

Etimologia ficcional contextual: o léxico indianista em Alencar à luz da Linguística de Corpus

Contextual fictional etymology: the indianist lexicon in Alencar the light of Corpus Linguistics

Maria Virgínia Dias de Ávila¹
Ariel Novodvorski²

Recebido em: 30/10/2019

Aprovado em: 28/03/2020

Publicado em: 30/06/2020

RESUMO: O presente artigo tem por objetivo apresentar uma parcela da pesquisa desenvolvida em nossa tese de Doutorado, cujo estudo teve como base o léxico indianista de José de Alencar em *Iracema*, *O Guarani* e *Ubirajara*. O recorte ora apresentado trata de uma análise na perspectiva da Etimologia Ficcional Contextual, conceito elaborado durante nossa pesquisa, que consiste na busca de origem dos vocábulos, a partir da interpretação no contexto de uso. Para o tratamento da Etimologia, utilizamos os pressupostos teóricos de Casares (1992) e de Viaro (2014). Para processamento do *corpus*, utilizamos o programa *WordSmith Tools*, 6.0 (SCOTT, 2012). Para a análise etimológica, recorreremos também ao *Corpus do Português* (DAVIES, 2016) e a dicionários de consulta e exclusão. Verificamos que Alencar cria étimos em seus romances com o objetivo de nomear, conforme seus desejos literários e linguísticos. Consideramos que Alencar contribuiu, sobremaneira, para a formação da língua portuguesa brasileira, como, por exemplo, o nome *Moacir*, que foi criado pelo autor, conforme corroboramos em nossa pesquisa, e tem seu uso expandido na língua portuguesa. Alencar estabelece a Etimologia Ficcional Contextual em razão do desejo de aproximar o significado dos vocábulos ao elemento nomeado, ou seja, para atender aos propósitos literários e linguísticos.

Palavras-chave: Etimologia Ficcional Contextual; Indianismo em Alencar; Linguística de *Corpus*; Étimos indianistas.

ABSTRACT: This paper aims to present a portion of the research developed in our PhD thesis, whose study was based on the indianist lexicon of José de Alencar in *Iracema*, *O Guarani* and *Ubirajara*. The cutout presented here deals with an analysis from the perspective of Contextual Fictional Etymology, a concept developed during our research, which consists of the search for the origin of vocabulary, from the interpretation in the context of use. For the treatment of Etymology, we used the theoretical assumptions of Casares (1992) and Viaro (2014). To process the corpus, we used *WordSmith Tools*, 6.0 (SCOTT, 2012). For the etymological analysis, we also used the *Corpus do Português* (DAVIES, 2016) and dictionaries for consultation and exclusion. We found that Alencar creates etymologies in his novels with the aim of naming them according to his literary and linguistic desires. We believe that Alencar has greatly contributed to the formation of the Brazilian Portuguese language, such as, for example, the name *Moacir*, which was created by the author, as we corroborated in our research, and is still used today has its use extended to the Portuguese language. Alencar establishes the Contextual Fictional Etymology due to the desire to approximate the meaning of vocabulary to the named element, that is, to meet the literary and linguistic purposes.

Keywords: Contextual Fictional Etymology; Indianism in Alencar; Corpus Linguistics; Indianist Etymologies.

1. Graduação em Letras. Especialista em Linguística Aplicada. Mestrado em Linguística. Doutorado em Estudos Linguísticos. Docente da Faculdade FATRA – Uberlândia/MG. ORCID: 0000-0003-2657-024X. E-mail: virginia@ufu.br

2. Doutor em Estudos Linguísticos (UFMG). Professor Adjunto do Instituto de Letras e Linguística (ILEEL) e do Programa de Pós-graduação em Estudos Linguísticos (PPGEL) da Universidade Federal de Uberlândia (UFU). ORCID: 0000-0003-1370-8334 E-mail: arivorski@ufu.br

INTRODUÇÃO

José de Alencar é considerado um dos precursores dos ideais nacionalistas na prosa brasileira, por isso, também, é considerado um dos símbolos da literatura romântica no Brasil. Após a Independência do Brasil, 1822, período que coincide com o Romantismo brasileiro, os escritores tomaram consciência da necessidade de se construir uma literatura genuinamente brasileira, que se identificasse com as próprias cores e sabores, com as raízes históricas e linguísticas e, sobretudo, culturais.

Alencar deixou uma vasta obra, dentre as quais merecem destaque as consideradas indianistas: *O Guarani*, *Iracema* e *Ubirajara*. O autor, como estudioso da língua indígena, afirmou no prefácio de *Iracema* que “o conhecimento da língua indígena é o melhor critério para a nacionalidade da literatura. É nessa fonte que deve beber o poeta brasileiro; é dela que deve sair o verdadeiro poema nacional, tal como eu o imagino”. O autor objetivou um texto verdadeiramente nacional na perspectiva do nacionalismo e de uma literatura autenticamente brasileira.

Para isso, o autor recorreu ao léxico indianista, ora utilizando vocábulos já existentes, ora criando étimos indígenas. Ao criar étimos indígenas, o autor entremeia as explicações das expressões com a descrição de algumas palavras que, na sua perspectiva, necessitaria de auxílio para compreensão do leitor da época, como na nota para a palavra “Uiraçaba — Aljava, de *uira* — seta, e a desinência *çaba* — coisa própria”. No corpo do texto do romance, o vocábulo é utilizado no período “O sentimento que ele pôs nos olhos e no rosto, não o sei eu. Porém a virgem lançou de si o arco e a uiraçaba, e correu para o guerreiro, sentida da mágoa que causara” (ALENCAR, 1965, p. 05). Observamos que, neste caso, o autor não apresenta o significado do vocábulo uiraçaba na nota, mas o seu processo de composição, ou seja, cria um étimo.

A Etimologia é o estudo da origem e evolução das palavras. Expandimos o conceito de Etimologia para Etimologia Ficcional Contextual que consiste na análise ou a na busca da origem dos vocábulos, a partir da interpretação no contexto de emprego nas obras. Estudar o léxico indianista na perspectiva da Etimologia Ficcional Contextual, além de oferecer subsídios para a melhor compreensão das pretensões literárias do autor no contexto das obras indianistas, também proporciona melhor compreensão de itens lexicais que foram criados por Alencar e, atualmente, fazem parte da Língua Portuguesa. Sendo assim, o objetivo deste artigo é apresentar uma análise etimológica ficcional contextual com base nos itens lexicais indígenas de Alencar.

