídas do COAC.

Palavras-chave	Frequência	Palavras-chave	Frequência
<i>Arriação</i>	25	<i>Mutá</i>	16
<i>Aviamento</i>	12	<i>Pau</i>	32
<i>Bandêra</i>	60	<i>Paxiúba</i>	94
<i>Boca</i>	137	<i>Poronga</i>	70
<i>Borracha</i>	647	<i>Prancha</i>	112
<i>Estrada</i>	790	<i>Rodo</i>	142
<i>Espigão</i>	64	<i>Seringueiro</i>	135
<i>Madêra</i>	222	<i>Seringueira</i>	105
<i>Manga</i>	132	<i>Tigela</i>	79
<i>Matêro</i>	19		

Nas tabelas 3 e 4 apresentamos as CCs geradas a partir das palavras-chave selecionadas para este artigo: “borracha”, no CEVIJ e “estrada”, “espigão” e “madeira”, no COAC.

Tabela 3 – As colocações culturais do CEVIJ.

Colocação cultural	Frequência	Colocação cultural	Frequência
<i>“Batalha da borracha”</i>	1	<i>“Extratativismo da borracha”</i>	2
<i>“Barões da borracha”</i>	1	<i>“Homens da borracha”</i>	2
<i>“Boom da borracha”</i>	3	<i>“Monopólio da borracha”</i>	2
<i>“Ciclo da borracha”</i>	4	<i>“Peles de borracha”</i>	2
<i>“Coronel da borracha”</i>	1	<i>“Soldados da borracha”</i>	1
<i>“Crise da borracha”</i>	2	<i>“Terra da borracha”</i>	1
<i>“Corrida da borracha”</i>	2		

Tabela 4 – As colocações culturais do COAC.

Colocações culturais	Frequência	Colocações culturais	Frequência
<i>“Corte da estrada”</i>	4	<i>“Rodo da estrada”</i>	10
<i>“Estrada de centro”</i>	11	<i>“Espigão da estrada”</i>	3
<i>“Estrada de porta”</i>	21	<i>“Entrada do espigão”</i>	1
<i>“Fecho da estrada”</i>	8	<i>“Espigão de quinze madêra”</i>	1
<i>“Fim da estrada”</i>	1	<i>“Madêra de canto”</i>	10
<i>“Manga da estrada”</i>	1	<i>“Madêra de cinco arriação”</i>	1

“Perna da estrada”	7	“Madêra bruta”	1
--------------------	---	----------------	---

No escopo deste artigo, trabalhamos a análise de 2 colocações culturais do CEVIJ, a saber: “*barões da borracha*” e “*boom da borracha*”; e 2 colocações culturais do COAC: “*estrada de porta*” e “*espigão da estrada*”. Já destacamos inicialmente que o critério que usamos para a escolha das CCs tomadas como exemplo para análise foi a sua frequência, figuração e conotação no contexto em foco, bem como por estarem condizentes com os princípios teóricos aqui discutidos. Primeiramente, destacamos no quadro 1 a CC inserida no seu contexto autêntico no *corpus*, com a análise vindo em seguida. Iniciamos com a CC “*barões da borracha*”.

Quadro 1 – “*Barões da borracha*”.

A inserção no universo da história contada e literária sobre os *barões da borracha* e sua saga na Amazônia são os pilares que dão base a este trabalho (que é ao mesmo tempo de organização e autoral) que teve a duração de aproximados quatro anos. (CEVIJ)

“*Barões da borracha*” é uma colocação nominal: Sm. ‘barões’ + Prep. ‘da’ + Sf. ‘borracha’. O título de barão, durante a monarquia brasileira, era um distintivo destinado a pessoas importantes. Uma honraria. A CC “*barões da borracha*”, transplantada para os seringais, refere-se, portanto, aos poderosos e importantes homens que dominaram através do poder do capital, com métodos rígidos, a grande massa de seringueiros que trabalhava como mão-de-obra semiescrava nos seringais, ao mesmo tempo em que ostentavam vida luxuosa nascidas de Belém e Manaus.

O pesquisador e romancista Souza (1978, p. 112) é enfático ao afirmar que, “*ao contrário do mercantilista português [...], os modernos extrativistas sustentavam-se nas casas de crédito, nas hipotecas e na exportação. Mostrariam um pendor pela ostentação por estarem irremediavelmente alienados pelo fácil enriquecimento*”. Entendemos que “*barões da borracha*” é uma CC, representada, conceitualmente, pelo colocado “*barões*” como uma personalidade do poder político e econômico da cultura brasileira. Os “*barões*” estavam no topo da pirâmide social do seringal, evidenciando, assim, a clara separação de classes e o domínio pelo léxico de “*barões*” como lugar e origem do poder.

