

O léxico da fermentação latina: escolhas e padrões do corpus fermentopedia

The Latin fermentation lexicon:
choices and standards in the corpus fermentopedia

Wildon César Rodrigues dos Santos

Universidade Federal de Minas Gerais

Igor Francisco Chagas dos Santos

Universidade Federal de Minas Gerais

Resumo: Neste trabalho, objetiva-se apresentar a construção e as análises do *Corpus* de Estudo Fermentopedia, composto por 7.030 textos sobre 157 alimentos fermentados latino-americanos. Após o tratamento do *corpus*, foram realizados os levantamentos estatísticos de termos e de nomes de alimentos fermentados explorados pela literatura científico-tecnológica da Fermentação, dos quais foram analisados os agrupamentos semântico-lexicais, observando-se os aspectos probabilísticos e categóricos. Também, utilizando-se de dois casos terminológicos, foram discutidas as possibilidades dos estudos terminológicos-cognitivos aplicados aos textos sobre a alimentação e fermentação.

Palavras-chave: Linguística de corpus; Linguística cognitiva; Terminologia; Fermentação

Abstract: In this work, the objective is to present the construction and analyzes of the Fermentopedia Study Corpus, composed of 7,030 texts on 157 Latin American fermented foods. After the treatment of the corpus, were carried out statistical surveys of terms and names of fermented foods explored by the Fermentation scientific-technological literature, from which were analyzed semantic-lexical groupings, observing the probabilistic and categorical aspects. Also, using two terminological cases, were discussed the possibilities of terminological-cognitive studies applied to texts on food and fermentation.

Keywords: Corpus linguistics; Cognitive linguistics; Terminology; Fermentation

Este trabalho levanta questões sobre a fermentação, em um projeto interdisciplinar denominado Fermentopedia, em desenvolvimento, que parte do interesse de organizar informações, histórias e conhecimentos tradicionais e científicos sobre os alimentos fermentados presentes nas culturas e sistemas alimentares. Atualmente, o projeto se encontra na Fase Latina, que investiga os nomes de alimentos fermentados e suas particularidades nos léxicos da alimentação na América Latina.

Inicialmente, devemos considerar que os alimentos fermentados são constituídos a partir de diversos tipos de substratos (alimentos iniciais), que são nativos, dispersos e/ou distribuídos em um espaço geográfico, ou seja, são elementos significativos de sistemas humanos complexos, que se organizam em uma cultura alimentar. Entende-se que a cultura alimentar, partindo-se de dimensões sociais, culturais e suas linguagens, refere-se aos mecanismos lógicos e abstratos em que são interpretados e representados os conjuntos de ferramentas, tecnologias, práticas, espaços, experiências e ensinamentos do comer, em que o alimento é o mediador. Quando colocada em um quadro esquemático, a cultura alimentar pode ser conceituada e elaborada pela perspectiva de sistemas, o sistema alimentar, no qual o alimento é relacionado e entendido a partir da sequência de controles sociais e individuais, isto é, são operações e mecanismos em que o alimento é produzido, consumido, distribuído, sentido e representado (CONTRERAS e GARCIA, 2015).

Nesse ponto, evoca-se as noções sobre a alimentação, das quais podem ser observadas sob diversas perspectivas, tais como: econômica, política, biológica, antropológica, artística, dentre outras, que, independentemente da orientação seguida, o objeto alimentar pode ser entendido como uma unidade carreadora de cultura e de natureza, um híbrido. Nesse trabalho, a alimentação comporta-se como um receptáculo, que sustenta significados, funções, sensações, histórias e compreensões sobre o alimento e suas relações com o sujeito que o come, o Comedor¹. Também, exige um olhar interdisciplinar orientado por processos metodológicos que envolvem a representação, os usos, os significados, a organização das informações, desde o nível lexical ao cultural, de um alimento e dos estados conceituais e sensoriais que um alimento apresenta em um sistema alimentar, assim como, nas interdependências com outros sistemas.

Em observação sincrônica, os nomes de alimentos, assim como de outros objetos linguísticos que circundam a alimentação, estão associados aos léxicos especializados (terminologias) das Ciências Biológicas e subáreas, visto que, pelo processo de categorização, carregam as particularidades dessas áreas e disciplinas, que modificam, intensificam, ampliam ou suprimem outros conhecimentos ligados aos nomes de alimentos. Especificamente, pensando nos registros dos nomes de alimentos fermentados nos léxicos latino-americanos, observa-se que há uma abundância documental (textual, oral e narrativo-histórica) em que prevalece um *sentido tecnológico* sobre os alimentos e a fermentação, principalmente produzido pelas Ciências Biológicas, isto é, os significados que os alimentos podem assumir em uma cultura alimentar têm a tendência de

¹ O comer constitui-se como um conceito derivado dos estudos do comensalismo, em que a alimentação é pensada na perspectiva do sujeito que come, o Comedor, inserindo-se noções sobre simbólico, representação e identidade, que atravessam a memória, tradição e inovação.

estar ancorados à ideia de tecnologia. Por outro lado, existe uma raridade documental que aborda os alimentos com elementos da linguagem, tratados como unidades carreadoras de significados que modificam a experiência humana com a natureza. Assim, surge o interesse de pensar criticamente a alimentação, questionando-se os ruídos documentais sobre a fermentação e os alimentos fermentados como objetos diretos dos estudos da linguagem e, também, o porquê do sentido tecnológico ser ressaltado e dedicado.

Para a compreensão dos mecanismos que levam à raridade documental sobre a fermentação como um processo da linguagem, por hipótese, podemos partir da reflexão que os alimentos fermentados parecem permanecer em estado de continuação, um *continuum alimentar*, uma ideia próxima ao que Giddens (2012), diz sobre a coexistência de tradições, que, por meios de mecanismos interacionais e multifacetados da alimentação, resistem a um eminente esquecimento na língua e nas culturas alimentares, no entanto, não são intocáveis, pois são itens em modificação sensorial, social e cultural. Tratam-se de partículas instáveis, em que seus nomes estão em estado de tensão designativa e conceitual na linguagem, comportando-se como metáforas, pois seus significados podem ser expandidos, isso, porque “o significado não está linguagem; ao contrário, a linguagem é como um recipiente para construção do significado” (CANÇADO, 2008, p.107 *apud* FAUCONIER, 1985,1994).

Para compreensão dos mecanismos que levam à abundância documental, observa-se que há uma parcela discreta de alimentos fermentados que são exaustivamente explorados e que foram escolhidos pelos interesses da Ciência, em geral, de caráter econômico, político e religioso, que regem à univocidade (ANDRADE, 2010. p. 116-120). Esses interesses de exploração de um número restrito de alimentos fermentados pela Ciência, aqui, referindo-se às Ciências Biológicas e subáreas, foi o orientador desse trabalho. Contudo, ainda não são evidentes as motivações e implicações que selecionam um alimento fermentado como item privilegiado pela exploração e divulgação das Ciências e que, em algum grau, legitima a existência documental desses alimentos. Em contrapartida, grande parte dos alimentos fermentados latino-americanos são inventariados na linguagem cotidiana, que é extremamente complexa.