A estrutura deste artigo consiste, além desta introdução, na apresentação dos pressupostos teóricos sobre Etimologia e, principalmente, na apresentação do conceito “Etimologia Ficcional Contextual” elaborado durante a pesquisa de Doutorado (AUTOR, 2018), o qual fundamenta as análises propostas. Em seguida, descrevemos a metodologia da Linguística de *Corpus* utilizada para o desenvolvimento da pesquisa, seguida da apresentação de análise de cinco étimos indianistas de Alencar. Por fim, apresentamos as considerações finais e as referências.

ETIMOLOGIA FICCIONAL CONTEXTUAL

As palavras fazem parte do léxico de uma língua, porém não de forma arbitrária. O léxico vai se formando a partir de uma evolução caracterizada por empréstimos, por criação de novas palavras ou por ressignificação das já existentes. O conhecimento da origem das palavras é importante, porque também é uma forma de conhecer a própria história de uma sociedade. Entretanto, descobrir a origem das palavras de uma língua não é uma tarefa fácil. Em muitos casos, afirmar a verdadeira origem de determinada palavra é impossível, dada a dinamicidade de uma língua viva.

Não obstante a essa dificuldade, a etimologia se ocupa de saber não só a respeito da origem da palavra, mas também do seu significado atual, sua história e seus processos de formação. Etimologia é, então, o estudo da origem das palavras, juntamente com as sucessivas mudanças pelas quais essa palavra passou (CASARES, 1992). Por sua vez, étimo pode ser entendido como sinônimo de etimologia, porém há de se considerar o contexto de emprego dos dois termos como tal. Dubois (2007) também traz uma definição para étimo de que “é qualquer forma dada ou estabelecida de que se pode derivar uma palavra; o étimo pode ser radical, base a partir da qual se criou, com um afixo, uma palavra recente. O étimo também pode ser a forma antiga de que se origina uma forma recente.” (DUBOIS, 2007, p. 251).

Para se estabelecer o étimo de uma palavra, nesse sentido, é necessário ter os *corpora* datados. Embora conhecer a data da criação de uma palavra seja praticamente impossível, a datação da ocorrência escrita mais antiga é um limite importante para se saber que, em termos daquela sincronia, a palavra já era usada (VIARO, 2014). É o caso dos étimos alencarianos, que foram empregados pela primeira vez em textos escritos em língua portuguesa pelo autor em seus romances indianistas. Dados consultados no *Corpus do Português* (DAVIES, 2016) e em dicionários anteriores à publicação dos romances, comprovam que alguns vocábulos foram empregados, ou pelo menos

publicados pela primeira vez em língua portuguesa, por Alencar. Esses *corpora* de consulta são tão importantes para a etimologia que Viaro (2014) os coloca, em termos de importância, no mesmo patamar dos dados coletados pela Arqueologia ou pela Paleontologia.

Estudar os étimos alencarianos, especialmente os indianistas, é um campo frutuoso, dada a diversidade de vocábulos indianistas que o autor emprega. Além disso, é possível considerar a sua vasta criatividade ao criar étimos contextuais ficcionais, ou seja, o autor cria étimos indígenas para atender aos seus desejos literários e linguísticos para serem empregados em suas obras, a fim de atribuir aos personagens e aos cenários o caráter indígena desejado.

Étimos ficcionais contextuais são, portanto, as formas criadas por um autor, com base na necessidade de atribuir mais do que um nome a um objeto ou a um personagem, mas de atribuir-lhes significado por meio dos vocábulos. O vocábulo, ao mesmo tempo em que nomeia, atribui também características. Nesse sentido, a análise não estará restrita ao vocábulo, mas atrelada ao contexto de uso nos textos de ficção literária.

Etimologia Ficcional Contextual é, portanto, o estudo dos étimos ficcionais contextuais considerando o processo de formação, atrelado ao contexto de produção em uma obra de ficção, como também o significado desse étimo relacionado às características do ser nomeado. Em Alencar, é possível falar de etimologia ficcional contextual, já que o autor buscou explicar os vocábulos empregados nos romances, para os quais não encontrou explicação nos estudos realizados nos livros e dicionários. As razões para as criações dos étimos ficcionais contextuais se deve ao fato de que o autor se baseia em associações de sentido, muitas vezes, a partir da decomposição de palavras oriundas de diversas fontes de pesquisa como Aires de Casal (1754?-1821?), Varnhagen (1816), Gonçalves Dias (1858), dentre outros.

A Etimologia Ficcional Contextual pode atuar tanto na forma quanto no significado das palavras, ou seja, o autor poderá ressignificar uma palavra por aproximação de sentido ou por aproximação fonética, quanto poderá compor uma nova palavra a partir de elementos já conhecidos. Na Etimologia Ficcional Contextual, o autor, por analogia, explica as mudanças ou cria vocábulos por meio do contexto de uso no interior do texto, ou seja, naquele ambiente da narrativa do romance. Por associações conscientes, o autor revela a essência da palavra no contexto da obra produzida. No caso de Alencar, o autor procura explicar os vocábulos indígenas criados por meio das notas explicativas ao final

dos romances e das frases explicativas dos vocábulos, ao decorrer da narrativa, buscando, assim, tornar os étimos compreensíveis para os leitores.

Consideramos Etimologia Ficcional Contextual, primeiro, porque alguns vocábulos podem ser considerados étimos do autor; segundo, pois pode ser um vocábulo utilizado para designar narrativas imaginárias ou referir-se a obras criadas a partir de “elementos imaginários calcados no real e/ou de elementos da realidade inseridos em contextos imaginários” (HOUAISS, 2009), como também pelo fato de tratar da criação e emprego de um vocábulo no interior de uma obra de ficção.

Etimologia Ficcional Contextual é, então, a análise ou a busca da origem dos vocábulos a partir da interpretação no contexto de emprego nas obras indianistas de Alencar. Portanto não temos pretensão de tratar os vocábulos em sua formação na história da língua portuguesa, que cabe ao Etimólogo, mas tratar da etimologia dos vocábulos na evolução dos romances de Alencar, ou seja, baseamo-nos na formação das palavras na história das obras.

PROCEDIMENTOS METODOLÓGICOS

O *corpus* linguístico de estudo reúne as três obras consideradas indianistas do autor José de Alencar: *O Guarani*, *Iracema* e *Ubirajara*. Para identificar o léxico indianista do autor utilizamos o programa *WordSmith Tools* (WST), de Mike Scott (2012), na versão 6.0, com as ferramentas *Concord*, *WordList* e *KeyWords*.