A próxima CC em análise é “*boom da borracha*”. Conforme quadro a seguir, percebemos que o “*boom da borracha*” e os altos dividendos gerados à nação brasileira não conseguem resolver os problemas da região e a população acreana, em franco crescimento, clama ao governo brasileiro por autonomia administrativa.

Quadro 2 – “*Boom da borracha*”.

No ano de 1909, no auge do *boom da borracha*, acontece o movimento autonomista que conta com o protagonismo dos homens de patente dada pelo Governo Federal e da elite política local juruaense. (CEVIJ)

“*Boom da borracha*” é uma colocação nominal: Sm. ‘boom’ + Prep. ‘da’ + Sf. ‘borracha’. O sítio *Linguee*¹⁰ traduz *boom* como explosão, lança, auge, expansão, estrondo. O “*boom da borracha*” foi, de fato, como nos sugere a tradução, uma grande explosão, um estrondo que foi ouvido nos quatro cantos do planeta. Vindo de uma lenta, mas crescente evolução, que se estendeu a partir da segunda metade do século XIX, teve seu auge na primeiradécada do século XX, em 1910. O “*boom da borracha*” significou a alta demanda, expansão e crescimento da borracha nos mercados industriais da Europa e América do Norte. O “*boom da borracha*” justificou a implementação da ferrovia Madeira-Mamoré em plena floresta amazônica, fato inusitado que possibilitou pela literatura o surgimento da expressão “trem fantasma”, pois surgia do meio da floresta, não teve a utilidade prevista e sua construção dizimou milhares de vida.

Seguimos a análise colocacional com a CC “*estrada de porta*”, extraída do COAC.

Quadro 3 – “*Estrada de porta*”.

Como é que o senhor chamava a estrada que o senhor pegava da sua casa para as estradas de seringa? L - estrada de porta D - o que é uma *estrada de porta*? L - porque é de porta ... fecha o corte em casa D - sim e as outras? L - a estrada de cento ... tem espigão ... faiz boca na mata D - então chega na mata entra por algum lugar? L - é entra ... entrada por duas perna. (COAC)

“*Estrada de porta*” é uma colocação nominal: Sf. ‘estrada’ + Prep. ‘de’ + Sf. ‘porta’. Porta é meio de entrada ou saída ou abertura que dá acesso. A “*estrada de porta*” representou uma facilidade ao seringueiro, pois lhe possibilitou começar e fechar o trabalho do corte ao final do terreiro de sua casa, no seu ‘quintal’ (*fecha o corte em casa*). Possibilitava-lhe, também, ganhar tempo para o descanso entre o corte e a colheita.

Encerramos a análise com a CC “*espigão da estrada*”. Nas cidades, “espigão” refere-se a um prédio alto. Vejamos como o seringueiro adaptou esta lexia para a sua realidade cultural.

Quadro 4 – “*Espigão da estrada*”.

Aí volta pra estrada ... aí chega novamente na boca dela ... que é onde fecha novamente ... aí tem ... antes de chegá na boca dá o nome de espigão D - certo L - *espigão da estrada* D - então o senhor começa cinco horas da manhã ... e chega que horas [...] L - deiz minuto de espigão se fô longe ... aí corta ... quando dé nove hora tá em casa espigão é as madêra distante D - é a madeira da seringueira?...L - é ... madêra distante d'uma seringuêra pra ôtra sabe ... D - hum rum ... L - você ... tem uma madêra aqui ... aí você vai MUITO ... anda assim: cinquenta ... sessenta ... quase sessenta metro ... oitenta metro prá chegá nôtra madêra aí é o espigão. (COAC)

“*Espigão da estrada*” é uma colocação nominal: Sm. ‘espigão’ + Prep. ‘da’ + Sf. ‘estrada’. Nesta colocação, o sentido da lexia “espigão” é adaptado à realidade da “estrada”, para indicar a distância entre uma árvore e outra. Normalmente, as madeiras (seringueiras) em uma “estrada” estão situadas relativamente próximas (mas tratando-se de seringal nativo, isso não depende da vontade do “*mateiro*”). Na procura pelas

seringueiras em meio à cerrada hileia, não se devia criar “*espigão*”. Entende-se por “*espigão*”, portanto, uma ‘reta’ no meio da floresta ligando uma madeira à outra, onde a proximidade das madeiras foge ao convencional.

Considerações finais

Pelo exposto, nota-se que as colocações culturais com características incomuns analisadas neste artigo, à luz da teoria da Fraseologia e da Linguística de *Corpus*, mantêm, em português, os mesmos sentidos de colocações criativas, com características únicas conforme proposta de Hori (2004), possuindo, dessa forma, condições de análise similar.