Nessa investigação, considerando-se os aspectos sócio discursivos, lexicais e semânticos, focaremos a análise na produção textual (escrita) abundante sobre a fermentação, definido como a literatura científico-tecnológica da fermentação, que é predominantemente publicada pelas Ciências Biológicas e subáreas. Contudo, ressalta-se que essas produções são resultados de escolhas semântico-lexicais de palavras estabelecidas em linguagem comum, que se transpõe (metaforicamente) a uma linguagem especializada ou temática, e vice-versa, dessa forma, esse trabalho parte dos postulados da Terminologia Textual, especificamente da Teoria Sociocognitiva da Terminologia, em que termos são entendidos como elementos da linguagem natural, assim, revisitando e analisando esses textos especializados, podemos entender a organização do pensamento sobre a fermentação e, possivelmente, entender o porquê de alguns alimentos fermentados permanecerem à margem dos interesses da Ciência.

O termo e as posições teóricas

Nesse trabalho, destaca-se a abordagem probabilística de análise de um *corpus* de estudo composto por textos especializados, ou seja, o uso de análises estatísticas descritivas, analíticas e multivariadas para observar as condições em normalidade e não-normalidade das escolhas de unidades linguísticas. Aqui, o termo é descrito a partir de frequências, principais escolhas e agrupamentos semântico-lexicais, no entanto, tratam-se de observações lógico-matemáticas que se referem aos comportamentos dos componentes lexicais, para que, posteriormente, possam ser geridos e discutidos em uma posição teórica.

Partindo-se dos pressupostos da Teoria Geral da Terminologia de Wüster, os termos são objetivos, precisos e exibem traços necessários que os distinguem de outros itens normalizados em um léxico específico, ou seja, são particulares e diferem-se do léxico cotidiano. Tais pressupostos foram discutidos pela Teoria Comunicativa da Terminologia, no entanto, foi superado o sentido de univocidade e intocabilidade dos termos, que passam a ser entendidos como parte do léxico comum e com função comunicativa, em que há polissemia e similaridade, “o que significa que os termos não são entidades autônomas que formam um léxico especializado diferenciado, mas unidades denominativo-conceptuais que têm capacidade de referência, podendo exercer funções distintas” e ainda “têm um caráter poliédrico, desempenhando uma função cognitiva, uma função social ou comunicativa, como também uma função linguística” (LARA, 2005 p.5). Também, essa orientação considera os aspectos textuais e discursivos, ou seja, que os componentes lexicais são resultantes dos aspectos linguísticos e extralinguísticos, que se referem aos fatores sociais, históricos, discursivos, geográficos, dentre outros, expressos na língua. Agora, aplicando-se a uma análise de *corpus*, poderiam ser utilizadas para traçar situações terminológicas no contexto especializado e com função comunicativa, que indicariam cenários de como os cientistas ou especialistas pensam um determinado objeto, aqui, a fermentação e os alimentos fermentados, considerando que há variação conceitual e denominativa. Embora a Teoria Comunicativa da Terminologia assumisse outras funções do termo e do texto especializado, outras correntes emergiram e emergem com os avanços das pesquisas terminológicas, pela problematização da linguagem e sua inaptidão por definições e arquiteturas delimitadas, o que é de interesse desse trabalho, pois pensando a alimentação e a fermentação como linguagem, ambas não se satisfazem com poucas definições, parâmetros e funções.

Também, destaca-se a posição terminológica-cognitiva, em que se observa a confluência de modelos, postulados e proposições metodológicas para o entendimento das (inter/intra) relações de unidades linguísticas, que se embebem da dimensão social, comunicacional, discursiva, cultural e histórica da linguagem e da própria mente humana. Assim, a Teoria Sociocognitiva da Terminologia (doravante, TST) torna-se um ambiente relevante para pensar o termo centralizado, descentralizado e/ou padronizado, que, inicialmente, privilegia a compressão do termo na descrição linguística, seja pela sua singularidade, persistência, frequência, variação e/ou escolha, permitindo-se dizer que o termo é uma consequência e/ou uma causalidade, ou ainda, uma provocação ou um conteúdo mental.

A TST parte das contribuições da Semântica Cognitiva, em que “o mundo é melhor compreendido no marco dos modelos cognitivos nos quais se relacionam diferentes unidades de compreensão ou entendimento estruturadas de forma prototípica” (LARA, 2005) e ainda “admite a sinonímia e a polissemia no processo de compreensão e de comunicação justificadas pela funcionalidade, pela flexibilidade (os modelos cognitivos se alteram constantemente) e pela diversidade dos processos de categorização” (LARA, 2005 apud TEMMERMANN, 2001), em que os itens são unidades em constante alteração, maleáveis e carreadores de características não limítrofes. Analisando um léxico especializado, as unidades são parte do que é comum na linguagem e que podem ser resumidas com unidades de experiência da linguagem: cotidianas, mutáveis e que apresentam diversas finalidades não restritivas.

Essa discussão leva à crítica ao conhecimento condicionador do pensamento sobre a fermentação, que prescreve os sentidos que um alimento pode apresentar em sistema, seja alimentar, linguístico, econômico etc. Com o foco no alimento (fermentado ou não) e na alimentação percebida como linguagem, o signo alimentar permite que o alimento atravesse a cultura, o social, a mente e o sentido de restrição imposto pelas Ciências, sem priorizar apenas um dos aspectos coerentes do seu entendimento. Esse trabalho anuncia o debate sobre a fermentação como a unidade da linguagem que provocava a instabilidade do isolamento, da restrição e da especialização dos sentidos, assim, indicando caminhos para o diálogo interdisciplinar, para isso, inicialmente, busca-se entender os padrões, as descrições e os pensamentos das áreas que predominam sobre a produção de conhecimentos sobre a fermentação. Assim, esse trabalho se trata de uma investigação terminológica de textos produzidos pela grande área da Ciências Biológicas sobre a fermentação, objetivando-se apresentar o *Corpus* de Estudo Fermentopedia (doravante, CEF) e descrever os comportamentos de termos selecionados e agrupamentos semânticos presentes em textos da literatura científico-tecnológica da fermentação.