Para consulta dos candidatos a itens lexicais indianistas, fizemos a consulta a quatro dicionários e ao *Corpus do Português* (Davies, 2016), especificamente na sua versão **Gênero/Histórico**. A versão histórica possui em torno de 45 milhões de palavras de, aproximadamente, 57.000 textos em português, entre os séculos XIV e XX, o que permite verificar a datação do emprego dos vocábulos na língua portuguesa. Segundo Viaro (2014, p. 101), “nos estudos etimológicos, os *corpora* são de grande importância, sobretudo os organizados de forma que se possa obter alguma informação diacrônica”.

Nosso *corpus* de estudo composto pelas obras indianistas de Alencar foi compilado do Portal do Domínio Público¹ em formato PDF. Em seguida, foram convertidos ao formato *Word* e aplicados procedimentos para limpeza e preparação. Depois de limpos e organizados em pastas, efetuamos a conversão para *TXT* para processamento com as ferramentas do *WST*.

¹ Disponível em: www.dominiopublico.gov.br. Acesso em: 10 abr. 2018.

A limpeza do *corpus*, cujo objetivo é manter o conteúdo textual de cada obra retirando os elementos não processados pelo *WST*, compreende uma das etapas de preparação do *corpus*. No caso desta pesquisa, retiramos os elementos que não são de interesse, como, por exemplo, nome do autor, números de página ou de capítulos, notas, sumários, índices e sumários, dentre outros. Dessa forma, retiramos tudo o que não compõe o corpo dos textos, ou seja, o que não pertence às narrativas dos romances.

Após a limpeza do *corpus*, iniciamos os trabalhos de processamento pelo *WST* com a geração de uma lista de palavras, por meio da ferramenta *WordList*. Em seguida, iniciamos a análise com suporte da ferramenta *WordList*, no intuito de identificação dos candidatos a vocábulos indianistas. Após os procedimentos de limpeza iniciais, a lista resultante computou 758 itens e foi armazenada para posterior extração e corroboração em dicionário do eventual léxico indianista.

A primeira consulta se deu ao dicionário *Houaiss* em versão eletrônica 2009. Se esse dicionário apontasse a etimologia da palavra e ela não fosse indígena, essa palavra seria excluída da lista dos candidatos a vocábulos indianistas. Mantivemos, nesse momento, todos os vocábulos, adotando os critérios: palavras, cuja origem não fosse explicitada pelo dicionário *Houaiss*, inclusive os nomes próprios; palavras não constantes da lista dos verbetes do dicionário; e as palavras declaradas de origem indígena pelo dicionário. Esse procedimento foi adotado com os 758 itens candidatos a vocábulos indianistas.

Após esse procedimento de consulta, restaram 408 palavras cuja etimologia não fora mencionada no dicionário *Houaiss*. Para pesquisar sobre a etimologia dos 408 vocábulos, utilizamos quatro dicionários anteriores às publicações das obras indianistas de Alencar: um, exclusivamente de língua indígena, *Dicionário de Língua Tupy*, de A. Gonçalves Dias (GD) (1858); e três de Língua Portuguesa: *Dicionário da Língua Brasileira*, de Luiz Maria da Silva Pinto (LMP) (1832); *Diccionario da Língua Portuguesa*, de Antonio de Moraes Silva (AMS) (1789) e *Vocabulario Portuguez e Latino*, de Raphael Bluteau (RB) (1712). Ao final da consulta aos dicionários mencionados, restou uma lista de 367 vocábulos indianistas.

Para identificação dos étimos indianistas do autor José de Alencar, recorreremos também ao *Corpus do Português*, (DAVIES, 2016), em sua versão histórica e diacrônica, que nos proporcionou uma das possibilidades de contato com os textos mais antigos escritos em português, recurso que favoreceu para a revelação do processo de criação de étimos indianistas pelo autor José de Alencar.

Procedemos, então, à análise dos vocábulos com base nos procedimentos mencionados, porém, em razão da dimensão e estrutura deste artigo, apresentamos os procedimentos e análises de cinco vocábulos: **Iracema**, **Moacir**, **jacarecanga**, moquém e **carimã**.

ETIMOLOGIA FICCIONAL CONTEXTUAL EM ALENCAR

Alencar afirma que, desde cedo, quando começaram os desejos de escrever, “a raça selvagem indígena” despertava-lhe interesse. Por não possuir conhecimentos suficientes para apreciar esta raça, ele se empenhou em estudos por meio de textos já publicados sobre o tema, porém não encontrou, dentre eles, uma “poesia nacional”, tal como ele percebia “os selvagens”. Em razão disso, Alencar procurou apropriar-se do linguajar indígena para tematizá-los em suas obras literárias.

Apresentamos análise de cinco vocábulos na perspectiva do conceito de Etimologia Ficcional Contextual, considerando a análise ou a busca da origem dos vocábulos, a partir da interpretação no contexto de emprego nas obras. Ressalta-se que Alencar não fizera distinção entre as línguas Tupi e Guarani, pois ele utiliza vocábulos das duas línguas e de outros dialetos indígenas em suas obras.

Iniciamos nossa análise com o nome da personagem principal que também nomeia o romance: **Iracema**. Há algumas tentativas de explicar a etimologia do vocábulo, dentre elas, o próprio Alencar, em nota, ao final do romance, afirma que “**Iracema** - Em guarani significa lábios de mel - de *ira*, mel e *tembe* - lábios. *Tembe* na composição altera-se em *ceme*, como na palavra *ceme-iba*”. Já Navarro (2013) não menciona a língua guarani e propõe que **Iracema** viria do nhengatu (uma língua amazônica), e foi utilizado por Alencar com o significado original (*ira*, abelhas + *sema*, saída) e alterada para “lábios de mel”. O dicionário de GD não possui o verbete **Iracema** como antropônimo, mas traz as palavras utilizadas por Alencar no processo de composição *ira* ou *yra* – mel; e *tembê* – beijo.

A fim de verificar a datação da primeira ocorrência do vocábulo **Iracema** em textos escritos no português brasileiro, constatamos a inexistência de ocorrências anteriores a Alencar no *Corpus do Português* (DAVIES, 2016). Apesar de os dados apresentados nesse *corpus* não serem a garantia de que o vocábulo, de fato, ocorreu pela primeira vez em Alencar, é altamente significativo que nenhum registro do vocábulo **Iracema** tenha sido encontrado no *Corpus do Português* (DAVIES, 2016), em sua versão histórica. O fato de o vocábulo **Iracema** não ter sido registrado, no *Corpus do Português* (DAVIES, 2016), anteriormente à publicação da obra *Iracema* é relevante, dada a extensão do *corpus* (45

milhões de palavras). Esse já poderia configurar um indício quanto à etimologia do vocábulo *Iracema*, criado na trama ficcional e no contexto da narrativa da obra. A Figura 1, a seguir, ilustra uma vista parcial da busca pela palavra **Iracema** no *Corpus do Português* (DAVIES, 2016):

Figura 1: Busca por *Iracema* no *Corpus do Português*



The screenshot shows the search results for the word "Iracema" in the "Corpus do Português: Gênero/Histórico" interface. The interface includes a search bar, navigation tabs (PESQUISAR, FREQUÊNCIA, CONTEXTO, CONTEXTO +), and a list of search results. The results are displayed in a table with columns for document ID, author, and text snippet. The word "Iracema" is highlighted in green in the text snippets.

