

**FRASEOLOGIA EM CORPUS ORAL E ESCRITO: A
LINGUAGEM CULINÁRIA EM REDE SOB A LENTE DA
LINGUÍSTICA DE CORPUS**

**PHRASEOLOGY IN AN ORAL AND WRITTEN CORPUS: THE LANGUAGE
OF CULINARY ARTS ON THE INTERNET UNDER CORPUS LINGUISTICS
APPROACH**

Renata Maria Odorissio¹

renata.odorissio@gmail.com

RESUMO: Nesse artigo discorremos sobre o estudo de unidades fraseológicas (UFs) a partir de um corpus formado por receitas escritas e transcrições de vídeos, representando, portanto, as modalidades escrita e oral de um domínio, a Culinária. Baseado nos princípios da Linguística de Corpus associada aos conceitos da Fraseologia, nosso estudo, de cunho comparativo, visa a analisar a complementariedade das informações veiculadas pela Internet, na forma escrita e na forma oral. Os conceitos de chavidade, de Biber (1998) e de Sinclair (1991), e de convencionalidade e regularidade, de Tagnin (2011) e de Berber Sardinha (2004), orientam os estudos empíricos, formando a base referencial para todas as operações e análises. Sobre a compilação do material analisado, selecionamos 280 vídeos para transcrição, e compilamos as receitas do website onde tais vídeos estão inseridos. Listamos as palavras-chave de cada modalidade, em seu correspondente subcorpus: CR para o subcorpus de receitas e CT para o subcorpus transcrito. As palavras-chave comuns às modalidades formaram uma lista que orientou as buscas pelas UFs e seus contextos frasais. Alcançamos 406 palavras-chave comuns, das quais três são expostas neste artigo como exemplos do trabalho empírico em desenvolvimento. A porcentagem de palavras-chave comuns aos subcorpora é de aproximadamente 75%. O recorte do objeto adotado é analisado e ainda refletimos sobre duas questões essenciais: como definir o objeto linguístico diante do panorama comunicativo da Culinária? Onde se dá o limite da linguagem de especialidade enquanto objeto, considerando que a esfera oral é integrante da comunicação em seu pleno senso? São estas as questões de fundo que norteiam a pesquisa desenvolvida ao longo do doutorado.

PALAVRAS-CHAVE: fraseologia; Linguística de Corpus; culinária; corpus oral; corpus escrito.

ABSTRACT: In this article, we are discussing about the phraseological units study from a corpus which is grounded by written recipes and video transcriptions, in so representing a written and an oral domain, the culinary arts. Based on Corpus and as well as taking into

¹ Doutoranda pelo programa de pós-graduação em Linguística da Universidade Federal de São Carlos (UFSCar). Bolsista CAPES.

account Phraseology concepts, our comparative study aims at analysing written and oral internet shared information complementarity. Biber (1998) and Sinclair (1991) keyness concept (1998), and the conventionality and regularity, by Biber (1998), by Tagnin (2011) and also by Berber Sardinha (2004) guide empiric studies, being a theoretical basis for analytical operations. Taking into consideration the analysed material compilation, we have selected 280 videos for transcription and we have compiled the website's recipes in which such videos are inserted. We have listed the key-words in each modality for each corresponding subcorpus. CR for the recipe subcorpus and CT for the transcribed subcorpus. The usual modality keywords have formed a list which guided the phraseological units searches and their phrasal contexts. We have reached 406 common keywords, from which, three were exposed in this article as an example of the empiric work development. The percentage of common keywords to both corpora is nearly 75%. The part of the object analysed made us reflect on two fundamental questions: How can we define a linguistic object into the culinary arts communicative environment? Where does the language limit speciality hinder, considering that the oral skills are integrated into communication itself? These were both questions which backgrounded this research over the years in order to complete the doctoral studies.

KEYWORDS: phraseology; Corpus Linguistics; culinary arts; oral corpus; written corpus.

INTRODUÇÃO

Esse artigo foi inspirado no projeto de pesquisa de doutoramento, cujo título é “Oral e Escrita da Linguagem de Especialidade: a Fraseologia e a Linguística de Corpus na Análise da Culinária em Rede”. A pesquisa está em fase de conclusão.

