

# Macabéa

Revista Eletrônica do Netli, Volume 8, Número 2, Jul.-Dez., 2019

## CONTRIBUIÇÕES DA LINGUÍSTICA DE CORPUS PARA UMA PESQUISA SITUADA NA INTERFACE ENTRE SEMÂNTICA LEXICAL COGNITIVA E LEXICOGRAFIA ONOMASIOLÓGICA



## CONTRIBUTIONS OF CORPUS LINGUISTICS TO RESEARCH IN THE FIELD IN COGNITIVE LEXICAL SEMANTICS AND ONOMASIOLOGICAL LEXICOGRAPHY

DIEGO SPADER DE SOUZA  
UNIVERSIDADE DO VALE DO RIO DOS SINOS, Brasil

[RESUMO](#) | [INDEXAÇÃO](#) | [TEXTO](#) | [REFERÊNCIAS](#) | [CITAR ESTE ARTIGO](#) | [O AUTOR](#)  
RECEBIDO EM 14/05/2019 • APROVADO EM 01/06/2019

---

### Abstract

---

The present article is part of the author's doctorate dissertation, which aimed at developing the bases for onomasiological dictionaries based on Cognitive Lexical Semantics, using, especially, Charles Fillmore's Frame Semantics. In this article, the focus

is on the aspects related to the use of Corpus Linguistics, especially related to the potential of corpus usage for the understanding of the domain addressed in the dissertation, which is that of the Italian immigration cooking. From both the dissertation and the article, it was possible to verify the high relevance of the notion of corpus for research in Cognitive Linguistics, considering the possibility of working with naturalistic data that, apart from the linguistic dimension, provide social and historical information.

---

## Resumo

---

O presente artigo é um recorte da Tese de doutorado do autor, cujo objetivo foi desenvolver bases para dicionários onomasiológicos baseados na Semântica Lexical Cognitiva, utilizando, em especial, a Semântica de *Frames*, de Charles Fillmore. Neste artigo, o enfoque são os aspectos relacionados ao uso da Linguística de *Corpus*, especialmente relacionados ao potencial da utilização de *corpus* para o entendimento do domínio abordado na Tese, que é o da culinária de imigração italiana. A partir tanto da Tese quanto deste artigo, foi possível verificar a alta relevância da noção de *corpus* para a pesquisa em Linguística Cognitiva, tendo em vista a possibilidade de se trabalhar com dados naturalísticos que, para além da dimensão linguística, fornecem informações de cunho social e histórico.

---

## Entradas para indexação

---

**KEYWORDS:** Frame Semantics. Lexicography. Corpus Linguistics.

**PALAVRAS CHAVE:** Semântica de Frames. Lexicografia. Linguística de Corpus.

---

## Texto integral

---

### INTRODUÇÃO

O presente artigo é um recorte da Tese de doutorado do autor (SOUZA, 2019), cujo objetivo foi desenvolver bases para dicionários onomasiológicos baseados nos construtos teóricos da Semântica Lexical Cognitiva, utilizando, em especial, a Semântica de *Frames*, de Charles Fillmore.

Neste artigo, voltamo-nos para os aspectos relacionados ao uso da Linguística de *Corpus*, especialmente relacionados ao potencial da utilização de *corpus* para o entendimento do domínio abordado na Tese, que é o da culinária de imigração italiana.

A onomasiologia, em linhas gerais, designa um método para a pesquisa lexical em que se parte, inicialmente, de um **conceito** – que se trata de uma ideia, um pensamento, um esquema abstrato –, para que se chegue, então, às possibilidades de realização linguística, isto é, às **formas linguísticas** atreladas a esse conceito.

Nesse sentido, importa salientarmos que a relação entre a Lexicografia e a onomasiologia não é, de modo algum, uma novidade. Sendo esse um conceito pertencente à escola estruturalista da Lexicologia, a Lexicografia sempre esteve presente nas discussões sobre onomasiologia. Como já dissemos, neste artigo, abordamos uma interface com os estudos da Semântica Lexical Cognitiva, subárea da Semântica Cognitiva, que, por sua vez, pertence ao quadro maior da Linguística Cognitiva. Além disso, utilizamos a teoria da Semântica de *Frames* (FILLMORE, 1982) como principal aporte teórico dentro da Semântica Lexical Cognitiva.

O presente artigo está dividido em 5 seções. A primeira busca articular conceitos da Semântica Lexical Cognitiva com o campo da onomasiologia. A segunda seção aborda o encontro entre a área da Linguística Cognitiva e a Linguística de *Corpus* como método. Dando seguimento, a terceira seção trata dos procedimentos metodológicos, focando no processo de compilação de *corpora*. A quarta seção aborda a utilização do *corpus* de apoio e o seu potencial na descrição do domínio estudado, que inclui, inegavelmente, uma dimensão histórica e social. A última seção apresenta nossas considerações finais.

## ARTICULANDO SEMÂNTICA LEXICAL COGNITIVA E ONOMASIOLOGIA

### SOBRE A ONOMASIOLOGIA E A LEXICOGRAFIA ONOMASIOLOGICA

A onomasiologia designa um fenômeno lexical, no qual se parte de um determinado conceito ou ideia e se chega às formas de realização linguística daquele conceito. A onomasiologia, portanto, é o oposto da semasiologia, no qual se parte de uma palavra ou expressão para então chegar à definição. Baldinger (1966) aborda os fenômenos semasiológico e onomasiológico no contexto da evolução da Linguística no final do século XIX e início do XX, quando se passou do **som** à **palavra**. A partir do impacto causado por Ferdinand de Saussure, a Linguística não só se estabeleceu como uma ciência moderna e independente, mas também iniciou uma fase de estudos voltados aos aspectos estruturais estrutura das línguas.

Os conceitos de semasiologia e a onomasiologia surgem nesse contexto histórico, no início do século XX, inicialmente atraídos pela pesquisa histórica (diacrônica), mesmo que também a serviço dos estudos estruturalistas (que seguiam a perspectiva sincrônica). Segundo Baldinger (1966), a onomasiologia implica desde o início uma preocupação de ordem estrutural.

Foi no campo da Semântica Estrutural que se desenvolveu a teoria dos campos lexicais (ou campos semânticos), de Jost Trier, que segue orientação onomasiológica, uma vez que trata de agrupamentos lexicais em redor de um conceito comum. Para Ullmann (1964), a teoria de Trier fornece meios valiosos para se lidar com um problema que é, ao mesmo tempo, difícil e de muita importância, que é a influência da linguagem no pensamento.