Corpus de estudo Fermentopedia

O *corpus* de Estudo Fermentopedia possui uma abordagem irnpessionística, segundo os critérios de Sardinha (2004, p. 24 apud LEECH, 1991, p.8-29) caracterizado como de taxa usual, de modo escrito, de tempo sincrônico, de seleção de amostragem, de conteúdo especializado e de finalidade de estudo. Na construção do CEF, foram obtidas fontes em língua portuguesa, inglesa e espanhola, por meio de explorações e recuperações documentais com método livre e sistemático, por exaustão de buscas, de 157 nomes de alimentos fermentados latino-americanos que estão descritos na literatura da Fermentação do tipo científico-tecnológico (Ciências Biológicas e suas subáreas, como Nutrição, Ciência de Alimentos, Microbiologia e Biotecnologia), em duas bases dados que apresentam essa literatura como escopo: a *Web of Science* (WOW) e a *Cab Direct* (CAB), contemplando trabalhos do gênero acadêmico-científico publicados entre 2010 a 2020, exceto em casos de raridade documental, em que não se aplica o critério de período de publicação. Ao todo, foram coletados 3.823 publicações da base *Cab Direct* e 3.207 publicações da *Web of Science*

relativos aos nomes, que resultou em 157 *corpora*, utilizando-se apenas os resumos e títulos em língua inglesa dos trabalhos. Esse critério foi aplicado devido esses textos apresentarem a síntese de um texto principal, ou seja, indicam as informações essenciais dos trabalhos que descrevem ou discutem algum dado sobre os alimentos fermentados e a fermentação. Os *corpora* foram tratados, excluindo-se interferentes da análise, como algumas palavras do relatório (resumo, autor, data de publicação etc.), e salvos em formato de arquivo de texto (.txt). Para a análise, agrupou-se cada arquivo em um único, em que cada *subcorpus* constituiu uma variável da análise, construindo-se os *corpora* Fermentopedia, que agrupados, nomeou-se *Corpus* de Estudo Fermentopedia.

Abordagem Probabilística

No CEF, foi realizado um levantamento estatístico de termos, bem como, a composição e distribuição probabilística das unidades lexicais e terminológicas, tais como formas reduzidas de nomes, adjetivos e verbos. Para análise do CEF, utilizou-se o software IRAMUTEQ versão 0.7 alpha 2, que apresenta diversos módulos de análise, aqui, foram utilizados: a) Análises Lexicográficas, que realiza a etiqueta morfossintática das formas lexicais e gera as estatísticas textuais descritivas do *corpus* em estudo e suas variáveis (*corpora*), tais como frequências e distribuição de formas ativas e suplementares, quantidade de textos, segmentos de textos e lemas, utilizando o método de classificação descrito por Reinert (1983); b) Classificação Hierárquica Descendente (doravante, CHD), principal módulo utilizado nesse trabalho, que se trata da análise de agrupamento que produz a categorização de palavras em classes semântico-lexicais, em que se observa a similaridade dos itens, e c) Análise Fatorial de Correspondência (doravante, AFC), que “permite, por meio de gráficos, visualizar a proximidade das palavras e das classes oriundas da CHD; não se trata de contar palavras, mas de relações entre palavras” (CARVALHO et al. 2020).

Estatísticas lexicográficas

Nas análises, foram encontrados valores muito próximos aos de Carvalho et al. (2020), tais como, aproveitamento de 75,66 % dos segmentos de texto durante as análises lexicográficas, o valor do qui-quadrado de Pearson acima de 3,85 do CHD, como descrevem os autores “representando uma separação entre as classes satisfatória”, e a soma dos fatores dos eixos dos gráficos da AFC de 92,5%. Conforme visualizado na Tabela 1, o CEF compreende 7.030 textos, que foram redimensionados pelo IRAMUTEQ em 230 textos e 125.368 segmentos de textos. Não foram avaliados os impactos do redimensionamento no comportamento distribucional das formas lexicais. O *corpus* apresenta a ocorrência de 4.321.624 termos e de 60.714 formas lexicais (após a lematização), das quais, 48.842 são formas ativas e 443 são formas suplementares. Além disso, apresenta 55.836 lemas e foram geradas 5 classes semântico-lexicais. Ao todo, 94.978 segmentos de textos foram classificados.

Os dez termos ativos mais frequentes foram *fermentation* “fermentação” (frequência.: 28.997/ etiqueta: substantivo), *cocoa* “cacau” (frequência.: 24.597/ etiqueta: substantivo), *food* “alimento, alimentação ou comida” (frequência: 17.515/ etiqueta: substantivo), *ferment* “fermentar” (frequência.: 14.842/ etiqueta: verbo), *acid*, “ácido” (frequência.: 14.277/ etiqueta: adjetivo), *affiliation* “afiliação” (frequência.: 14.234/ etiqueta: substantivo), *bean* “grão” (frequência.: 13.834/ etiqueta: substantivo), *production* “produção” (frequência.: 13.715/ etiqueta: substantivo), *high* “elevado” (frequência:11.725/ etiqueta: adjetivo) e *cassava* “mandioca” (frequência.: 11.613/ etiqueta: substantivo). As formas suplementares obedeceram ao Princípio de Menor Esforço (ZIPF, 1965), em que se espera a elevada ocorrência de preposições, conjunções, advérbios, pronomes e artigos, como *the* “o”, *of* “de”, *and* “e”, *in* “em” e *for* “para”.

Tabela 1 – Estatística descritiva do *corpus* de estudo Fermentopedia

ESTATÍSTICA DESCRITIVA DO <i>CORPUS</i> DE ESTUDO FERMENTOPEDIA	
UNIDADES	QUANTIDADE
<i>CORPORA</i>	157
TEXTOS (WOW)	3.207
TEXTOS (CAB)	3.823
TEXTOS REDIMENSIONADOS	230
SEGMENTOS DE TEXTOS	125.368
FORMAS	60.714
OCORRÊNCIAS	4.321.624
LEMAS	55.836
FORMAS ATIVAS	48.842
FORMAS SUPLEMENTARES	443
MÉDIA DE FORMAS POR SEGMENTO	34,42
<i>CLUSTERS</i>	5
SEGMENTOS CLASSIFICADOS	94.978

Fonte: autor, 2021

A AFC é uma análise multivariada, que se baseia na noção de *token* (GRIES, 2019, p.47), que pode ser aplicada aos estudos de *corpora* (SARDINHA, 2004, p.94), identificando o grau de complexidade e variedade de escolhas lexicais. Nessa análise, avaliou-se os termos que apresentavam a raiz $\sqrt{\text{ferment}}$ em relação as classes geradas pelo CHD, para identificar o melhor exemplar nas 157 variáveis do CEF. As formas lexicais mais utilizadas podem estabelecer uma noção de exemplar, ou seja, são prioridades dentre as escolhas lexicais possíveis e entendidas como variantes terminológicas (VIEIRA, 2006, p.31-32), sendo elas: *fermentation*, “fermentação”, e *ferment*, “fermentar ou fermento”, respectivamente foram etiquetadas morfossintaticamente pelo IRAMUTEQ como substantivo e verbo, sendo a segunda variante também etiquetada como substantivo (forma nomi-

nal). Foram encontradas outras 22 formas lexicais, que, na maioria, são resultados de derivação (prefixal, sufixal ou parassintética), composição e intensificação, ou de outras línguas como espanhola (*fermentación*), italiano (*fermentati*) ou latim (*fermentum*). Essas últimas também podem ser entendidas, especificamente, como variantes terminológicas competitivas, em que se refere aos itens lexicais de línguas diferentes (VIEIRA, 2006, p.32).