Tratamos do objetivo principal do estudo que é a realização do levantamento de unidades fraseológicas lexicais e frasais extraídas de dados empíricos reunidos num corpus de transcrições de vídeos de receitas e de textos das receitas referentes a tais vídeos. A escolha por esse recorte linguístico será justificada neste artigo, assim como elucidaremos as etapas do trabalho, desde a concepção e seleção do material para produção do corpus, até as bases epistemológicas que orientam a metodologia que envolve essa descrição terminológica em ambas as línguas.

O interesse por esse tema surgiu durante a pesquisa do Mestrado, quando nos concentramos no estudo de predicados semântico em corpus bilíngue, francês e português, da Culinária (Odorissio, 2011).

Como abordagem metodológica, o aporte operacional do corpus e os parâmetros quantitativos do estudo foram baseados na Linguística de Corpus. A definição de corpus, sua dimensão, o nível de representatividade do corpus para a linguagem em estudo e outras bases referenciais são conceitos que norteiam a pesquisa a partir das concepções de Biber (1998) e Berber Sardinha (2004).

Quanto ao conceito de Fraseologia, ao longo de séculos até os dias atuais, são diversas as acepções impressas a ele, e dentre elas enfatizamos duas preponderantes. A primeira, de viés epistemológico, é compreendida como subárea da Lexicografia (Rey; Garabatos, 2000: 119-20); e a segunda, que concebe a Fraseologia como um fenômeno da língua. Esse fenômeno se traduz pela atratividade que se manifesta mutuamente entre determinadas palavras, e que, conseqüentemente, são refratárias a outras palavras. Muitos autores usam o termo no plural – fraseologias – para se referir às unidades lexicais ou às unidades gramaticais que se formam em razão dessa atratividade.

Rey e Garabatos (2000: 119) afirmam que as diversas apelações denunciam certa imprecisão quanto às prioridades atribuídas aos tais fenômenos, objeto que já foi caro tanto à Gramática como à Lexicografia, e por esse motivo, adotamos em nosso trabalho a noção de disciplina ou de subárea das Ciências da Linguagem, assim como o faz Tagnin (2011). Fraseologia se define, doravante, como um enfoque da Lexicografia, e para os fatos linguísticos que essa disciplina observa, conceitualizamos como “unidade fraseológica” ou UF. Em nossa pesquisa, delimitamos as buscas nas UFs de natureza nominal, excluindo das análises, portanto, as UFs de natureza frasal.

O método de análise se emoldura em duas noções, a de unidade fraseológica e de chavicidade ou *keyness* (Berber Sardinha, 2004). Pelo potencial de chavicidade² que uma palavra apresenta para o domínio, comprovado quantitativamente no corpus, são selecionadas as unidades lexicais que dão sustentação às buscas por unidades fraseológicas no corpus.

Para Bevilacqua et al. (2006) e Rebecchi (2015) as unidades fraseológicas têm atraído o interesse de vários acadêmicos e mais recentemente o estudo centrado em unidades fraseológicas especializadas surge no debate entre as disciplinas das Ciências da Linguagem, em especial a Terminologia e a

² Não aprofundamos o conceito de *chavicidade* aqui por razões de limite de tempo e de espaço. Definimos o termo, brevemente, como o potencial que uma palavra adquire quando empregada em determinados conteúdos linguísticos. Esse potencial pode ser alterado, maximizado ou minimizado, segundo os critérios de elaboração do corpus, assim como pela filtragem feita nesse corpus, para a seleção entre as palavras-chave e as palavras sem relevância para o domínio analisado. Para tal operação automatizada, se faz uso de um corpus de referência, cujo conteúdo deve ser mais volumoso e mais diversificado que o do corpus de estudo. Esse procedimento consiste numa filtragem automática. Em nosso caso, consideramos que a chavicidade das palavras já é influenciada pela natureza dos documentos selecionados para a montagem do corpus. Porém, não tratamos desse debate neste artigo. Para maior explanação do conceito, indicamos a leitura do artigo de Berber Sardinha (1999), acessível em <http://www2.lael.pucsp.br/direct/DirectPapers40.pdf> Acesso em: 03 ago. 2017.

Lexicologia. Seguimos nessa direção tratando por UF a unidade fraseológica extraída do corpus, que comporta um termo-chave como núcleo e outros termos colocados que, associados a ele, resultam na estabilidade semântica dessa unidade lexical composta.

Após a coleta dos termos-chave de maior frequência ao corpus, realizamos o levantamento das unidades fraseológicas no subcorpus oral, em seguida, no subcorpus escrito, observando assim a complementaridade dessas modalidades da linguagem de especialidade. Imprimimos maior rigor à busca por unidades fraseológicas representativas da linguagem culinária com o emprego das duas vertentes epistemológicas, diferentemente do que foi realizado no estudo anterior, de predicados semânticos do francês e do português.