Corpus do Português: Gênero/Histórico			
PESQUISAR	FREQUÊNCIA	CONTEXTO	CONTEXTO +
SECCÕES: s19 (275) AMOSTRA: 100 200 PÁGINA: << < 1 / 3 > >>			
CLIQUE NO TÍTULO PARA MAIS CONTEXTO <input type="checkbox"/> [?] SHOW DUPLICATES			
1	18:Anjos:Eu	A B C	acordando na desgraça, Viu toda a podridão de sua raça.. Na tumba de Iracema .. Ah! Tudo, como um lúgubre ciclone, Exercia sobre ele a
2	18:Alencar:Guarani	A B C	. A palmeira arrastada pela torrente impetuosa fugia.. E sumiu-se no horizonte. 0006-01155.TXT## Iracema , de José de Alencar Edição de
3	18:Alencar:Iracema	A B C	da praia um eco vibrante, que ressoa entre o marulho das vagas: - Iracema . O moço guerreiro, encostado ao mastro, leva os olhos preso:
4	18:Alencar:Iracema	A B C	II Além, muito além daquela serra, que ainda azula no horizonte, nasceu Iracema . Iracema, a virgem dos lábios de mel, que tinha os cabe
5	18:Alencar:Iracema	A B C	, muito além daquela serra, que ainda azula no horizonte, nasceu Iracema. Iracema , a virgem dos lábios de mel, que tinha os cabelos ma
6	18:Alencar:Iracema	A B C	esparziam flores sobre os úmidos cabelos. Escondidos na folhagem os pássaros ameigavam o canto Iracema saiu do banho; o aljôfar d' é
7	18:Alencar:Iracema	A B C	ignotos cobrem-lhe o corpo. Foi rápido, como o olhar, o gesto de Iracema . A flecha embebida no arco partiu Gotas de sangue borbulham
8	18:Alencar:Iracema	A B C	que rápida ferira, estancou mais rápida e compassiva o sangue que gotejava. Depois Iracema quebrou a flecha homicida: deu a haste ao
9	18:Alencar:Iracema	A B C	dos tabajaras, senhores das aldeias, e à cabana de Araquém, pai de Iracema . III O estrangeiro seguiu a virgem através da floresta. Quand
10	18:Alencar:Iracema	A B C	bosque, então seu olhar como o do tigre, afeito às trevas, conheceu Iracema e viu que a seguia um jovem guerreiro, de estranha raça e k
11	18:Alencar:Iracema	A B C	cabana. O mancebo sentou-se na rede principal, suspensa no centro da habitação. Iracema acendeu o fogo da hospitalidade; e trouxe o
12	18:Alencar:Iracema	A B C	vibrou o maracá e saiu da cabana, porém o estrangeiro não ficou só. Iracema voltara com as mulheres chamadas para servir o hóspede i
13	18:Alencar:Iracema	A B C	traga luz a teus olhos, alegria à tua alma. E assim dizendo, Iracema tinha o lábio trêmulo, e úmida a pálpebra. - Tu me deixas?

Fonte: *Corpus do Português* (DAVIES, 2016).

É possível notar que a primeira linha com o registro do vocábulo **Iracema** é de um texto de Augusto dos Anjos. Contudo, esse autor é um sucessor de José de Alencar. Ao consultar o texto de Anjos, notamos que ele faz referência à índia Iracema, personagem de Alencar, como se observa na Figura 2.

Figura 2: Iracema em Augusto dos Anjos

The screenshot shows the 'Corpus do Português: Gênero/Histórico' interface. At the top, there are navigation tabs: 'PESQUISAR', 'FREQUÊNCIA', 'CONTEXTO', and 'CONTEXTO +'. Below the tabs, the 'FONTE:' section contains a table with the following information:

Data	
Título	Eu
Autor	Augusto dos Anjos

Below the table, the 'Contexto ampliado:' section displays a text excerpt from the work 'Eu' by Augusto dos Anjos, with the word 'Iracema' highlighted in green.

Fonte: *Corpus do Português* (DAVIES, 2016).

A linha de concordância Nº 2 da Figura 1 apresenta detalhes de uma edição de *O Guarani* e que, provavelmente, não corresponderia à primeira edição, uma vez que faz referência ao romance *Iracema*, como atesta a Figura 3.

Figura 3: Vocábulo *Iracema* na obra *O Guarani*

The screenshot shows the 'Corpus do Português: Gênero/Histórico' interface. At the top, there are navigation tabs: 'PESQUISAR', 'FREQUÊNCIA', 'CONTEXTO', and 'CONTEXTO +'. Below the tabs, the 'FONTE:' section contains a table with the following information:

Data	
Título	O Guarani
Autor	José de Alencar

Below the table, the 'Contexto ampliado:' section displays a text excerpt from the work 'O Guarani' by José de Alencar, with the word 'Iracema' highlighted in green.

Fonte: *Corpus do Português* (DAVIES, 2016).

Cabe destacar, também, que nos dicionários RB, AMS e LMSP não há registro do vocábulo *Iracema*, já GD traz o verbete *Iracema* referindo-se a uma cidade do Ceará. Outros aspectos importantes em relação ao vocábulo *Iracema* é que: primeiro, Navarro (2013) discorda de Alencar em relação à origem do nome, afirmando ser da língua nhengatu, já que Alencar afirma ser de origem guarani. Segundo, Navarro afirma que

Iracema tem o significado original *ira* – abelha e *sema* – saída, ou seja, seria a saída das abelhas, ou seja, o próprio mel, porém Alencar deixou clara a etimologia que pretendia para o nome da heroína indígena era *ira* – mel e *tembe* – lábios: a virgem dos lábios de mel. É possível afirmar que Alencar compôs o vocábulo **Iracema** a partir da definição de GD, porém atribuiu o significado pretendido por ele.

Assim sendo, com base nos documentos consultados, o vocábulo **Iracema** é utilizado pela primeira vez por Alencar, como antropônimo. **Iracema** é, portanto, um étimo criado por José de Alencar, trata-se, portanto, de uma Etimologia Ficcional Contextual.