Em relação à semasiologia, Baldinger (1966) apresenta considerações de cunho mais histórico, ainda que defenda que “[...] o estabelecimento dos campos semasiológicos é a tarefa central de qualquer léxico alfabético e sincrônico” (p. 13). Ao pensarmos no significado de uma palavra, somos levados a pensar nos usos e nos variados contextos em que tal palavra é utilizada. Em um dicionário alfabético e semasiológico, substitui-se contextos por uma definição. E essa definição, que é a estrutura semasiológica da palavra, é de grande importância para a interpretação de qualquer um dos contextos que uma lexia possa apresentar.

Posto que ainda tivesse muito a se dizer sobre a estrutura semântica de uma determinada palavra, é fato que, primeiramente, nos voltaríamos à estrutura semasiológica. Conforme Guiraud (1972), toda palavra tem uma espécie de “nó semântico” que é mais ou menos denso, mais ou menos volumoso, que é cercado por uma auréola de associações secundárias, afetivas ou sociais. O nó é o próprio campo semasiológico, é a definição. Todo o resto compõe o conjunto de contextos de usos pessoais.

Somente a semasiologia, portanto, permite, partindo de um grande número de contextos de uso, estabelecer as generalizações de uma significação específica. A semasiologia trata, assim, da passagem do nível periférico para o central, ou do específico para o genérico. Um dicionário semasiológico, organizado alfabeticamente, não costuma se deter a cada uso particular que uma palavra possa apresentar.

O percurso onomasiológico, por sua vez, inicia com um elemento mais geral, que é o conceito, e parte em busca daquilo que é mais específico, os itens linguísticos de um léxico em questão que se relacionam a esse conceito. Pode-se entender melhor a onomasiologia a partir da relação que esta mantém com a semasiologia e vice-versa.

A interdependência entre a semasiologia e a onomasiologia, segundo Baldinger (1966), está relacionada à própria estrutura do léxico das línguas. Vejamos como isso se dá no campo da diacronia: ao verificarmos o campo semasiológico de uma determinada palavra ao longo do tempo, teremos possivelmente uma lista de significados. Espera-se que, em alguma medida, esses significados possam ou não apresentar relações. Em um estudo diacrônico, Baldinger (1966) percebeu que “trabalhar” e “laborar” faziam parte do conjunto de significações da palavra “trabalho”. A posição dessas significações no campo semasiológico determinada também a sua posição no campo onomasiológico de “trabalho”.

A semasiologia e a onomasiologia olham para as duas microestruturas fundamentais do léxico. A onomasiologia, segundo Baldinger (1966), promete resultados mais inovadores, principalmente levando em consideração a relação intrínseca que se mantém com aspectos sociais e pragmáticos. Dessa forma, o autor já via a necessidade de que fossem considerados aspectos exteriores ao puro sistema linguístico, ou seja, que perpassam a estrutura linguística.

Baldinger (1966) defende que se tenha dois tipos distintos de dicionários: um que parte de uma listagem alfabética (ou fonológica) e que liste significados – ou seja, o dicionário semasiológico, ao qual já estamos mais do que habituados – e um

que parta de conceitos, que seja onomasiológico. Falemos, brevemente, sobre o dicionário onomasiológico.

Segundo Babini (2001), o dicionário onomasiológico foi concebido para dar conta das dificuldades que usuários enfrentam ao consultar um dicionário tradicional. Mesmo que a lista alfabética seja consideravelmente fácil de lidar tanto para quem produz o dicionário quanto para quem o utiliza, é possível que o usuário tenha uma ideia do significado da palavra que quer encontrar, mas não a sua forma. É possível, ainda, que ele esteja procurando palavras relacionadas a um dado verbete, que podem ser sinônimos, antônimos etc. Nesses casos, o dicionário tradicional pode não atender às necessidades do usuário. O dicionário onomasiológico, portanto, ao apresentar uma estrutura que parte dos conceitos para os verbetes, pode resolver tais dificuldades.

Uma das vantagens do dicionário onomasiológico é a de apresentar muitas palavras semanticamente relacionadas, aumentando a capacidade de expressão do usuário, uma vez que este tem acesso a um inventário muito maior de palavras e expressões a sua disposição. Um dos objetivos do dicionário onomasiológico, nesse sentido, é, também, auxiliar a escrita e a expansão do vocabulário. É esse, por exemplo, um dos objetivos do *Oxford School Thesaurus* (2016, p. vi)<sup>1</sup>:

Um tesouro oferece alternativas – muitas vezes mais interessantes e coloridas – às palavras que você já conhece e usa; estes são conhecidos como sinônimos. Em alguns casos, ele também fornece o oposto de palavras, conhecidas como antônimos. Usar um tesouro regularmente irá estender o seu vocabulário e ajudá-lo a ser mais preciso e imaginativo na maneira como você se expressa.

Consoante Babini (2006), o dicionário onomasiológico tem o dever de resolver o problema inverso daquele de um dicionário semasiológico. Em outras palavras, para uma determinada ideia ou conceito, o dicionário de orientação onomasiológica deve apresentar a relação de palavras presentes em um determinado léxico que se usem a – ou evocam – essa ideia ou conceito. No que diz respeito ao percurso onomasiológico e à estruturação de um dicionário desse tipo, Babini (2001) lista seis possibilidades (que podem ou não ser utilizadas simultaneamente): (i) pelo sistema nocional ou plano de classificação dos conceitos, apresentados no início da obra lexicográfica, (ii) pela classificação sistemática das entradas, (iii) pelo conteúdo semântico das entradas, (iv) pela sinonímia, (v) pela antonímia, ou (vi) por analogia.

Essas diferentes possibilidades dão vida a obras igualmente diferentes, o que nos leva a uma tipologia dos dicionários onomasiológicos, que inclui os *thesaurus*, os dicionários de sinônimos, os dicionários ilustrados etc. Na próxima subseção, voltamos-nos à Semântica Lexical Cognitiva e a teoria da Semântica de *Frames*.

Considera-se a Semântica Lexical Cognitiva como uma subárea da Semântica Cognitiva que aplica os construtos teóricos desenvolvidos no âmbito dessa abordagem aos estudos lexicais. Nas palavras de Taylor, Cuyckens e Dirven (2003, p. 1, grifo nosso),<sup>2</sup> para a Linguística Cognitiva, “[...] itens lexicais, bem como classes de palavras e construções gramaticais, são **categorias conceptuais** que devem ser estudadas e investigadas com respeito a sua função cognitiva [...]”, o que contrasta diretamente com modelos mais formais de descrição semântica. Por conseguinte, o léxico, tanto quanto as estruturas gramaticais, designa um repositório de conhecimento de mundo, cuja estrutura não apenas armazena informações relevantes e significativas sobre experiências passadas, como também auxilia na forma como lidamos com novas experiências (cf. GEERAERTS, 1995).