Os predicados semânticos são conceitos aplicados aos dados da língua geral, a fim de se obter unidades semânticas da língua geral. Partindo do que Gross (1994) realizou, na pesquisa anterior, optamos por empregar esses conceitos para uma linguagem de especialidade, com o suporte da Linguística de Corpus e dessa combinação empírica, elaboramos um método híbrido, observacional e experimental, aos moldes de Laporte (2007).

No entanto, foi possível selecionar e descrever somente predicados semânticos. Alguns eventos linguísticos característicos do gênero *receita* não puderam ser analisados por não serem objeto do modelo teórico e essas ocorrências desprezadas eram unidades lexicais presentes nas receitas com muita frequência e caracterizam a escrita culinária. Notamos, pelos dados coletados no estudo anterior (Odorissio, 2011), que os exemplos são numerosos: *sal a gosto, ponto de bala, claras em neve, forma untada, forno preaquecido, fogo médio, manteiga com consistência de pomada, pão em fatias, tomate sem pele e sem sementes, napê espesso.*

Outros estudiosos da Culinária atribuem ao texto escrito do gênero *receita* o estatuto de texto culinário por excelência. Entretanto, Santos (2014) faz uma ressalva sobre essa visão dicotômica estabelecida historicamente entre linguagem oral e linguagem escrita, importante para se refletir sobre os fenômenos da linguagem:

Biber (1998) parece-me ter chegado à conclusão de que é o gênero que provoca as diferenças (pelo menos, as diferenças que ele quis medir), e

não a dicotomia oral escrito, por isso não parece desfasada a ideia de tentar verificar, com base numa classificação por géneros, se os orais e os escritos se encontram próximos. (Santos, 2014: 01)

Embora existam, de fato, géneros preponderantemente orais e géneros escritos, Santos aponta para a insuficiência das classificações que não conseguem abarcar com precisão e amplitude os conteúdos difusos que se sobrepõem no ato comunicativo humano. Essas questões perpassam por todo o estudo, sabendo-se que essas concepções acabam por influenciar a natureza do corpus e o desenvolvimento global da observação da linguagem.

Além de questões de natureza epistemológica, a pesquisa visa contribuir para o suprimento de dados para ferramentas automáticas de processamento da língua natural (PLN). Tagnin (2011: 277), referindo-se à Fraseologia como sinónimo de convencionalidade, defende a ideia de que a língua é um sistema que se realiza em termos de probabilidade de ocorrência, mas não se restringe ao que se mostra provável: “Nem tudo que é gramatical faz parte do uso cotidiano da língua, ao passo que, mesmo não sendo gramatical, há unidades lexicais aceitas e empregadas pela comunidade linguística”.

Para a pesquisadora, as pesquisas envolvendo unidades fraseológicas especializadas ou da língua geral colaboram significativamente para o funcionamento de ferramentas computacionais que existem para observar uma grande quantidade de ocorrências de uma só vez. Tagnin (2011: 280-1) cita seis grandes corpora disponíveis na rede para consulta e desses apenas um, o Corpus do Português, é constituído de material oral. Não há garantia de que uma produção com finalidade ficcional represente de fato os eventos reais e espontâneos da comunicação em funcionamento da língua viva.

Santos (2014) expõe um gráfico do corpus Gramateca em que é mensurado o material oral compilado. A autora ressalta que não se trata exclusivamente de material transcrito, posto que apenas as entrevistas compiladas são documentos transcritos e todo o resto é composto por texto oral do género monólogo, que não pode ser considerado como modalidade genuinamente oral. Enquanto há 5.559.594 de palavras compiladas de entrevistas, por exemplo, de monólogos existem 78.543.738 *tokens*, isto é, uma quantidade 14 vezes maior.

Os dados que temos coletado evidenciaram que as características do género textual *receita culinária* como um género técnico-instrucional (Teixeira,

2008) priorizam certos aspectos da linguagem de especialidade. A recorrência de verbos na forma imperativa na terceira pessoa do singular ou o emprego do infinitivo canônico são apenas dois exemplos mais evidentes e esperados para o gênero em questão.