Classes

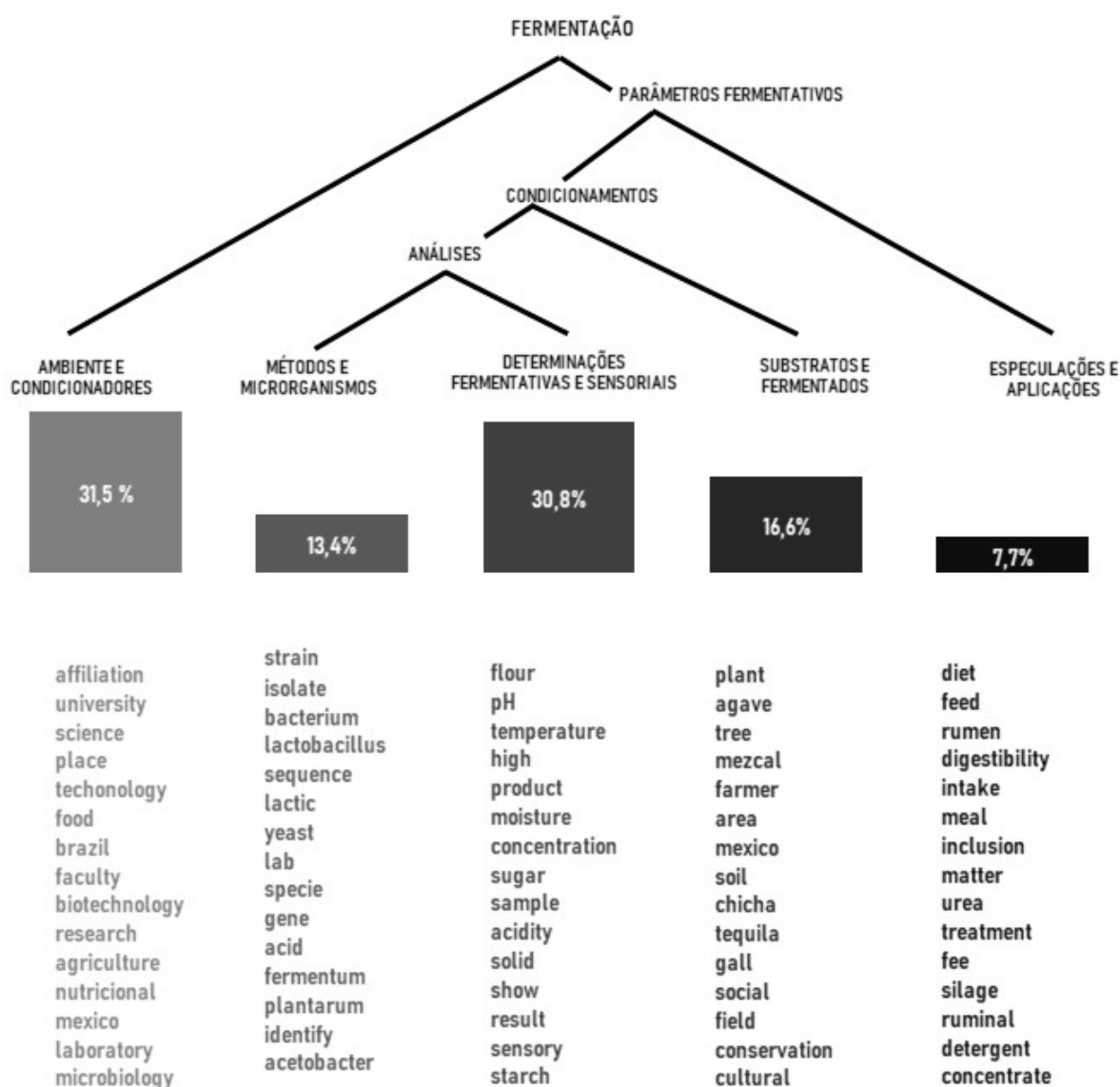
Na CHD, foram geradas cinco classes semântico-lexicais, que podem ser visualizadas no Gráfico 1. A CHD é uma análise hierárquica de cluster, isto é, um conjunto de itens inseridos em grupos que apresentam semelhanças e diferem-se de outros grupos. Segundo Gries (2019, p. 296), é “frequentemente a forma ideal para detectar padrões em conjuntos de dados grandes e aparentemente cheios de ruídos/caóticos”. Essas cinco classes apontam, principalmente, para organização das escolhas lexicais quanto a aproximação ou similaridade lexical e de sentidos, são sistemas representacionais e/ou cognitivos, dessa forma, a análise indica a relação entre formas reduzidas (lemas) e sentidos. As classes foram representadas de forma hierárquica em um dendrograma e rotuladas conforme a distribuição e os julgamentos de similaridade. Como descrição, o CEF apresenta a temática “fermentação”, o item em posição mais elevada no dendrograma, que, conforme o comportamento semântico-lexical, subdivide-se em dois grandes grupos ou aglomerados, sendo o primeiro também uma classe: 1) Ambientes e Condicionadores e 2) Parâmetros fermentativos.

Em análise, a classe Ambiente e Condicionadores apresenta um padrão distribucional mais definido e corresponde a 31,5% das ocorrências no CEF. Os três termos mais frequentes nessa classe são: *affiliation* “afiliação” (Freq.:11.192, $p < 0,0001$), *university* “universidade” (Freq.: 6.391, $p < 0,0001$) e *science* “ciência” (Freq.: 5.715, $p < 0,0001$). Ela apresenta, principalmente, os itens que personificam, caracterizam, condicionam e especificam o conhecimento sobre a fermentação nos textos especializados. O grupo Parâmetros Fermentativos representa 68,5% das ocorrências no CEF e possui um padrão distribucional difuso, sendo assim, foi subdividido em outros dois grupos, o primeiro mais geral, Condicionamentos, e o segundo mais delimitado, Especulações e Aplicações. Ainda, o grupo Condicionamentos subdivide-se em três grupos mais definidos: a) Métodos e Microrganismos; b) Determinações Fermentativas e Sensoriais; e c) Substratos e Fermentados.

A classe *Especulações e Aplicações* corresponde a 7,7% das ocorrências no CEF, sendo os três termos mais frequentes: *diet* “dieta” (Freq.: 2.392, $p < 0,0001$), *feed* “alimento” (Freq.: 1.931, $p < 0,0001$) e *rumen* “rúmen” (Freq.: 832, $p < 0,0001$). Essa classe agrupa, principalmente, os termos que designam ou cumprem função, finalidade e intenção do objeto de estudo (alimento/fermentação) nesse campo do conhecimento. A classe *Métodos e Microrganismos* corresponde a 13,4% das ocorrências no CEF, sendo os três termos mais frequentes: *strain* “cepa” (Freq.:4.430, $p < 0,0001$), *isolate* “isolado” (Freq.: 3.523, $p < 0,0001$) e *bacterium* “bactéria” (Freq.: 3.356, $p < 0,0001$).

Essa é uma classe muito interessante, pois agrupa tanto os termos relacionados à produção e constituição biológica de alimentos fermentados, quanto os termos que são prioritariamente relacionados às análises laboratoriais de alimentos fermentados, exibindo o sentido de técnica ou de objeto tecnológico.