Geeraerts (2009) aborda a Semântica Lexical Cognitiva como parte da tensão entre as visões **minimalistas** e **maximalistas** dos estudos lexicais. A abordagem estruturalista, a partir da qual se desenvolveu o modelo tradicional de onomasiologia, representa uma visão minimalista, em que se tenta a todo custo manter uma diferenciação clara entre o que é o território da Semântica e o que é o território da Pragmática. Busca-se, nessa visão de semântica lexical, impor uma distinção entre o que deve ser levado em conta e o que deve ser rejeitado em relação à descrição semântico-lexical. Segundo o autor (2009, p. 182),<sup>3</sup> essa tensão pode tomar diferentes formas e se relacionar a diferentes aspectos das teorias semânticas, tais como:

[...] a velha questão da fronteira entre o conhecimento da palavra e o conhecimento do mundo. [...] a linha divisória entre semântica e pragmática. [...] uma escolha metodológica entre um modo de investigação orientado para a estrutura ou orientado para o uso. Ou [...] o grau de realismo cognitivo que uma teoria do significado das palavras deveria tentar alcançar.

A Semântica Lexical Cognitiva, nesse sentido, assume um posicionamento maximalista, uma vez que (i) defende a hipótese de que a estrutura semântica é a própria estrutura conceptual, (ii) rejeita uma distinção precisa entre estrutura semântica e estrutura pragmática e (iii) assume uma postura voltada para o uso da linguagem. Isso não significa dizer, contudo, que a Semântica Cognitiva não trata das estruturas linguísticas; significa, contudo, dizer que as estruturas linguísticas são pontos de acesso para estruturas cognitivas, conceptuais, que vêm da experiência com o mundo, numa relação contínua entre linguagem e conhecimento.

Um dos primeiros pontos levantados por Geeraerts (2009) em relação a uma Semântica Lexical Cognitiva é a teoria dos protótipos, que caracteriza a forma como o cognitivismo entende o fenômeno da categorização. Assim, para Taylor, Cuyckens e Dirven (2003), a Semântica Lexical Cognitiva tem se mantido um

campo vibrante de investigações dentro do cognitivismo desde quando, na década de 80, pesquisadores foram capazes de transpor as descobertas da Psicologia Cognitiva acerca da estrutura interna de categorias – estruturas de propósito e estruturas de semelhança de família – para o estudo das categorias lexicais. Assim, a teoria prototípica proposta por Eleanor Rosch no campo da Psicologia Cognitiva foi o ponto de partida para uma visão fresca e inovadora da estrutura lexical.

Dessa forma, a partir das pesquisas de Rosch sobre categorias prototípicas e a sua influência no arcabouço teórico-metodológico da Semântica Cognitiva como um todo, os grandes interesses da Semântica Lexical Cognitiva, nas últimas décadas, têm sido:

[...] (i) estrutura interna das categorias lexicais (estrutura prototípica, estrutura de semelhança de família, estrutura de rede radial); (ii) a natureza polissêmica de itens lexicais e os princípios cognitivos (e.g., metáfora, metonímia, transformações de esquemas de imagem) que motivam as relações entre os diferentes sentidos dos itens lexicais, (iii) estruturas conceptuais mais amplas (e.g., pesquisas sobre metáfora, Semântica de *Frames*). (TAYLOR, CUYCKENS, DIRVEN, 2003, p. 2)<sup>4</sup>

A Semântica de *Frames*, de fato, tem sido um dos grandes interesses da Semântica Lexical Cognitiva nos últimos anos, tendo em vista o seu potencial para a descrição das estruturas conceptuais evocadas por itens lexicais. Como já dito anteriormente, nossa proposta de modelo de um dicionário onomasiológico baseado nos conceitos da Semântica Lexical Cognitiva parte, principalmente, do raciocínio que orienta e fundamenta a Semântica de *Frames* concebida por Charles J. Fillmore.

A Semântica de *Frames* surge a partir da proposta de Fillmore de se tratar a análise do significado em termos de *frames* semânticos. Um *frame*, segundo o autor (1975, 1982), caracteriza um sistema coerente de conceitos, de forma que a compreensão de um desses conceitos depende da compreensão do sistema como um todo. Nesse sentido, diz-se que o *frame* funciona como uma **esquematização da experiência**: todo o conhecimento que um falante tem sobre um determinado evento ou situação ganha representação no nível conceptual; essa representação cognitiva é entendida por Fillmore como sendo um *frame*.

A teoria caracteriza um modelo empírico para a análise semântica (FILLMORE, 1982; PETRUCK, 1996). Em consonância com os princípios da Semântica Cognitiva, e, em consequência disso, da Semântica Lexical Cognitiva, a Semântica de *Frames* adota uma posição enciclopédica. Segundo Petruck (1996), o modelo concebido por Fillmore enfatiza as continuidades entre linguagem e experiência.

Fillmore (1982) liga o conceito de *frame* ao de **protótipo** – e às pesquisas de Eleanor Rosch (1973, 1975). Contrariando a ideia de descrever o significado em termos de condições de verdade, uma perspectiva baseada em *frames* deve supor que a compreensão dos significados comumente apela para um protótipo. Uma das

formas de unir esses dois conceitos, para Fillmore, é através da tese de que *frames* são instâncias prototípicas da experiência humana. O entendimento dos itens lexicais pelos falantes subjaz as instâncias prototípicas dos *frames*. Consideremos o pequeno diálogo a seguir:

Paula: “Ontem ganhei um monte de presentes!”

Pedro: “Nossa! Ontem era seu aniversário?”

Quando Paula fala para Pedro que ganhou um monte de presentes presentes, não é necessário que ela diga que o dia anterior era seu aniversário para que ele acesse o *frame*. Isso porque, dentre todos os elementos que pertencem ao *frame* de aniversário, está lá o que diz que aniversariantes ganham presentes. Ainda, Pedro poderia ter complementado “Teve bolo?”, uma vez que a ideia de ganhar presentes no aniversário está comumente relacionada a de festas de aniversário, em que há convidados e bolo (“Apagou as velinhas?”). Nesse sentido, portanto, o *frame* é um sistema de conceitos relacionados, de modo que a ativação do *frame* ativa não só um certo aspecto dele, mas a estrutura como um todo (FILLMORE, 1982).