As “novidades linguísticas”, antes de serem incorporadas pelo registro escrito como no caso dos dicionários, se reverberam em nível oral. A lentidão com que palavras estrangeiras ou neologismos são “transferidos” para textos impressos e documentos, ao contrário da agilidade das conversas, é prova dessa tendência entre as modalidades que constituem a comunicação. Numa era em que a língua está completamente incrustada no universo da comunicação digitalizada por meio de áudios, textos escritos, vídeos e outros incrementos automáticos, a dicotomia entre as modalidades oral e escrita se torna objeto de reflexão.

São essas as implicações do estudo que, iniciado na pesquisa do Mestrado, ganha maior robustez nessa nova etapa de investigação científica.

1. UM MÉTODO OBSERVACIONAL BASEADO NOS CONCEITOS DE FRASEOLOGIA E NO EMPIRISMO DA LINGUÍSTICA DE CORPUS

Quando da pesquisa no Mestrado (Odorissio, 2011), fizemos uso do concordanciador Unitex (Paumier, 2002)³ para aferir as ocorrências relativas a 20 verbos da língua de partida, o francês, obtidos a partir de uma amostra de 505 verbos do corpus. Essas unidades lexicais cumpriram a função de palavra-chave para as buscas dos predicados representativos do domínio em questão. A título de ilustração, citamos alguns empasses que serviram de ponto de partida para a reflexão exposta neste artigo.

No corpus de receitas em francês, o verbo *poivrer*, cujo número de ocorrências é de 420 em mais de 451.762 *tokens* ao total (ou seja: 0,092%), não apresentava equivalente evidente no corpus de receitas em português. Diferentemente do verbo *frir*, por exemplo, que aparece com definição dicionarizada e casos de ocorrências muito similares ao *fritar* encontrado no

³ O processador UNITEX encontra-se disponível gratuitamente nos seguintes endereços eletrônicos: em inglês <<http://www-igm.univ-mlv.fr/~unitex/index.php?page=0>> ou em português <<http://infolingua.univ-mlv.fr/brasil/Ferramentas/Unitex.html>>. Seu manual também se encontra disponibilizado em ambos os endereços. Acesso em: 12 out. 2014.

corpus em português, o que garantia a validação da equivalência entre tais unidades lexicais.

No caso de *poivrer*, pela definição dicionarizada, esse verbo traduz o ato de adicionar pimenta num alimento ou num prato, e ele deriva do termo *poivre noire* que equivale a pimenta do reino. Pela consulta feita ao *Trésor de la langue française*, pensou-se que *apimentar* poderia ser um termo concorrente à equivalência de *poivrer* para o português. Pelas buscas no corpus de receitas em português, *apimentar* aparece em oito ocorrências apenas, encontradas em receitas de origem portuguesa (português europeu), que estavam inseridas no corpus cedido à pesquisa pelos responsáveis do corpus de Culinária, resultante do projeto COMET⁴.

Aprofundando as buscas, notamos que a frequência de *pimenta* correspondia a um número superior a 7400, ou 0,209% em números relativos. Ao analisar as ocorrências de predicados com *pimenta* no concordanciador, obtivemos primeiramente uma frequência inexpressiva de predicados, e em seguida foi possível observar que esse argumento aparecia associado a outro argumento majoritariamente. As unidades lexicais encontradas no corpus da pesquisa como, por exemplo, *sal e pimenta a gosto* e *sal e pimenta do reino*, *pimenta dedo de moça*, *pimenta biquinho*, *pimenta rosa*, *pimenta em pó*, *pimenta moída na hora* são de uso recorrente em textos culinários, conforme apontam os dados do corpus.

Essas unidades lexicais complexas são chamadas de *unidades de significação* ou *unidade de tradução* por Teixeira (2008). Para Xatara e Parreira (2011), a essa atração entre certas unidades lexicais dá-se o nome de *unidade fraseológica*, fenômeno recorrente em todas as línguas e que não precisam ocorrer necessariamente junto a um verbo. Para nossa pesquisa, unidade fraseológica é toda formação lexical composta por uma palavra-chave comum aos subcorpora e que co-ocorre com ao menos mais um elemento linguístico. Ou seja, a UF é resultado da co-ocorrência de, no mínimo, dois elementos lexicais, dos quais um deles é comum às duas modalidades representadas no corpus. As UFs são exclusivamente de natureza nominal,

⁴ Para acessar ao projeto COMET e o Corpus de Culinária, acesse <http://comet.fflch.usp.br/publicacoes>. Acesso em: 05 jun. 2017.

embora os contextos de ocorrência possam interessar no que tange aspectos definitórios das palavras-chave analisadas.