O programa *WordSmith Tools* facilitou sobremaneira nosso trabalho, ao nos proporcionar dados estatísticos, como: listas de palavras, palavras-chave e concordâncias, o que nos possibilitou ter uma visão quantitativa e, dessa maneira, possibilitar uma análise qualitativa dos *corpora* de modo mais preciso.

Os dados resultantes e as conclusões aqui expostos, possibilitam um resgate léxico-cultural, os quais poderão ser explorados em sala de aula por professores, haja visto que a Fraseologia e, mais especificamente, as colocações, prezam pelo ensino contextualizado e fundamentado em formas recorrentes na língua em uso, que podem colaborar para o enriquecimento do léxico mental do aluno.

As CCs analisadas, “*boom da borracha*”, “*barões da borracha*”, “*estrada de porta*”, “*espigão da estrada*”, atestam a presença de combinados fixos em relatos de vida de comunidades seringueiras (regionais), preocupação ressaltada inicialmente em nossos objetivos.

Este artigo, nos leva a concluir que, pesquisas de caráter diacrônico e de resgate histórico-lexical, mostram a versatilidade da língua portuguesa, sua capacidade de adaptação a meios culturalmente marcados e confirmam o ponto de vista teórico anteriormente destacado, que afirma a interligação do léxico, com a cultura e a história de uma dada comunidade.

Referências

BEBER SARDINHA, T. **Linguística de corpus**. Barueri: Manole, 2004.

_____. **Pesquisa em linguística de corpus com WordSmith Tools**. Campinas: Mercado das Letras, 2009. (As Faces da Linguística Aplicada).

BEVILACQUA, C. R. Unidades fraseológicas especializadas: estado da questão em relação a sua definição, denominação e critérios de seleção. **Tradterm**, Porto Alegre, n. 11, p. 237-253, 2005.

BIBER, D.; CONRAD, S.; REPPEN, R. **Corpus linguistics: investigating language, structure and use.** Cambridge: Cambridge University Press, 1998.

BRAGANÇA JUNIOR, A. A. Os provérbios na Idade Média Latina: por um jogo especular histórico-social. In: SILVA. (Org.) **Fraseologia & Cia: entabulando diálogos reflexivos.** Londrina: Universidade Estadual de Londrina, p. 51-67, 2012.

HORI, M. **Investigating Dickens' Style: a collocational analysis.** New York: Palgrave Macmillan, 2004.

KENNEDY, G. **An introduction to corpus linguistics.** London: Longman, 1998.

LESSA, L. G. O atlas etnolinguístico do Acre. **Comunicação...** CONGRESSO NACIONAL DE LINGÜÍSTICA E FILOLOGIA. UERJ, Rio de Janeiro, 2000.

ORENHA-OTTAIANO, A. O. **A compilação de um glossário bilíngue de colocações criativas, na área de jornalismo de negócios, baseado em corpus comparável.** 2004. 233 f. Dissertação (Mestrado em Estudos Linguísticos e Literários) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2004.

_____. The compilation of an online corpus-based bilingual collocations dictionary: motivations, obstacles and achievements. In: **Proceedings. E-LEX CONFERENCE 2017.** Leiden, p. 458-473, 2017.

ORTIZ ALVAREZ, M. L. Apresentação ou entabulando a conversação. In: SILVA. (Org.) **Fraseologia & cia: entabulando diálogos reflexivos.** Londrina: Universidade Estadual de Londrina, p. 7-12, 2012.

SAUSSURE, F. **Curso de Linguística Geral.** São Paulo: Cultrix, 1973.

SCOTT, M. **WordSmith Tools.** Versão 6.0. Liverpool: Lexical Analysis Software, 2012.

SILVA, R. I. C. **Causos, histórias e memórias da Vila Japiim: 1912-1977.** Recife: Bagaço, 2013.

SOUZA, M. **A expressão amazonense: do colonialismo ao neocolonialismo.** 2. ed. São Paulo: Alfa-Ômega, 1990.

TAGNIN, S. E. O. **O jeito que a gente diz: combinações consagradas em Inglês e Português.** Barueri: Disal, 2013.

_____. **Expressões idiomáticas e convencionais.** São Paulo: Ática, 1989.

TOGNINI-BONELLI, E. **Corpus linguistic at work.** Amsterdam: John Benjamins, 2001.

XATARA, C. M. Mais um dedinho de prosa... In: SILVA. (Org.) **Fraseologia & cia: entabulando diálogos reflexivos.** Londrina: Universidade Estadual de Londrina, p. 7-12, 2012.

ZULUAGA, A. **Introducción al estudio de las expresiones fijas.** Frankfurt: Peter D. Lang, 1980.