Passemos, agora, à discussão sobre a relação entre o quadro teórico da Linguística Cognitiva (que inclui a Semântica de *Frames* e, de modo mais abrangente, a Semântica Lexical Cognitiva) e a Linguística de *Corpus*.

## A LINGUÍSTICA COGNITIVA ENCONTRA A LINGUÍSTICA DE *CORPUS*

Langacker (1987) define sua teoria, a Gramática Cognitiva, que designa um dos modelos mais influentes dentro da Linguística Cognitiva, como um **modelo baseado no uso**. Esse termo pode ser atribuído à Linguística Cognitiva como um todo, incluindo a Semântica Lexical Cognitiva. Segundo Langacker (1987, p. 220)<sup>5</sup>, “uma motivação para esse termo é a afirmação de que eventos de uso são a fonte de todas as unidades linguísticas.” No momento em que o cognitivista fala de unidades linguísticas, ele fala não só de unidades gramaticais, mas também de unidades lexicais, uma vez que a Linguística Cognitiva rejeita fronteiras rígidas entre os níveis da gramática e do léxico, apostando, por sua vez, em uma ideia de contínuo. De acordo com Bybee, em entrevista concedida a Torrent (2012), o cognitivista utiliza o termo para enfatizar o conhecimento específico que falantes de uma língua têm de expressões idiomáticas e construções, sendo que esse conhecimento vem da experiência de tais falantes no mundo. Essa mesma experiência interage com o aparato cognitivo no intuito de formar as representações mentais necessárias para que o uso produtivo da linguagem seja possível (TORRENT, 2012).

Sendo assim, compreendemos que a Linguística Cognitiva firma um compromisso com o estudo da linguagem a partir do uso real feito pelos falantes, o

que mostra a necessidade de que o cognitivismo se alie a métodos empíricos de análise. A pesquisa aqui descrita, estando pautada na proposta de um dicionário onomasiológico baseado na Semântica Lexical Cognitiva, se estabelece como um estudo lexical, de modo que adotamos a Linguística de *Corpus* como abordagem metodológica. O maquinário da Linguística de *Corpus* nos permite analisar o comportamento de unidades léxicas no uso real da linguagem, partindo de *corpus* que designam coleções de dados linguísticos (escritos ou falados) sistematizados de acordo com certos critérios e compilados com o propósito de fornecer resultados para a análise linguística (cf. BERBER SARDINHA, 2004). A Linguística de *Corpus*

[...] ocupa-se da coleta e exploração de corpora, ou conjuntos de dados linguísticos textuais que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem através de evidências empíricas, extraídas por meio de computador.

A Linguística de *Corpus* está contextualizada na era da informática, o que permite que grandes porções de texto sejam processadas de forma totalmente eletrônica. Cabe salientar que, dependendo do tamanho do *corpus*, o tratamento manual seria ou muito demorado ou até mesmo humanamente impossível. Vale ressaltar, no entanto, que a Linguística de *Corpus* não inventou o conceito de *corpora*, uma vez que esse termo, que significa “corpo”, já designava coleções de textos e documentos desde a Grécia Antiga (BERBER SARDINHA, 2000).

McEnery e Hardie (2012) abordam duas formas de se utilizar a Linguística de *Corpus*: a partir de uma orientação **baseada** em *corpus*, ou *corpus-based*, e uma orientação **dirigida** pelo *corpus*, ou *corpus-driven*. Pesquisas baseadas em *corpus* partem do uso dos dados obtidos em *corpora* como forma de validar hipóteses já existentes na literatura. É, assim, a ideia de se usar a Linguística de *Corpus* como *método*. Já as pesquisas dirigidas pelo *corpus*, no entanto, rejeitam a concepção de *corpus* como metodologia, compreendendo que o *corpus* deve ser usado como fonte de formulação de hipóteses sobre a linguagem. O *corpus* é, em si mesmo, sua própria teoria (McENERY; HARDIE, 2012). No âmbito deste artigo e da pesquisa maior na qual nos baseamos, adotamos a Linguística de *Corpus* como um **método**, de forma que os *corpora* aqui utilizados servem ao propósito de fornecer os dados de que precisamos para tratar da proposta do dicionário onomasiológico baseado na Semântica Lexical Cognitiva.

A relação entre a Linguística de *Corpus* e as teorias cognitivistas não é total novidade. Podemos citar, por exemplo, a plataforma FrameNet, base de dados baseada em *frames*, cujo trabalho semântico-computacional se dá através de conjuntos de dados oriundos de *corpus*. Gries (2010) é um autor que se posiciona a favor de um encontro entre Linguística de *Corpus* e abordagens cognitivas, ressaltando que, em dado momento, para a Linguística de *Corpus*, assumir uma perspectiva cognitivista é necessário, uma vez que “[...] as coisas só entram em discurso quando um falante as processou e ‘decidiu’ expressá-las e, assim, torná-

las parte do discurso, e a forma como um ouvinte processa essa informação também é determinada pela estrutura interna do ouvinte.” (2010, p. 333). Do ponto de vista da Linguística Cognitiva, também é fácil perceber a relevância da Linguística de *Corpus*, uma vez que utilizá-la como metodologia permite ao pesquisador embasar empiricamente as hipóteses e generalizações, ainda que a abordagem baseada em *corpus* – isto é, que utiliza o *corpus* como método – não exclua totalmente o papel da introspecção do linguista (GRIES, 2010). Utilizar o *corpus* como uma ferramenta da pesquisa fornece os recursos necessários para que as reflexões ocorram a partir de conjuntos de dados naturalísticos; no entanto, ainda cabe ao linguista, fundamentado pela sua base teórica, interpretar os dados.

Tecidas essas considerações a respeito do uso da Linguística de *Corpus* como aporte metodológico às pesquisas desenvolvidas no âmbito da Linguística Cognitiva, passamos à descrição dos procedimentos metodológicos.

## PROCEDIMENTOS METODOLÓGICOS: COLETA E TRATAMENTO DE DADOS

Os procedimentos metodológicos da presente pesquisa foram pautados na coleta de dois *corpora*: um *corpus* de apoio e um *corpus processável*. O *corpus* de apoio faz parte da metodologia de trabalho dos projetos do grupo de pesquisa SemanTec, da UNISINOS, coordenado pela Profa. Dra. Rove Chishman. Designa uma coletânea de textos compilados com o propósito de servir de ferramenta para um estudo preliminar do domínio estudado, possibilitando a familiarização com as estruturas que o compõem (CHISHMAN et al., 2017). O *corpus processável*, por sua vez, é compilado com o propósito de servir para análise e processamento eletrônico, através dos *softwares* da Linguística de *Corpus*. Neste trabalho, utilizamos o *software* Sketch Engine.