Alencar coloca **Iracema** diante da imaginação do leitor por meio das descrições em forma de comparação “Iracema, a virgem dos lábios de mel, que tinha os cabelos mais negros que a asa da graúna e mais longos que seu talhe de palmeira”. E é essa a imagem que Alencar deseja para sua heroína, a companhia doce e servil ao seu amor, Martim; chama-a de “virgem”, uma condição importante para a mulher indígena na visão das tribos. A virgindade da índia é “doada” também para o jovem português em demonstração de amor incondicional. Passa a ser considerada uma espécie de santa nacional e referência no Brasil ao se utilizarem, em contextos diversos, expressões como “terra de Iracema” e “filhos de Iracema”, as quais, inicialmente, designavam os que nasciam no Ceará, porém, tempo depois, denominou a todos os brasileiros.

Iracema, a virgem dos lábios de mel, conquistou a todos e exerceu nos seus tempos futuros um efeito não pensado por Alencar. Seu nome já foi utilizado extensivamente para nomear as meninas que nasciam no Ceará e, em menor proporção, nos outros cantos do Brasil. Nomeia também praias, estabelecimentos comerciais de todos os tipos, ruas, enfim saiu do livro para ganhar notoriedade no país

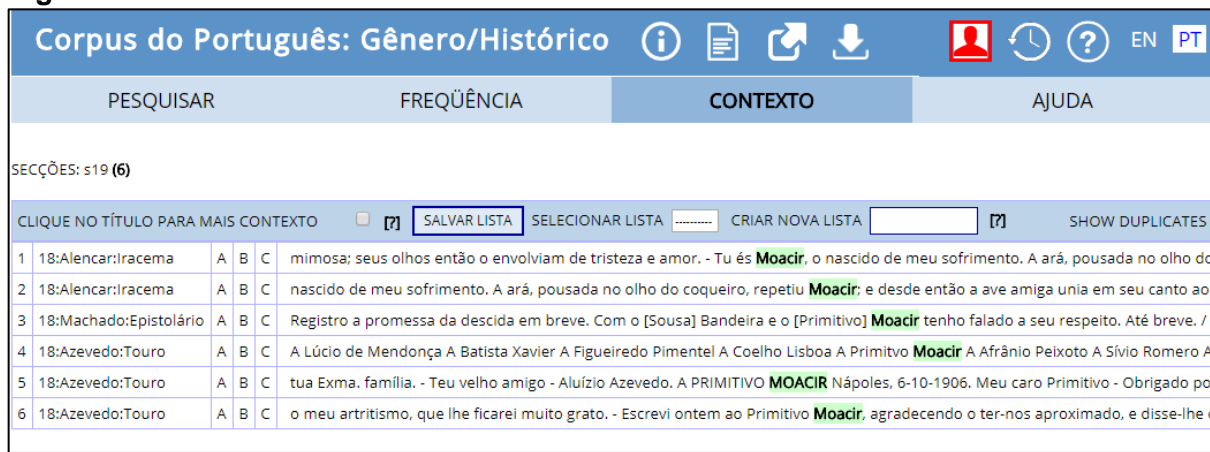
O direcionamento de Alencar para compor seus personagens e, imbricado de ambições ideológicas e literárias, cria também o personagem **Moacir**. Não é um personagem determinante no romance, porém não é menor o seu valor, especialmente no que tange aos propósitos deste trabalho. Pode-se dizer que o enredo é que direciona o autor a criar e a nomear o personagem **Moacir**, filho de **Iracema** e **Martim**.

Alencar traz, na seção de notas do livro *Iracema*, a descrição “**Moacir** — Filho do sofrimento: de *moaci* — dor, e *ira* — desinência que significa — saído de”. Devido à necessidade de explicitar o emprego dos vocábulos em seus romances, o autor cria o étimo ao propor um processo de formação e um significado ao vocábulo para designar, por meio do nome, o que desejava para o filho de **Iracema**, ou seja, aquele que nasceu em decorrência do sofrimento físico e emocional da mãe. Além de ressaltar no próprio

romance pela fala de **Iracema**, “- Tu és Moacir, o nascido do meu sofrimento” o autor, para reforçar a intenção de explicar que **Moacir** é fruto de um amor que trouxe sofrimentos à mãe, também traz a nota de composição do vocábulo para garantir seu propósito.

A fim de corroborar a análise proposta, buscamos o vocábulo **Moacir** no *Corpus do Português* em sua versão histórica, que confirma a primeira ocorrência no romance **Iracema**², conforme a Figura 4.

Figura 4: Linhas de Concordância com o vocábulo **Moacir**



The screenshot shows the 'Corpus do Português: Gênero/Histórico' interface. The top navigation bar includes 'PESQUISAR', 'FREQÜÊNCIA', 'CONTEXTO', and 'AJUDA'. Below the navigation bar, there are search options like 'SECCÇÕES: s19 (6)', 'CLIQUE NO TÍTULO PARA MAIS CONTEXTO', and buttons for 'SALVAR LISTA', 'SELECIONAR LISTA', 'CRIAR NOVA LISTA', and 'SHOW DUPLICATES'. The main content area displays a table with 6 rows of concordance results. Each row includes a line number, the source text, and the word 'Moacir' highlighted in green.

Line	Source	Word	Context
1	18:Alencar:Iracema	A B C	mimosa: seus olhos então o envolviam de tristeza e amor. - Tu és Moacir , o nascido de meu sofrimento. A ará, pousada no olho do
2	18:Alencar:Iracema	A B C	nascido de meu sofrimento. A ará, pousada no olho do coqueiro, repetiu Moacir ; e desde então a ave amiga unia em seu canto ao r
3	18:Machado:Epistolário	A B C	Registro a promessa da descida em breve. Com o [Sousa] Bandeira e o [Primitivo] Moacir tenho falado a seu respeito. Até breve. / C
4	18:Azevedo:Touro	A B C	A Lúcio de Mendonça A Batista Xavier A Figueiredo Pimentel A Coelho Lisboa A Primitvo Moacir A Afrânio Peixoto A Sívio Romero A
5	18:Azevedo:Touro	A B C	tua Exma. família. - Teu velho amigo - Aluizio Azevedo. A PRIMITIVO MOACIR Nápoles, 6-10-1906. Meu caro Primitivo - Obrigado por
6	18:Azevedo:Touro	A B C	o meu artritismo, que lhe ficarei muito grato. - Escrevi ontem ao Primitivo Moacir , agradecendo o ter-nos aproximado, e disse-lhe co

Fonte: *Corpus do Português* (DAVIES, 2016).