Gráfico 1 – Dendrograma dos agrupamentos/classes semântico-lexicais gerados pela Classificação Hierárquica Descendente dos termos ativos (palavras ativas)

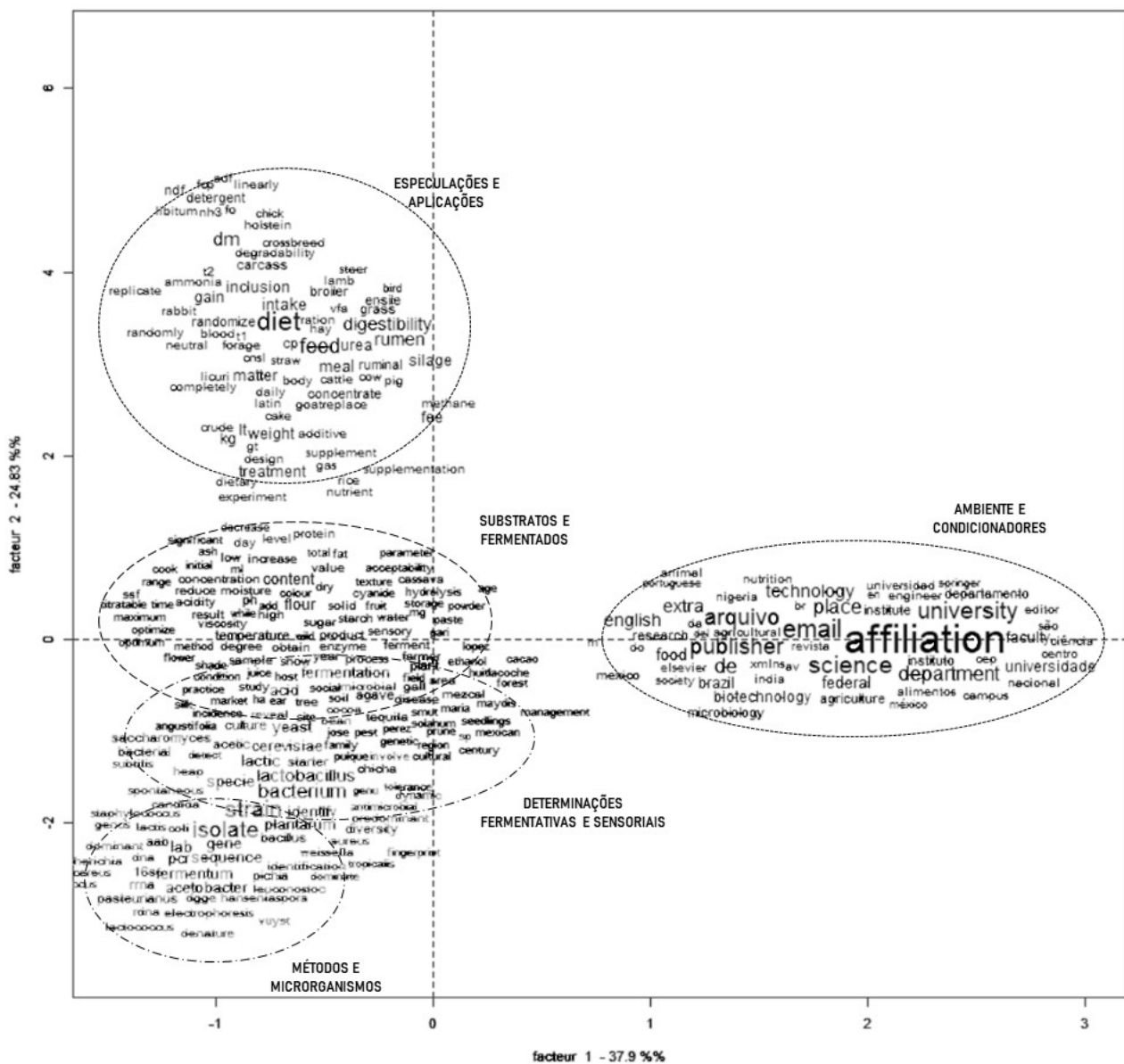


Fonte: autor, 2021

A classe *Determinações Fermentativas e Sensoriais* corresponde a 30,8% das ocorrências no CEF, sendo os três termos mais frequentes: *content* “teor” (Freq.: 5.027, $p < 0,0001$), *flour* “farinha” (Freq.: 3.202, $p < 0,0001$) e *pH* (Freq.: 3.505, $p < 0,0001$). Essa classe agrupa, principalmente, os

termos relacionado à classificação, determinação e atributos sensoriais da fermentação e dos alimentos fermentados. A classe *Substratos e Fermentados* corresponde a 16.6% das ocorrências no CEF, sendo os três termos mais frequentes: *plant* “plantar” ou “vegetal” (Freq.: 2.330, $p < 0,0001$), *agave* (Freq.: 1.510, $p < 0,0001$) e *tree* “árvore” (Freq.: 629, $p < 0,0001$). Essa classe reúne, principalmente, os nomes de alimentos fermentados e seus substratos, uso, consumo, distribuição geográfica, dentre outros. A classe possui um comportamento distribucional irregular no ponto de inter-relação de termos e ocorrência no *corpus*, no entanto, por exemplo, poderia ser utilizada para o rastreamento terminológico de nomes alimentos, pois é possível indicar e identificar outros nomes pela similaridade, frequência e aproximação.

Gráfico 2 – Análise Fatorial de Correspondência dos termos ativos e a distribuição em cada classe



Fonte: Gerado no IRAMUTEQ e alterado pelo autor, 2021

No Gráfico 2, visualiza-se a Análise Fatorial de Correspondência dos termos ativos em cada classe semântico-lexical obtida pela CHD. Observa-se na parte negativa do eixo X, a distribuição de termos que se relacionam ao grupo mais geral *Parâmetros Fermentativos*, e a oposição, na parte positiva do eixo X, da classe mais específica *Ambientes e Condicionadores*, em que se visualiza uma clara distinção com as outras classes. O eixo Y indica o padrão de similaridade dos itens que já apresentam um maior grau de similitude, ou seja, agrupados em uma classe, que quanto mais distantes na orientação do eixo sugerem menos semelhanças.

O destaque por tamanho de fonte representa os termos mais frequentes em cada classe e, como comentário, podem ser entendidos como metáforas conceptuais centrais ou (em noção de prototipicidade) itens centrais, isto é, representativos em cada um dos conjuntos. Com isso, há uma uniformidade em questão de ocorrências de itens nas classes *Substratos e Fermentados* e *Determinações Fermentativas e Sensoriais*, o que indica que são termos muito mais discretos do que os observados nas outras classes, contudo, não significa que são pouco representativos.