O Sketch Engine designa “[...] uma ferramenta de *corpus* que toma como entrada um *corpus* de qualquer linguagem (com marcação linguística apropriada), e que, então, gera, entre outras coisas, *word-sketches* para as palavras dessa linguagem.” (KILGARRIFF et al., 2004, p. 1). A partir de um determinado *corpus*, portanto, o usuário do Sketch Engine tem a possibilidade de fazer a manutenção e a análise dessa coletânea de textos através de um conjunto de recursos, como as *word-sketches*, o concordanceador, a geração de lista de palavras por frequência, entre outras. Neste trabalho, para vias de análise, utilizamos o concordanceador, as *word-sketches* e a lista de palavras.

O Sketch Engine também foi utilizado para a compilação do *corpus processável* em si, a partir da ferramenta *Create corpus*. Ao acessar a função *Create corpus* na página inicial do Sketch Engine, o usuário é levado a uma primeira tela, em que deve inserir as informações iniciais sobre o *corpus* que está a compilar, sendo elas: o nome do *corpus*, o idioma (com a possibilidade de criar um *corpus* multilíngue) e a descrição sobre o *corpus*.

Build your own private corpus from texts on the web or from your own documents.

Name  required

Language  🔍

**MULTILINGUAL**

Description

Storage used: 345,458 of 1,000,000 words (34%)

Available features ▼

BACK NEXT

Ao selecionar a primeira opção, o Sketch Engine dá a opção de o usuário escolher entre URLs, *sites* específicos ou busca na *web*. Além disso, o usuário deve prover ao SE palavras-chave, ou *input-words*, para que o sistema faça a busca por textos para o *corpus*.

← TEXTS FROM WEB

Input type  Web search ?  
 URLs ?  
 Website ?

house X apartment X life X home X ?  
 real estate X

Number of Bing searches: 10

Folder name ?

Web search settings ▼  
 Black list settings ▼  
 White list settings ▼  
 Size restrictions ▼

Compile when finished ?

CANCEL GO

Ao clicar em *Go*, o Sketch Engine inicia o processo de compilação do *corpus*, que fica disponível ao usuário e permite a utilização das ferramentas antes abordadas. Como falamos anteriormente, o Sketch Engine e seus recursos são

utilizados apenas no tratamento do nosso *corpus* processável, de modo que o *corpus* de apoio tem apenas o objetivo de nos permitir conhecer o domínio estudado, servindo de base para a sistematização inicial dos conceitos, ou *frames*, que compõem a especialidade estudada. Seguimos a lógica da Semântica de *Frames* de representação de conceitos em redes conceituais. Consideramos viável, nesse sentido, usar, como recurso, mapas conceituais, uma vez que permitem uma visualização da organização de um domínio ou assunto determinado. Sendo assim, abordamos, também, o *software* CmapTools, utilizado para o desenvolvimento de mapas conceituais.

O Cmap Tools é um *software* gratuito para a criação de mapas conceituais. O método de mapeamento conceitual permite que se criem conceitos, organizando-os em uma estrutura cognitiva, em uma rede de relações entre esses conceitos, sendo tais relações não restritas ao modelo do pensamento cartesiano (DALEY et al., 1999).

A ferramenta CmapTools permite que o usuário crie ou importe mapas conceituais, que podem, depois de prontos, ser exportados em formatos diversos, como PDF e PNG. Também é possível salvar o mapa em formato aberto, para que o usuário continue trabalhando nele em outras ocasiões. A Figura 9, presente no segundo capítulo, caracteriza um mapa conceitual feito no CmapTools. Como podemos perceber pela imagem, o *software* permite a criação de várias relações entre conceitos, que podem ser nomeadas de acordo a natureza da relação.

Em relação à natureza dos *corpora*, iniciemos pelo *corpus* de apoio. Nosso *corpus* de apoio se subdivide em três categorias: *websites* de gastronomia e turismo, *websites* jornalísticos e materiais acadêmicos. As duas primeiras categorias compreendem textos escritos para blogs e jornais, com o intuito de comentar a culinária da imigração italiana na Serra Gaúcha, muitas vezes sob o ponto de vista do turismo, utilizando essa gastronomia como atrativo. A terceira categoria, materiais acadêmicos, compreende artigos e trabalhos de conclusão de curso que tomam a culinária de imigração como objeto de estudo científico, identificando aspectos históricos e culturais, o que nos auxilia muito no entendimento do domínio.

Os itens elencados foram selecionados para o *corpus* de apoio a partir de buscas feitas no Google, de modo que essas buscas seguissem alguns critérios: primeiramente, as buscas foram todas feitas com palavras-chave inseridas dentro de aspas, para que os sites retornados na busca contivessem essas palavras-chave obrigatoriamente. Para os materiais dentro da categoria de websites de gastronomia e turismo, utilizamos as seguintes palavras-chave: “culinária”, “imigração italiana”, “serra gaúcha”, “turismo” e “gastronomia”. Foram feitas algumas buscas a partir desse grupo de palavras-chave, trocando “culinária” por “gastronomia”, por exemplo. Os materiais selecionados estavam presentes nas primeiras cinco páginas de resultados de cada busca.

Os materiais jornalísticos também foram coletados dessa maneira. Contudo, o critério maior foi a sua presença na aba *Notícias* do Google, em que o buscador lista os conteúdos pertencentes a *websites* de jornais e revistas.

Os materiais acadêmicos foram compilados também a partir da busca por palavras-chave, porém adicionando “pdf” no final da busca, o que retornou arquivos de artigos e outros trabalhos acadêmicos. Novamente, selecionamos os documentos mais pertinentes ao nosso objetivo presentes nas primeiras cinco páginas de resultados.

O *corpus* processável, por sua vez, como já vimos, foi compilado automaticamente a partir do Sketch Engine, seguindo os passos descritos previamente. O *corpus* processável partiu de uma busca na *web*, feita pelo próprio Sketch Engine, que considerou as seguintes palavras-chave, ou *input-words*: “culinária”, “imigração italiana”, “serra gaúcha”, “gastronomia”, “história”, “cultura” e “turismo”. Utilizamos a configuração padrão do *software* para limites de tamanho dos documentos e quantidade máxima e mínima de palavras-chave por texto. O *corpus* processável possui 341.591 palavras divididas em 236 documentos. Na imagem abaixo, podemos ver algumas estatísticas do nosso *corpus*, como contagem de *tokens*, sentenças e parágrafos, além das informações já mencionadas.