Alencar, portanto, desejava que o nome do filho de Iracema fosse um reflexo de sua vida durante a gestação e o parto: sofrimento.

A obra de Alencar teve uma repercussão tão expressiva que a palavra **Moacir**, apesar de apenas duas ocorrências no romance, foi incorporada à língua portuguesa e ganhou tamanha popularidade que, a partir da publicação de *Iracema*, passou a ser utilizada no batizado de meninos. Assim sendo, a popularidade do personagem preponderou sobre o significado.

Entre os dicionários consultados, GD define “**Moacy** – magoar-se, estimular-se; agravado, sentido, doente” (p. 445) sem mencionar a etimologia. O dicionário apresenta definição semelhante para **Moacir**, porém não menciona a decomposição do vocábulo. Apesar de semelhante, o significado não é exatamente como Alencar propõe “o nascido do meu sofrimento”. Sendo assim, **Moacir**, como **Iracema**, são considerados etimologia ficcional contextual de Alencar.

Outro exemplo de étimo de Alencar é **jacarecanga**. O autor assim o explica, em sua nota “**jacarecanga** – morro de areia na praia do Ceará, afamado pela fonte de água

² Para esta afirmação, consultamos as linhas de concordância, a fim de verificar a datação dos textos.

fresca puríssima. Vem o nome *jacaré* – crocodilo e *acanga* – cabeça” (ALENCAR, 1965, p. 159).

O vocábulo **jacarecanga** não consta dos dicionários de consulta, porém o vocábulo **jacaré** já havia sido dicionarizado por lexicógrafos anteriores à publicação de *Iracema*. Por exemplo, RB define “jacare’, f. m. ou jacareo, (o primeiro mais comum no Brasil) o mesmo, que o crocodilo” (p. 740); e LMSP define “Jacare – s. m. Assim chamão no Brasil ao crocodilo” (s/d). Em textos escritos no Brasil do século XIX, a primeira ocorrência do vocábulo **jacarecanga** foi também por Alencar em *Iracema*, conforme Figura 5.

Figura 5: Linhas de concordância do vocábulo **jacarecanga**

The screenshot shows the 'Corpus do Português: Gênero/Histórico' interface. The 'CONTEXTO' tab is selected. The search results are displayed in a table with columns for document ID, page, and context. The word 'jacarecanga' is highlighted in green in the context snippets.

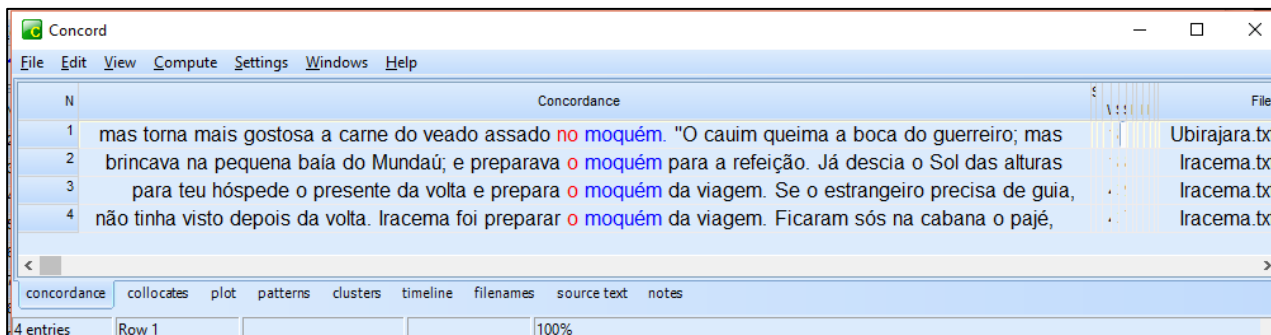
Corpus do Português: Gênero/Histórico			
PESQUISAR	FREQÜÊNCIA	CONTEXTO	AJUDA
SECCÖES: s19 (3)			
CLIQUE NO TÍTULO PARA MAIS CONTEXTO		<input type="checkbox"/> [?]	SHOW DUPLICATES
1	18:Alencar:Iracema	A B C	morro de areia; pela semelhança com a cabeça do crocodilo o chamavam os pescadores Jacarecanga . Do seio das brancas areias escald
2	18:Alencar:Iracema	A B C	banho. Segue na areia o rasto de Coatiabo, e sobe ao alto da Jacarecanga . Aí encontra o guerreiro em pé no cabeço do monte, com os o
3	18:Alencar:Iracema	A B C	, caminhava a seu lado. Oito luas havia que ele deixara as praias de Jacarecanga . Vencidos os guaraciabas, na baía dos papagaios, o guer

Fonte: *Corpus do Português* (DAVIES, 2016).

Sendo assim, Alencar compõe o vocábulo **jacarecanga**, a partir de *jacaré* e *acanga*, porque seu desejo era, por meio do nome, propor também uma descrição do morro cuja característica se assemelha a uma cabeça de jacaré. Portanto, *jacarecanga* é um étimo ficcional contextual indígena de Alencar.

Outro vocábulo, **moquém**, foi analisado em Autor (2004) como “tipo de comida indígena, que também era carregada em viagens”. Contudo, um estudo mais aprofundado da origem e do emprego desse vocábulo nos levou a uma análise distinta. Essa percepção foi motivada pela análise do vocábulo em *Ubirajara*, no qual deixa pistas mais precisas do significado do vocábulo. Na Figura 6, é possível notar as linhas de concordância em que o vocábulo **moquém** é empregado em *Iracema* e *Ubirajara*. Na linha 1, está claro que se trata de um lugar onde se assam as carnes, porém nas demais linhas em que o autor o emprega em *Iracema*, não deixa claro se se trata de uma comida ou lugar onde se prepara a comida.