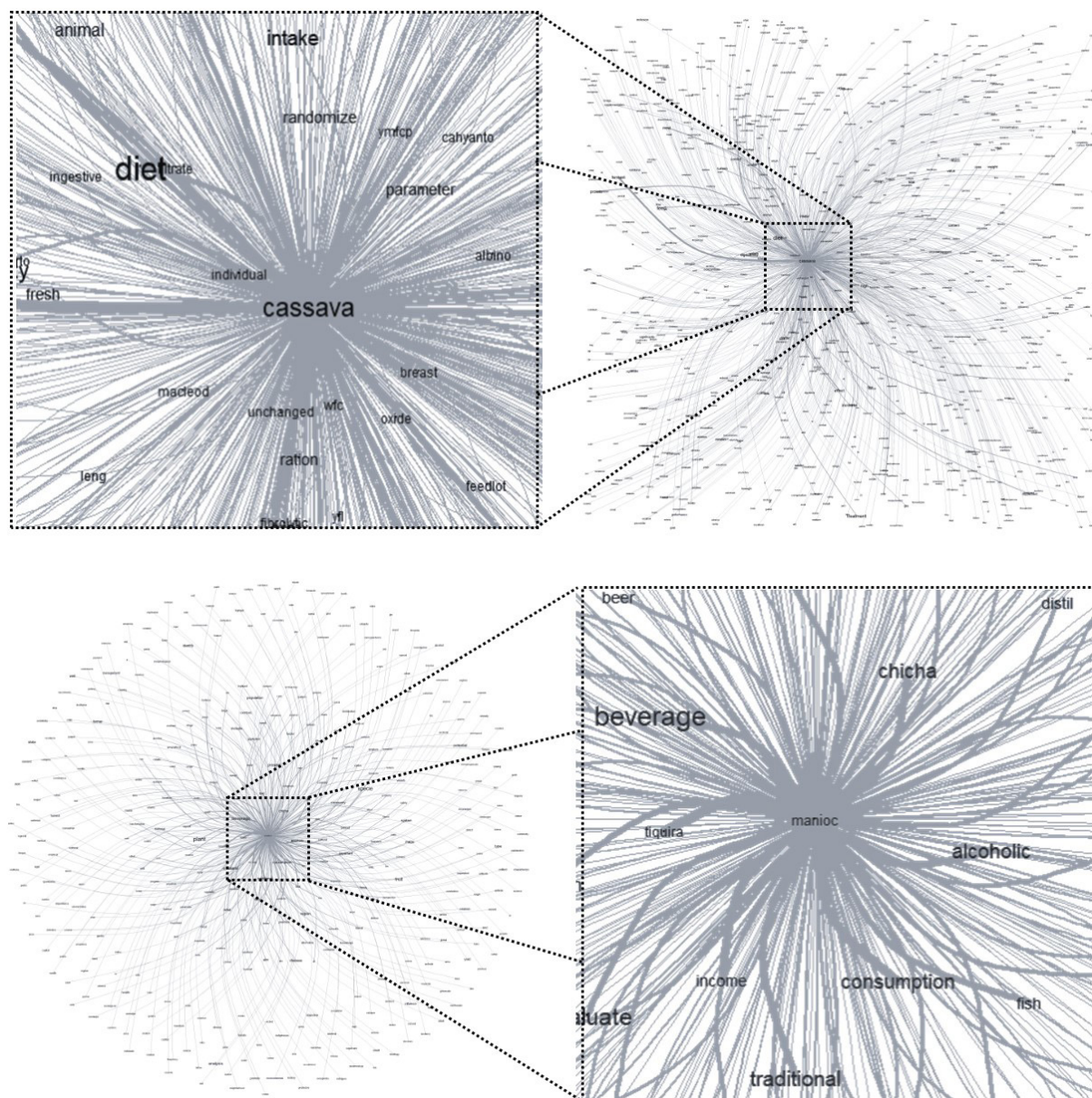
Na seção anterior, observou-se a especificidade e AFC dos termos que apresentavam a raiz *√ferment-*, como recorte, com os resultados da CHD, analisou-se quais classes contribuem para cada uma das variantes terminológicas. Exemplificando com o termo *fermentation* “fermentação”, item de primeira escolha, percebendo-se a predominância no agrupamento *Métodos e Microrganismos*, assim, comparando com as características dessa classe, pode-se estabelecer que o termo é associado a outras formas lexicais que conferem o sentido de técnica ou aparato tecnológico, sendo prioritariamente relacionada à análise microbiana. Em contrapartida, o termo parece se opor as escolhas que são agrupadas nas classes *Determinações Fermentativas e Sensoriais* e *Ambiente e Condicionares*, ou seja, nesses textos, a escolha do termo *fermentation* é mais significativo para designar um processo ou um artefato técnico-metodológico. Esse tipo de análise, comparando-se AFC e CHD, pode ser aplicada aos demais termos presentes no CEF, assim, permite entender os relacionamentos lexicais e semânticos dos termos, podendo ser utilizados para confecção de modelos cognitivos, culturais e linguísticos sobre os alimentos nos textos especializados, assim como de outros objetos de interesse linguístico e cultural.

***Cassava e manioc* “mandioca”**

Destaca-se os comportamentos dos termos *cassava* e *manioc* no CEF (Gráfico 3), que se referem ao substrato mandioca. Analisando suas distribuições gráficas, referente à análise de similaridade, esperava-se que esses termos possuíssem semelhanças e interações equivalentes por se tratarem de sinônimos, isso, considerando que ambos poderiam ser substituídos no contexto especializado (CANÇADO, 2008, p. 42), contudo, a análise revela que *cassava* se relaciona preferencialmente com itens lexicais agrupados nas classes *Especulações e Aplicações* e *Métodos e Microrganismos*, prioritariamente, exibindo o sentido tecnológico, enquanto que *manioc* se relaciona com os itens agrupados na classe *Substratos e Fermentados*, além disso, o termo é mais discreto em questão de ocorrência no CEF. Pela observação gráfica é impossível estabelecer

equivalência semântica entre esses dois termos, pois parecem permitir outros sentidos e possibilidades metafóricas.

Gráfico 3 – Representação gráfica da análise de similitude dos termos *cassava* e *manioc*