Tese_culinária		user/dspader/tese_culinaria	
Corpus processável tese			
GENERAL INFO		COUNTS	
Language	Portuguese	Tokens	419,064
Tagset	<a href="#">Description</a>	words	341,591
Word sketch grammar	<a href="#">Description</a>	Sentences	19,194
		Paragraphs	7,931
		Documents	236

Descritas as características de ambos os *corpora*, passamos agora às nossas reflexões sobre o potencial da Linguística de *Corpus* na parte analítica da pesquisa. Neste artigo, focamos especificamente no *corpus* de apoio. A metodologia e a análise completas se encontram na Tese (SOUZA, 2019), cuja referência consta ao fim deste trabalho.

### **CORPUS DE APOIO: FAMILIARIZAÇÃO COM O DOMÍNIO E PROPOSIÇÃO AMPLA DE FRAMES**

Esta seção do artigo tem como objetivo discutir o campo da culinária de imigração italiana na Serra Gaúcha de forma mais ampla, de modo a apresentar uma sistematização do domínio, permitindo verificarmos a importância e o papel central do *corpus* de apoio. Além disso, como vimos anteriormente, esta parte da

análise utiliza mapas conceituais como ferramenta, o que nos permite visualizar o domínio e estabelecer as relações entre as partes.

Dos itens selecionados para compor o *corpus* de apoio, os materiais de caráter acadêmico ofereceram uma visão sócio-histórica extremamente rica e importante para a compreensão do domínio. Isso porque a culinária de imigração italiana na região da Serra Gaúcha está intrinsecamente relacionada às condições de vida da comunidade de imigrantes italianos que chegaram ao Rio Grande do Sul no século XIX. Os materiais jornalísticos e relacionados à gastronomia e ao turismo forneceram uma perspectiva mais contemporânea, focada no papel social e econômico dessa culinária nos dias atuais. Existe, nesse sentido, uma perspectiva diacrônica, de transformação da culinária da imigração italiana no que diz respeito àquilo que ela representa em diferentes momentos da sociedade. Todo esse domínio da culinária da imigração, contudo, pressupõe o contexto maior da imigração italiana em si.

Segundo Pertile e Gastal (2012), a culinária da imigração italiana na Serra Gaúcha representa aquilo que se chama de “comida de pobre”, ou “comida empobrecida”. Essa ideia surge do fato de que, à época, a alimentação dos imigrantes italianos tinha o propósito principal de “encher barriga”. Os imigrantes eram pobres, saíram da Itália com a promessa de encontrar um país rico e cheio de oportunidades. Os imigrantes que chegaram ao Rio Grande do Sul, mais especificamente na região que hoje concentra os municípios de Bento Gonçalves, Garibaldi e Caxias do Sul, encontraram áreas montanhosas, rochosas, que demandaram trabalho árduo e contínuo. A alimentação era a forma de se manter de pé e trabalhando. Nesse sentido, não é possível abordar a culinária da imigração italiana sem considerar os aspectos geográficos, sociais e econômicos que trouxeram dificuldades aos imigrantes. É por isso, também, que, neste trabalho, não se pode falar de culinária *italiana*. Não se trata da culinária da Itália: se trata da culinária da *imigração*, que retrata as condições de vida das famílias que saíram do seu país de origem e desembarcaram em uma terra completamente nova e desconhecida.

Outro aspecto importante que deve ser levado em consideração é a influência dos hábitos alimentares dos imigrantes italianos na culinária que se originou em território brasileiro. Houve, nesse sentido, uma adaptação desses hábitos alimentares impulsionada pelas necessidades impostas. Segundo Scarpo (2011), quando os imigrantes chegaram ao Rio Grande do Sul, eles se alimentavam basicamente de frutos silvestres e pinhões; foi após as primeiras colheitas que começaram a se alimentar de pratos típicos da culinária camponesa italiana, como a polenta, resgatando os hábitos que trouxeram consigo. Um ponto curioso e interessante é a presença muito forte da abóbora na culinária da imigração italiana: o solo da região da Serra Gaúcha propiciava o cultivo do fruto em larga escala, e foi isso que fez com que o alimento adentrasse os hábitos alimentares do imigrante italiano, além do fato de que a abóbora alimentava sem ser cara, uma vez que era produção do próprio imigrante. A partir disso, a fruta foi *misturada* com elementos culinários que vieram da cultura italiana, como a forte presença das carnes na alimentação. Assim, o domínio da culinária da imigração italiana

articula dois domínios maiores, que é o da herança cultural trazida pelos imigrantes da Itália e o da realidade da imigração em si.

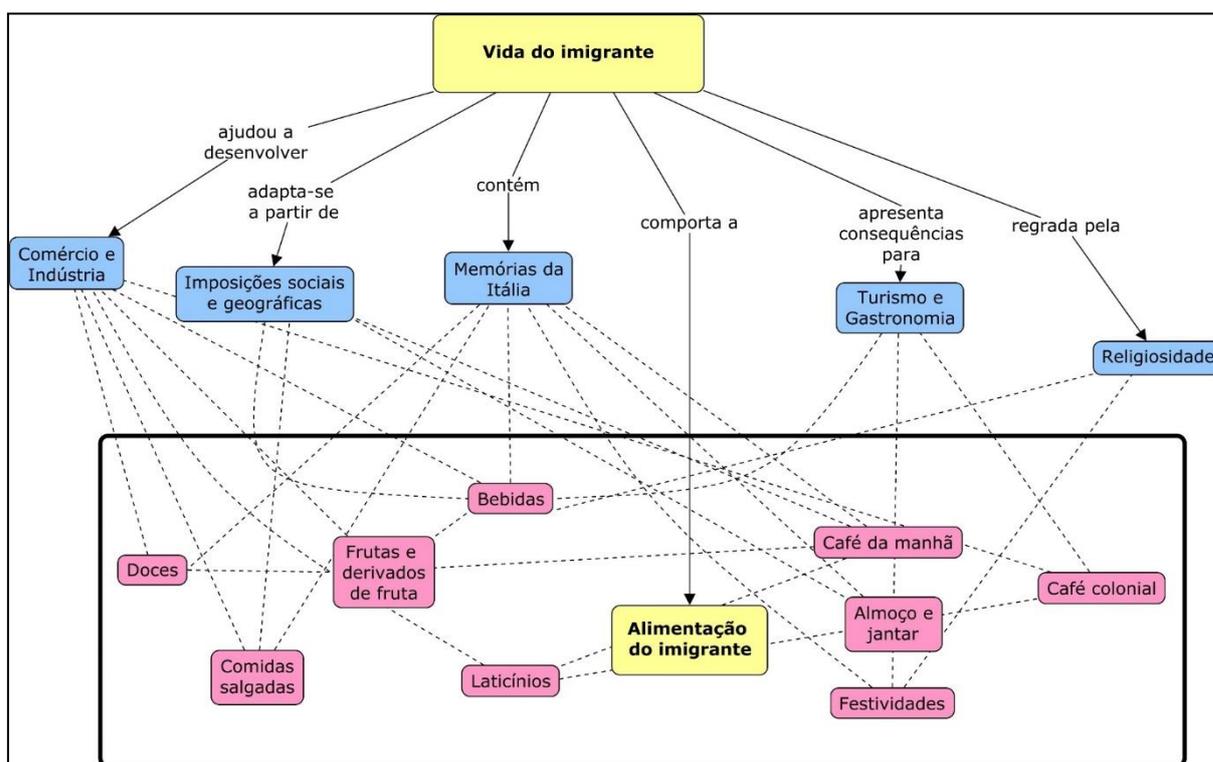
Associada à herança cultural do imigrante italiano, está a questão da religiosidade. O catolicismo foi o eixo organizador da experiência social nas colônias de imigração italiana. Para Manfroi (2007), a reconstituição, no Brasil, dos valores sociais mantidos pelos imigrantes na Itália permitiu que as comunidades de imigração não só se adaptassem melhor à nova terra, mas também superassem a saudade do país natal. Essa reconstituição de valores teve consequências para a alimentação também, na forma das grandes comemorações, fortemente baseadas na ideia de fornecer muita comida aos convidados. Essas celebrações tinham, na grande maioria das vezes, motivações religiosas, como era o caso dos casamentos e festividades católicas (PERTILE; GASTAL, 2012).