Figura 6: Linhas de concordância com o vocábulo **moquém**

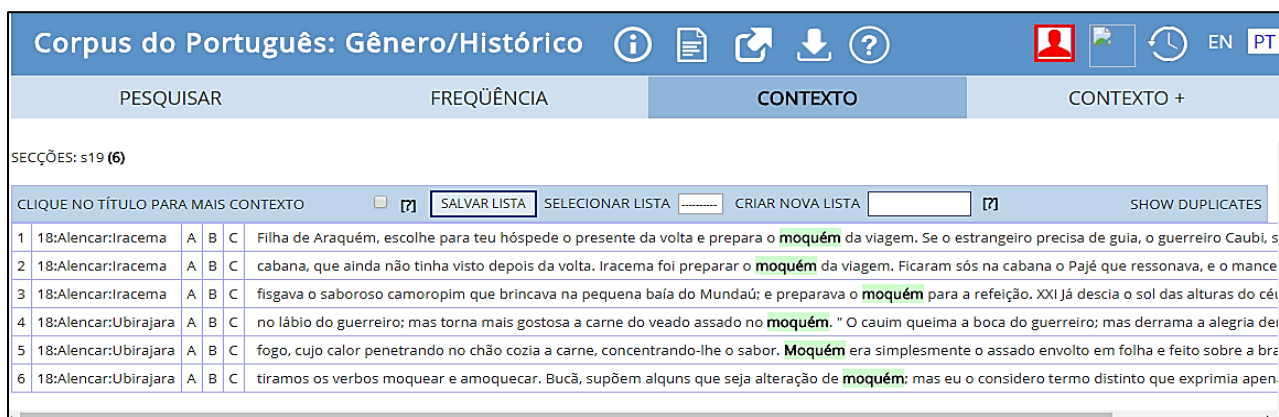


Fonte: Extraído por meio da ferramenta *Concord* do WST.

A fim de dirimir a dúvida em relação ao emprego em *Iracema*, inicialmente, em consulta aos dicionários, notamos que GD traz um verbete para o vocábulo e assim o define “**Moquém** – s. Vila do Estado de Goiás. De *mo-caê*, fazer assar, espécie de grelha de varas para assar peixe, carne. Var. Muquem”. (p. 571). Buscamos também as notas do autor em seus romances. Em *Iracema*, Alencar explica “**Moquém** — Do verbo *mocáem* — assar na labareda. Era a maneira por que os indígenas conservavam a caça para não apodrecer, quando a levavam em viagem. Nas cabanas a tinham no fumeiro” (ALENCAR, 1965, p. 152).

O vocábulo é explicado por processos de decomposição diferentes em Alencar e GD. Alencar afirma ser do verbo *mocáem* que já significa assar na labareda, porém GD diz ser formado por *mo-caê* que é uma espécie de grelha de varas para assar que viria de “fazer assar”. Embora o *Corpus do Português* (DAVIES, 2016) apresente ocorrências do vocábulo **moquém** apenas nas duas obras de Alencar, *Iracema* e *Ubirajara*, há uma diferença em relação à quantidade de ocorrências. Há quatro ocorrências no *corpus* de estudo e, no *Corpus do Português*, seis, como se constata na Figura 7.

Figura 7: Linhas de concordância com o vocábulo **moquém**



Fonte: Extraído do *Corpus do Português* (DAVIES, 2016).

Em razão desta diferença em relação à quantidade de ocorrências, buscamos ampliar o contexto de abonação do vocábulo no *Corpus do Português* em relação ao livro *Ubirajara* e constatamos que se trata de uma nota do autor. Como as notas foram retiradas do *corpus* de estudo, justifica-se, então, a diferença em relação às ocorrências. Notamos também que Alencar explica em *Ubirajara* que “**Moquém** era simplesmente o assado envolto em folha e feito sobre a brasa”, conforme Figura 8.

Figura 8: Contexto ampliado do vocábulo **moquém**

The screenshot shows the 'Corpus do Português: Gênero/Histórico' interface. The search results are displayed in a table with columns for 'PESQUISAR', 'FREQUÊNCIA', 'CONTEXTO', and 'CONTEXTO +'. The search results show the title 'Ubirajara' and the author 'José de Alencar'. Below the search results, there is a section titled 'Contexto ampliado:' which contains a paragraph of text from the book 'Ubirajara' by José de Alencar. The text describes the process of making 'moquém', a type of food made by wrapping meat in banana leaves and cooking it over a fire. The word 'moquém' is highlighted in green in the text.

Fonte: Extraído do *Corpus do Português* (DAVIES, 2016).

Não resta dúvida de que Alencar emprega **moquém** tanto para designar o processo de assamento ou o local onde se assam as carnes, como na definição de GD e sua nota, como o utiliza também como o alimento. Provavelmente, trata-se de carnes que passam pelo processo de assamento, criando uma etimologia ficcional contextual para **moquém** no sentido de comida.

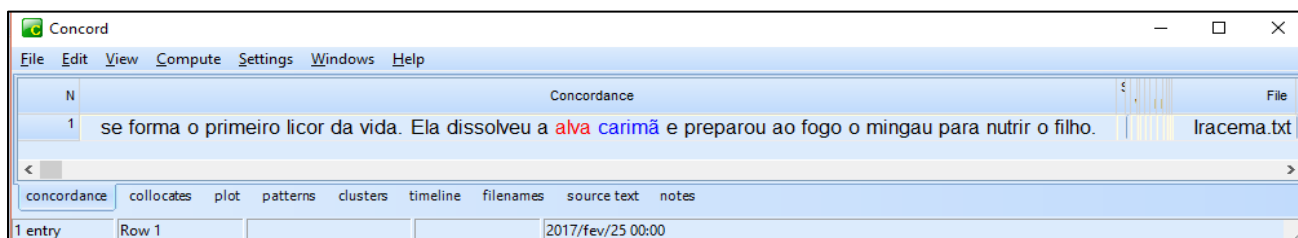
Outro vocábulo relacionado ao campo semântico alimentos é **carimã**, que é um vocábulo criado e utilizado pela primeira vez em textos escritos em língua portuguesa por Alencar. Isso pode ser comprovado no *Corpus do Português* (DAVIES, 2016), em que as duas ocorrências no século XIX estão em Alencar. A Figura 9 nos mostra as duas ocorrências de **carimã**, ambas em textos de Alencar, sendo a primeira em *Iracema*.

Figura 9: Linhas de concordância com o vocábulo **carimã**

The screenshot shows the 'Corpus do Português: Gênero/Histórico' interface. The search results are displayed in a table with columns for 'PESQUISAR', 'FREQUÊNCIA', 'CONTEXTO', and 'AJUDA'. The search results show the title 'Iracema' and the author 'José de Alencar'. Below the search results, there is a section titled 'SECCÕES: s19 (2)'. The text 'carimã' is highlighted in green in the text. The interface also shows a search bar and a list of search results.

Fonte: Extraído do *Corpus do Português* (DAVIES, 2016).

Figura 10: Linhas de concordância com o vocábulo **carimã**



Fonte: Extraído do *Corpus do Português* (DAVIES, 2016).

Além disso, os dicionários consultados GD, LMSP, RB e AMS não trazem o vocábulo **carimã** entre seus verbetes. Talvez, por isso, Alencar teve a preocupação, pelo fato de não ser de conhecimento dos leitores, de trazer uma nota explicando o vocábulo “**Carimã** — Uma conhecida preparação de mandioca. *Caric* — correr, *mani* — mandioca: mandioca escorrida” (ALENCAR, 1965, p. 160).