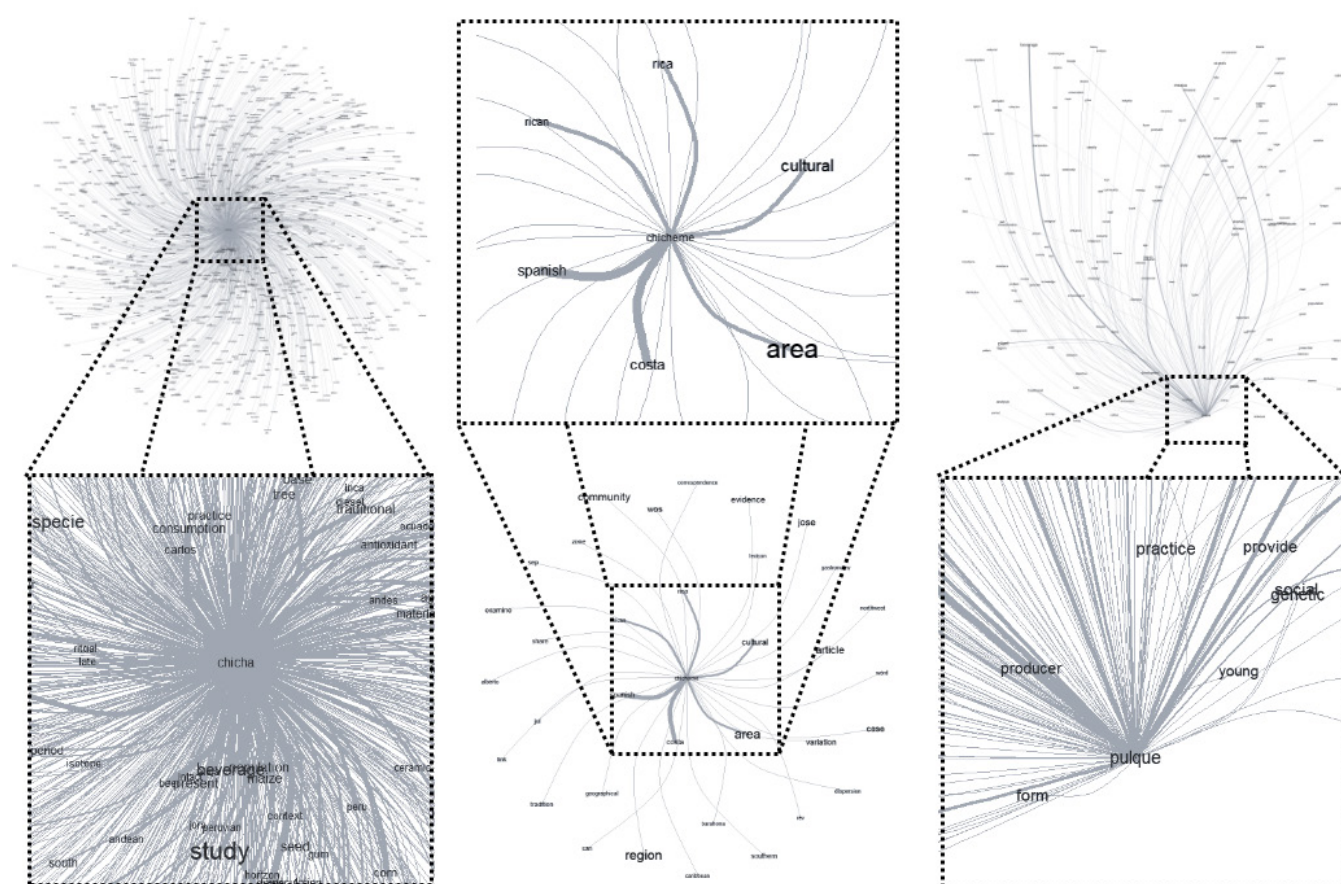
Fonte: gerado no IRAMUTEQ e compilado pelo autor, 2021

Chicha, chicheme e pulque

No CEF, observa-se que quando nomes de alimentos fermentados documentalmente pouco explorados são evidenciados pela análise gráfica (Gráfico 4), como *chicha*, *chicheme* e *pulque*, eles possuem comportamentos distribucionais muito semelhantes entre si, que, em geral, compreendem as relações com itens que designam tipos de substratos (como a *manioc*), localização geográfica

e sociocultural, tipo de conhecimento (como o tradicional), generalização (bebida), consumo, prática etc., isto é, são nomes carregados de atributos culturais e sensoriais organizados no texto especializado e que cumprem uma função comunicativa e terminológica.

Gráfico 4 – Representação gráfica da análise de similitude dos termos *chicha*, *chicheme* e *pulque*



Fonte: compilação do autor, 2021

Quando comparado aos nomes mais explorados, como a *tapioca*, percebe-se que ainda continuam esses tipos relacionamentos (padrões), no entanto, são sufocados pela especificação tecnológica, característica do gênero científico-tecnológico, conferindo distribuições gráficas irregulares aos itens de menor frequência e exploração. Dessa forma, essa análise é eficaz para indicar a complexidade de relações, das escolhas dos termos e como os alimentos fermentados estão organizados nos textos especializados, conforme a orientação teórica, também, indica caminhos interpretativos sobre esses alimentos na linguagem cotidiana. Além disso, demonstra ser um mecanismo muito preciso e estável para a rastreabilidade de nomes, bem como, de outros termos, entretanto, exige análises complementares para se evitar julgamentos que estabelecem arbitrariedade.

Provocação alimentar

Para o CEF, o ideal é realizar uma análise de *cluster* não hierárquica devido ao tamanho amostral, o que permitiria a geração de inúmeros agrupamentos pela flexibilidade de reagrupar os elementos durante as análises, usualmente, para observar com poucas restrições e profundamente as semelhanças e diferenças entre as unidades e classes, isso, utilizando-se de agrupamentos iterativos (*k*-médias) e difusos (*Fuzzy c*-médias). No entanto, o método hierárquico foi, inicialmente, utilizado para pensar o léxico especializado da fermentação, a restrição e como determinadas áreas e campos do conhecimento científico-tecnológico condicionam os sentidos dos alimentos fermentados.

Anteriormente apresentado, alguns poucos alimentos fermentados são objetos típicos do conhecimento científico-tecnológico, sendo incansavelmente descritos, caracterizados e sufocados pelo sentido tecnológico, que é corroborado pela análise do CEF. Já a maioria dos alimentos fermentados em observação no CEF (76% - 120 alimentos), parecem emitir sentidos mais próximos à diversidade, patrimônio e representação, porém não isentos do sentido tecnológico, em que a “culinária, as práticas e hábitos alimentares do local estabelecem aspectos representativos da comunidade, assim como o termo próprio daquele domínio faz referência a essas práticas e hábitos. Logo, como dito em outra ocasião, o léxico de uma língua nomeia as realidades” (RIBEIRO, 2016, p. 74). Assim, tratando-se da descrição da fermentação, por mais que o termo (nome do alimento) estabeleça o sentido de técnica, simultaneamente, também se incorporam outros significados ligados à cultura e linguagem.

Esse trabalho é parte de um projeto que coloca a fermentação em reflexividade interdisciplinar, sendo um recorte necessário para buscar modelos formais de entender o alimento e a própria fermentação como unidades/ elementos em hibridação com cultura humana e natureza, que, como suposição, provocam as incertezas das condições humanas. O léxico em foco, permite-nos entender os estados de escolhas, uso e preferência, que associados aos estudos interdisciplinares (como comensalidade, Teoria da História, Biotecnologia, Geografia Social, etc.) multiplica os sentidos sobre os alimentos.

Emergencialmente, a fermentação anuncia o pensamento em cooperação, opondo-se à especialização, ao produto tecnológico e dependência ao sentido restrito. Contudo, a dificuldade básica é iniciar esse debate, que esse trabalho pretende, mesmo instrumentalizado pelos aparatos estatísticos que são utilizados para o estabelecimento da univocidade. Em geral, os métodos semiquantitativos com tratamento estatístico são colocados em segundo plano nos estudos linguísticos, no entanto, são inúmeras as contribuições, principalmente quando aplicados aos estudos de *corpora*. Esse trabalho destaca a abordagem estatística pela possibilidade de privilegiar a unidade lexical/terminológica com precursor do debate sobre a linguagem e seus objetos, dessa forma, como representativo para questionar a realidade, o conhecimento e a mente. Também, aqui, a metáfora é o aviso quanto aos efeitos da restrição, sendo o ponto de partida para formular a necessidade de expandir os modelos de significado da fermentação, que, juntamente com as frequências e agrupamentos, indicam certas posições que um alimento pode apresentar em um conhecimento.