Também importa, neste trabalho, considerarmos o papel turístico e gastronômico da culinária de imigração italiana atualmente. Essa importância se deve, principalmente, ao fato de que a culinária da região da Serra Gaúcha, vendida hoje como atrativo turístico, não é a culinária de imigração italiana propriamente dita, que esteve e ainda está presente nos pratos e mesas das famílias descendentes de imigrantes. Um exemplo disso é o galeto *al primo canto*. A origem desse prato está na prática dos imigrantes italianos de caçarem pássaros para consumo. A partir do momento em que a caça dos pássaros foi dificultada, iniciou-se o hábito de abater o frango por volta de seus 25 dias de vida, época em que ele produz o primeiro canto. A questão é que comer galeto *al primo canto*, que compreende uma ave de no máximo 500 gramas, não faz parte dos hábitos alimentares das comunidades descendentes de imigrantes italianos, sendo, assim, algo mais relacionado à gastronomia turística. Na mesa dos descendentes de imigrantes estão muito presentes, da mesma forma, partes menos nobres dos animais, como patas de frango e de porco e pescoço de frango. Essas partes não são oferecidas em restaurantes típicos na região da Serra Gaúcha. Outro exemplo do apelo turístico da culinária da imigração italiana é aquilo que se chama, nas rotas gastronômicas, de café colonial. O café colonial é composto de um conjunto de elementos clássicos da culinária da Serra Gaúcha, como os pães, as cucas, o salame e o queijo colonial, as geleias de fruta de sabores variados, o vinho, o suco de uva artesanal etc. A fartura presente nos cafés coloniais é muito maior do que aquela que se apresenta na mesa das famílias de descendentes de imigrantes. A motivação do café colonial é turística, pois tem a intenção de mostrar a essência da culinária da Serra Gaúcha aos visitantes.

Nesse sentido, a partir dessas considerações, fica evidente que tratar da culinária de imigração italiana demanda considerar fatores que extrapolam a comida pura e simplesmente. São aspectos ligados à religião, à história, à geografia, à esfera social, econômica etc. Santos (2005, p. 15) salienta que “[...] comer é um ato social, pois constitui atitudes ligadas aos usos, costumes, protocolos, condutas.” Assim, é impossível que, a partir de uma perspectiva semântico-cognitiva, deixemos de lado todos esses aspectos, os quais constituem um arcabouço rico de conhecimento enciclopédico. Nossa sistematização do domínio da culinária de imigração italiana não pode negligenciar os aspectos relacionados às experiências e às condições de vida dos imigrantes italianos no Rio Grande do Sul.

Por conta disso, partindo do estudo do *corpus* de apoio, percebemos a necessidade de abarcar pelo menos cinco aspectos socioculturais ou sócio-históricos, sendo eles (i) as condições sociais, culturais e geográficas do imigrante italiano, (ii) as memórias e hábitos da terra natal, (iii) a vida religiosa, (iv) o turismo e a gastronomia e (v) o comércio e a indústria. A partir dessas cinco “esferas” é que podemos abordar a culinária em si, haja vista que todos os itens pertencentes ao quadro da culinária de imigração italiana estão relacionados a um ou mais de um dos aspectos mencionados.

Apresentamos o mapa conceitual da culinária de imigração italiana, elaborado a partir do estudo do *corpus* de apoio:



Como podemos ver pela figura, o *corpus* de apoio possibilitou a familiarização com o domínio e a compreensão dos conceitos que os compõem. O mapa conceitual inicia com **Vida do imigrante**. A partir dele, conseguimos organizar todos os outros subdomínios. Assim, a culinária da imigração italiana é parte de um grande domínio, que é o da vida dos imigrantes italianos no Rio Grande do Sul. O nome do conceito parte da ideia de que estamos lidando com experiências de pessoas, com costumes e hábitos que, sob diversas formas, influenciaram a formação da culinária característica da imigração italiana na Serra Gaúcha. **Vida do imigrante** relaciona-se com outros cinco conceitos do mapa: **Imposições sociais e geográficas**, **Memórias da Itália**, **Turismo e Gastronomia**, **Comércio e Indústria**, **Religiosidade** e **Alimentação do imigrante**. Esse último compreende

os *frames* específicos da culinária de imigração, divididos por categoria de elemento.

Todas essas reflexões se tornaram possíveis por conta da utilização de *corpora*, ainda que nesta fase da análise tenham sido através do *corpus* de apoio, que não passou pelo tratamento do maquinário da Linguística de *Corpus*. Avançamos, agora, às considerações finais.

## CONSIDERAÇÕES FINAIS

No presente artigo, foi possível verificarmos o potencial e a relevância da utilização da Linguística de *Corpus* como metodologia de pesquisa. Nesse sentido, advogamos fortemente a favor do conceito de *corpus* de apoio em trabalhos como este, que se voltam a domínios específicos. O *corpus* de apoio permite não só desenvolver uma familiaridade com o domínio estudado, mas também perceber os conceitos e relações envolvidos, o que combina muito bem com teorias de cunho cognitivista como a Semântica de *Frames*.