Note-se, então, que Alencar cria uma etimologia ficcional contextual ao explicar o significado do vocábulo, por meio da decomposição *caric* – correr + *mani* – mandioca.

Assim sendo, Alencar vai estabelecendo sua etimologia ficcional contextual na medida em que deseja aproximar o significado da palavra com o elemento nomeado. Um dos recursos utilizados pelo autor é a decomposição de vocábulos já existentes e, com os elementos adquiridos ele cria os étimos associando os significados dos fragmentos e estabelecendo um novo significado com o novo étimo. Por isso, a utilização de frases explicativas no próprio texto ou a recorrência a notas explicativas ao final dos romances.

Vale ressaltar que os étimos criados por Alencar não são aleatórios, existe, por trás de cada um, uma razão que é esclarecida no decorrer da narrativa, nas notas ou nos textos de defesa contra as críticas de outros autores escritos após a publicação de cada romance. Alencar cuida de dar nomes aos elementos do mundo indígena de acordo com a etimologia de palavras de origem Tupi ou Guarani, dentre outras línguas indígenas. A maioria dos vocábulos indígenas é utilizada para atender aos desígnios do autor e cumprir seu objetivo de criação de uma língua brasileira, a partir da figura do indígena.

CONSIDERAÇÕES FINAIS

Com base nos estudos e nas análises apresentadas, é possível afirmar que José de Alencar adquiriu um vasto conhecimento sobre a língua indígena, o que se comprova também ao se analisar as notas que o autor coloca no final dos seus romances.

Como se pode observar, os étimos ficcionais contextuais foram criados pelo autor a fim de nomear os personagens e elementos do cenário de suas obras. Especificamente

tratamos de alguns étimos indianistas. Com base nas teorias sobre Etimologia, nos conceitos sobre ficção e contexto, elaboramos a expressão “Etimologia Ficcional Contextual” para dar conta da análise das criações dos étimos por Alencar, com base em seus propósitos contextuais. Consideramos Etimologia Ficcional Contextual, porque Alencar cria os étimos para imprimir aos romances aspectos do seu intuito, no que se refere à criação literária e à emancipação da língua portuguesa brasileira, em relação aos europeus.

Ratificando nosso zelo em relação à Etimologia Ficcional Contextual de Alencar, procedemos à consulta de quatro dicionários, sendo um deles específico de língua indígena e três de língua portuguesa. Ambos os dicionários com publicação anterior à publicação dos romances de Alencar. Como complemento aos dicionários, utilizamos o *Corpus do Português* (DAVIES, 2016) para consulta dos vocábulos. Esse *corpus* nos revelou, por meio da consulta na versão histórica, que os vocábulos indianistas consultados foram empregados por Alencar, pela primeira vez, em textos escritos no Português do Brasil.

Os estudos sobre os étimos ficcionais contextuais de Alencar estão longe de se esgotarem neste texto, porém nosso estudo propõe um início para outras pesquisas que poderão ser proveitosas. Por exemplo, é possível estabelecer um estudo considerando apenas dos antropônimos, ou os topônimos; da mesma maneira, é possível, com base nos 367 vocábulos indígenas identificados, separá-los em campos semânticos e recortar o estudo.

Sendo assim, este texto pretende iniciar os estudos do léxico indianista de Alencar com base na Etimologia Ficcional Contextual, como também deixa o conceito aberto à utilização para análise de vocábulos de outros autores e de outras etimologias.

REFERÊNCIAS

AIRES DE CASAL, M. 1754?-1821?. **Corografia brasílica ou Relação histórico-geográfica do Reino do Brasil**. Belo Horizonte: Ed. Itatiaia, 1976.

AIRES DE CASAL, M. **Corografia Brasílica**. Tomo 1. Rio de Janeiro: Imprensa Nacional, 1945.

ALENCAR, J. **Iracema**. Ed. do centenário. Rio de Janeiro: Livraria José Olympio, 1965.

ALENCAR, J. **O Guarani**. Tomo 1º. Rio de Janeiro: Livraria José Olympio, 1951.

ALENCAR, J. **O Guarani**. Tomo 2º. Rio de Janeiro: Livraria José Olympio, 1951.

ALENCAR, J. **Ubirajara**. Rio de Janeiro: Edições de Ouro. s/d.

ÁVILA, M.V.D.; NOVODVORSKI. A

BLUTEAU, R. **Vocabulário Portuguez & Latino**. Coimbra: Collegio das Artes da Companhia de Jesus, 1712-1728.

CASARES, J. **Introducción a la Lexicografía moderna**. 3 ed. Madrid: Raycar, 1992.

DAVIES, M. **Corpus do Português**. 2016. Disponível em: www.corpusdoportugues.org. Acesso em: 18 out. 2019.

DUBOIS, J. et al. **Dicionário de Linguística**. 14 ed. São Paulo: Cultrix, 2007.

GONÇALVES DIAS. A. **Dicionário da língua Tupy**. Lipsia: F. A. Brockhaus, 1858.

HOUAISS, A. **Dicionário eletrônico Houaiss da língua portuguesa**. Versão 3.0, 2009.

NAVARRO, E. A. **Dicionário de tupi antigo: a língua indígena clássica do Brasil**. São Paulo: Global, 2013.

PINTO, L. M. S. **Dicionário da Língua Brasileira**. Ouro Preto: Typographia de Siva, 1832.

SILVA, A, M. **Diccionario da Lingua Portugueza**. Lisboa: Officina de Simão Thaddeo Ferreira, 1789.

VARNHAGEM. F. A. **História Geral do Brasil**. 1816-1878. 10 ed. Belo Horizonte: Ed. Itatiaia; São Paulo: Ed. Universidade de São Paulo, 1981.

VIARO, M. E. **Etimologia**. São Paulo: Contexto, 2014.

Como citar este artigo (ABNT)

ÁVILA, M.V.D.; NOVODVORSKI.A. Etimologia ficcional contextual: o léxico indianista em Alencar à luz da linguística de corpus. SELL, Uberaba, MG, v. X, n. X, p. XXX-XXX, 2019. Disponível em: <inserir link de acesso>. Acesso em: inserir dia, mês e ano de acesso. DOI: inserir link do DOI.

Como citar este artigo (APA)

Ávila, M.V.D. & Novodvorski.A. (2020). Etimologia ficcional contextual: o léxico indianista em Alencar à luz da linguística de corpus. SELL, X (X), XXX-XXX. Recuperado em: inserir dia, mês e ano de acesso de inserir link de acesso. DOI: inserir link do DOI.