Considerações finais

A Alimentação é multifatorial e descentralizada na discussão epistemológica, o que vale dizer que ela permeia em diversos campos e tipos de conhecimentos e que isoladamente não conseguem compreender o dinamismo sobre o que é o comer, a comida e a fermentação. Observa-se que nas escolhas semânticas e lexicais em textos sobre Alimentação, que o sentido de especificação tecnológica prevalece, provavelmente por apresentar uma função ou uma determinação sociobiológica, ou seja, o artefato tecnológico legítimo das Ciências, que formata e cria certezas sobre a natureza. Tais considerações desconsideram o fato de hibridação do alimento, no qual independe da existência humana, no entanto, é comido e seu comedor o faz existir enquanto cultura, seja como um mecanismo de sobrevivência, seja como uma unidade da linguagem. O alimento e alimentação devem ser tratados e pensados como equivalentes à natureza e cultura, dessa forma, faz-se necessário entender como um campo do conhecimento condicionante descreve e caracteriza a fermentação e os alimentos fermentados, assim, a Linguística de *corpus*, a Linguística Cognitiva e a Terminologia Textual emergem como potências para perceber os padrões e a organização sobre o pensamento fermentativo, devido seu arsenal probabilístico, quantificável, qualificável e interdisciplinar.

Dessa forma, propusera-se nesse trabalho, admitindo-se a função instrumental, em sentido metodológico, da Linguística de *corpus*, observar-se o cenário de ocorrências de escolhas lexicais e normalidades semântico-lexicais, por meio de detecção de padrões e similaridades de itens (unidades linguísticas) no texto especializado sobre a fermentação e os alimentos fermentados. Adicionando a esse sentido, apontou-se pelas contribuições da Terminologia e Linguística Cognitiva, que permitem inúmeras possibilidades interpretativas sobre a organização de termos e análise de agrupamentos. Com isso, foi apresentado o software IRAMUTEQ, que apresenta módulos de análises de textos que enfatizam os aspectos distribucionais dos componentes lexicais de um *corpus* de estudo, assim como, organiza e propõe agrupamentos semântico-lexicais, por meio da similaridade e distribuição. Finalmente, foram apresentados os resultados de cada módulo utilizado, apontando as características de cada agrupamento e os comportamentos de alguns itens do *corpus* de estudo em sentido de distribuição.

Referências

- ANDRADE, Adriano. A metáfora no discurso das ciências. Dissertação (Mestrado em Linguística). Programa de Pós-Graduação em Letras da Universidade Federal de Pernambuco, Recife, 2010. Disponível em: <https://repositorio.ufpe.br/handle/123456789/7650>. Acesso em: 15.mar.2021.
- CABRÉ, Maria. **La terminologia: teoria, metodologia, aplicaciones**. Barcelona: Ed. Antártida; Empúries, 1993.

CARVALHO, Maria.; LUZ, Madel; PRADO, Shirley. Comer, alimentar e nutrir: categorias analíticas instrumentais no campo da pesquisa científica. **Ciência & Saúde Coletiva**, 16(1):155-163, 2011. Disponível em: <https://www.scielo.org/article/csc/2011.v16n1/155-163/>. Acessado em: 15.04.2021.

CARVALHO, T. *et al.* **Utilização do software IRaMuTeQ na análise de contribuições da sociedade em processo regulatório conduzido pela Agência Nacional de Vigilância Sanitária.** Vigil. Sanit. Debate, 2020, p.10-21. Disponível em: <https://visaemdebate.incqs.fiocruz.br/index.php/visaemdebate/article/download/1429/1207/>. Acesso em: 01.mar.2021.

CANÇADO, Márcia. **Manual de Semântica:** noções básicas e exercícios. 2ª ed. revisada. Belo Horizonte: Editora UFMG, 2008.

CONTRERAS, Jesus; GRACIA, Maria. **Alimentação, sociedade e cultura.** Rio de Janeiro: Editora Fiocruz, 2015.

GIDDENS, Anthony. **A vida em uma sociedade pós-tradicional.** In: Modernização Reflexiva. GIDDENS, A; LASH, S; BECK, U. Tradução de Magda Lopes. São Paulo: Editora Unesp, 2012.

GRIES, Stefan. **Estatística com R para linguística: uma introdução prática.** MELLHO, H. (org.) e traduzido por MELLO, H.; PAIXÃO, C.; SOUZA, A.; ZARA, J. Belo Horizonte: FALE/UFMG, 2019.

FAUCONNIER, Gilles. **Mental Spaces.** Cambridge: The MIT Press, 1985.

FAUCONNIER, Gilles. **Mental Spaces:** aspects of meaning construction in natural language. Cambridge: University Press, 1994.

IRAMUTEQ. Disponível em: <http://iramuteq.org/>. Acesso em: mar. 2021.

JULLIEN, François. **Diálogo entre as culturas, do universal ao multiculturalismo.** Rio de Janeiro: Jorge Zahar; 2009.

LARA, Marilda. Uma teoria terminológica para um conceito contemporâneo de informação documentária. In: **Encontro Nacional de Pesquisa em Ciência Da Informação**, Florianópolis, SC, 2005. Disponível em: <http://repositorios.questoesemrede.uff.br/repositorios/handle/123456789/362>. Acesso em: 20.abr.2021.

LEECH, Geoffrey. Corpora and theories of linguistic performance. In: SVARTVIK, J.(org.). Directions in corpus linguistics. Proceedings of Nobel Symposium 82, vo. 65. rens in **Linguistics.** Studies and Monographs, Mouton de Gruyter, Berlin/New York, 1992.

LUZ, Madel. *Natural, racional, social: razão médica e racionalidade científica moderna*. Rio de Janeiro:Campus; 1989.

REINERT, Max. **Une méthode de classification descendante hiérarchique**: application à l'analyse lexicale par contexte. In: *Cahiers de l'analyse des données*, 8(2), 1983, p. 187-198.

RIBEIRO, Andreia. **A terminologia da culinária nos rituais em Bragança-PA**: Alimentação como Representação Simbólica e Identitária de uma Sociedade. Dissertação (Mestrado). Programa de Pós-graduação e Linguagens e Saberesna Amazônia da Universidade Federal do Pará, 2016. Disponível em: <https://www.ppls.a.prosp.ufpa.br/ARQUIVOS/dissertacoes/2014/Andréia%20da%20Silva%20Ribeiro.pdf>. Acessado em: 22. Abr. 2021.

SARDINHA, Tony. **Linguística de corpus**. Barueri: Manole, 2004.

TEMMERMAN, Rita. **Towards New Ways of Terminology Description**: The Sociocognitive Approach. Amsterdam/Philadelphia: John Benjamins Publishing Co., 2000.

TEMMERMMAN, Rita. **Sociocognitive terminology theory**. In: *Simpósio Internacional De Verano De Terminología: Terminologia y cognición*. Barcelona: IULA-UPF, 2001, p. 75-92.

ZIPF, George. **Human behaviour and the principle of least effort**; an introduction to human ecology. Facsim. of 1949, ed. New York, Hafner, 1965.