Além disso, salientamos que o uso de *corpus* se confirma como uma excelente alternativa metodológica para a Linguística Cognitiva, uma vez que o acesso a conjuntos estruturados de dados linguísticos naturalísticos fortalece o compromisso da Linguística Cognitiva com o uso da linguagem. Por fim, ressaltamos, também, que embora esta não seja uma pesquisa que se insere na área da Linguística Histórica, o *corpus* de apoio evidenciou aspectos sócio-históricos do domínio que não puderam ser ignorados.

## Notas

A thesaurus gives you alternatives – often more interesting and colourful ones – to the words you already know and use; these are known as *synonyms*. In some cases, it also gives you the opposite of words, which are known as *antonyms*. Using a thesaurus regularly will extend your vocabulary and help you to be more accurate and imaginative in the way that you express yourself.

<sup>2</sup> No original: “[...] lexical items, as well as word classes and grammatical functions, are conceptual categories that have to be studied and investigated with respect to their cognitive function [...]”.

<sup>3</sup> No original: “[...] the old question of the borderline between word knowledge and world knowledge. [...] the dividing line between semantics and pragmatics. [...] a methodological choice between a structure-oriented or a usage-oriented mode of investigation. Or [...] the degree of cognitive realism that a theory of word meaning should try to achieve.”

<sup>4</sup> No original: “[...] (i) the internal structure of lexical categories (prototype structure, family resemblance structure, radial network structure); (ii) the polysemous nature of lexical items and the cognitive principles (e.g., metaphor, metonymy, image-schema transformations) motivating the relations between the different senses of lexical items; (iii) larger conceptual structures (e.g., metaphor research, frame semantics).”

<sup>5</sup> No original: "One motivation for this label is the claim that usage events are the source of all linguistic units." (LANGACKER, 1987, p. 220).

---

## Referências

---

BABINI, M. Do conceito à palavra: os dicionários onomasiológicos. **Ciência e Cultura** (SBPC), São Paulo, v. 2, p. 38-42, 2001.

\_\_\_\_\_. **Onomasiologie et dictionnaires onomasiologiques**. São José do Rio Preto: Beatriz, 2001.

BALDINGER, K. Semasiologia e onomasiologia. Tradução de Ataliba T. de Castilho. **Alfa**. São Paulo, v. 9, p. 7-36, 1966.

BERBER SARDINHA, T. **Linguística de Corpus**. São Paulo: Editora Manole, 2004.

BYBEE, J. Usage-based models in linguistics: an interview with Joan Bybee. Entrevistador: Tiago Timponi Torrent. **Revista Linguística**, v. 8, n. 1, 2012.

FILLMORE, C. J. An alternative to checklist theories of meaning. In: **Proceedings of the first annual meeting of the Berkeley Linguistics Society**. Berkeley: Berkeley Linguistics Society, 1975.

\_\_\_\_\_. Frame semantics. In: The Linguistic Society of Korea (Eds.). **Linguistics in the Morning Calm**. Seoul: Hanshin, 1982.

GRIES, S. T. Corpus linguistics and theoretical linguistics: a love-hate relationship? Not necessarily... **International Journal of Corpus Linguistics**, v. 15, n. 3, 2010.

GUIRAUD, P. **A Semântica**. Rio de Janeiro: DIFEL, 1975.

GEERAERTS, D. Representational formats in Cognitive Semantics. **Folia linguística**, v. 29, n. 1, Mouton de Gruyter, 1995.

\_\_\_\_\_. Lexicography. In: GEERAERTS, D.; CUYCKENS, H. (Eds.). **The Oxford Handbook of Cognitive Linguistics**. Oxford: Oxford University Press, 2009.

LANGACKER, R. W. **Foundations of Cognitive Grammar**, volume 1. Stanford: Stanford University Press, 1987.

KILGARRIFF, A.; Rychlý, P.; SMRZ, P.; TUGWELL, D. **The Sketch Engine**. Lorient: Euralex, 2004.

McENERY, T.; HARDIE, A. **Corpus Linguistics: Method, Theory and Practice**. Cambridge: Cambridge University Press, 2012.

OXFORD School Thesaurus. Oxford: Oxford University Press, 2016.

PETRUCK, M. R. L. **Frame semantics**. Berkeley: University of California, 1996.

PERTILE, K.; GASTAL, S. **Turismo e Gastronomia: as vozes italianas e a culinária de imigração**. Anais do VII Seminário de Pesquisa em Turismo do Mercosul, 2012.

ROSCH, E. **Recuperando os conceitos**. Tradução de Dalby Dienstbach. In: SIQUEIRA, M.; OLIVEIRA, A. F. S. (Orgs.). **Cadernos de Tradução**. n. 31, 2012.

SCARPO, P.S. **Imigrantes Italianos na Serra do Nordeste do Rio Grande do Sul**. Porto Alegre: EST Edições, 2011.

SOUZA, D. S. **Entre conceitos e conce(p)tos**: uma proposta teórico-metodológica na interface entre a onomasiologia, a Lexicografia e a Semântica Lexical Cognitiva. 2019. 209 f. Tese de Doutorado (Doutorado em Linguística Aplicada). São Leopoldo: Universidade do Vale do Rio dos Sinos, 2019.

ULLMANN, S. **Semântica**: uma introdução à ciência do significado, Tradução de J. A. Osório Mateus, Lisboa: Fundação Calouste Gulbenkian, 1964.

---

### Para citar este artigo

---

SOUZA, D. S.. Contribuições da Linguística de Corpus para uma pesquisa situada na interface entre Semântica Lexical Cognitiva e Lexicografia Onomasiológica. **MACABÉA – REVISTA ELETRÔNICA DO NETLLI**, CRATO, V. 8., N. 2., 2019, p. 194-212.

---

### O Autor

---

**Diego Spader de Souza** é Doutor (CAPES/PROSUP) em Linguística Aplicada pela Universidade do Vale do Rio dos Sinos - UNISINOS (São Leopoldo/RS). Realizou período de Doutorado-Sanduíche (PDSE/CAPES) na Universidade do Texas em Austin, sob orientação do Prof. Ph.D. Hans C. Boas. É também membro do grupo de pesquisa SemanTec - Semântica e Tecnologia, atuando no projeto "Convergências entre Semântica de Frames e Lexicografia Computacional". Interesses de pesquisa e estudo incluem: Semântica Lexical, Semântica Cognitiva (com ênfase em Semântica de Frames) e Lexicografia Computacional.