São esses parâmetros que garantem a representatividade (chavicidade) dos termos candidatos a UFs de estudo. Acreditamos que a concepção de corpus, assim como de gênero textual e recorte linguístico atuam como condicionantes para a chavicidade dos dados coletados. Também advogamos que a limitação do objeto de estudo por um gênero textual pode acarretar em resultados pouco representativos para a linguagem de especialidade, e isso se torna evidente quando tratamos da Culinária.

A linguagem culinária se mostra amplamente acessível, embora seja uma linguagem de especialidade cuja natureza é técnica e instrucional, assim como o texto culinário. Isso é constatado pela propagação de programas culinários na mídia televisiva, na grande variedade de cadernos especiais da imprensa escrita, em revistas especializadas nacionais e internacionais impressas ou digitais. Esse volume de produtos voltados ao grande público atente a uma demanda crescente que se comprova pela propagação de cursos técnicos e superiores. As escolas profissionalizantes e especializadas vão da Culinária básica à alta gastronomia, o que também movimenta o mercado editorial voltado à área didática e científica sobre o domínio. Todo esse acervo com alto potencial informativo do ponto de vista linguístico tem sido pouco explorado academicamente, à medida que é possível observar que muitos estudiosos optam por restringir seu material de pesquisa a livros impressos justificando a escolha pela padronização das receitas impressas. Não temos conhecimento sobre pesquisas que incluam conteúdo da rede em seus corpora.

1.1. CORPUS DE CULINÁRIA ORAL E ESCRITO: A LINGUAGEM PARA ALÉM DOS LIVROS DE RECEITAS

Para se produzir um corpus parte-se da definição dos parâmetros de sua compilação a fim de que as demandas e propósitos da pesquisa sejam atendidos a contento. Para nosso objetivo, o corpus deveria conter dados que permitissem observar a relação de complementaridade, como interfaces, das esferas oral e escrita da linguagem culinária.

O objetivo central é estabelecer uma comparação entre o conteúdo textual das receitas disponibilizadas em rede e a transcrição dos vídeos em que o preparo dessas mesmas receitas fosse realizado. É nesse momento que as unidades fraseológicas atuam como meio de se concretizar a análise comparativa: a fim de compreender a interação das esferas oral e escrita, se adota, como parâmetro de análise, a unidade fraseológica extraída de cada uma das modalidades.

Diante do espectro linguístico que envolve a difusão do saber culinário, a escolha das receitas e do seu preparo se justifica por se tratar de uma espécie de protótipo da comunicação em questão, enquanto outros registros sobre o universo da Culinária possuem outras funções e, portanto, outras características discursivas – e linguísticas.

Ao definir que dentro da esfera da linguagem há um gênero textual que a caracteriza, de certa forma, o que estamos adotando é um novo recorte linguístico do domínio propriamente dito.

A evidente riqueza da comunicação desse saber se evidencia na medida em que observamos que há uma grande liberdade de escrita na redação culinária sem que a essência do texto instrucional esteja ameaçada ou sem que se perca na sua essência. Acrescentando-se a esse aspecto, há o elemento histórico-temporal que imprime marcas à modalidade escrita da Culinária, como ilustramos em seguida.

Uma receita extraída do livro *Açúcar* de Gilberto Freyre (Odorissio, 2011: 31), tratando do Brasil no século XVI e da sociedade canavieira, não corresponde ao texto culinário padrão dos dias atuais, o que corrobora a ideia de que certos aspectos textuais não anulam a essência de um gênero. Nessa receita constam todos os elementos que caracterizam um texto culinário impresso: o título da receita inicia o texto, que descreve o modo de preparo juntamente com a sequência dos ingredientes necessários. Ao final há uma observação sobre o aspecto do bolo, como se dá em receitas atuais, presentes em qualquer livro impresso ou blog disponível em rede.

BOLO DE BACIA À MODA DE PERNAMBUCO

Batem-se 12 gemas de ovos com 20 colheres de açúcar; depois de tudo bem batido, põe-se 1/2 libra de manteiga, leite de 2 cocos e deita-se massa de mandioca até ficar em boa consistência de assar. A massa deve ser bem seca.

Nossa observação se justifica em razão de haver estudiosos que condicionam esses elementos do corpo do texto à definição de “texto culinário por excelência” (Teixeira, 2008).

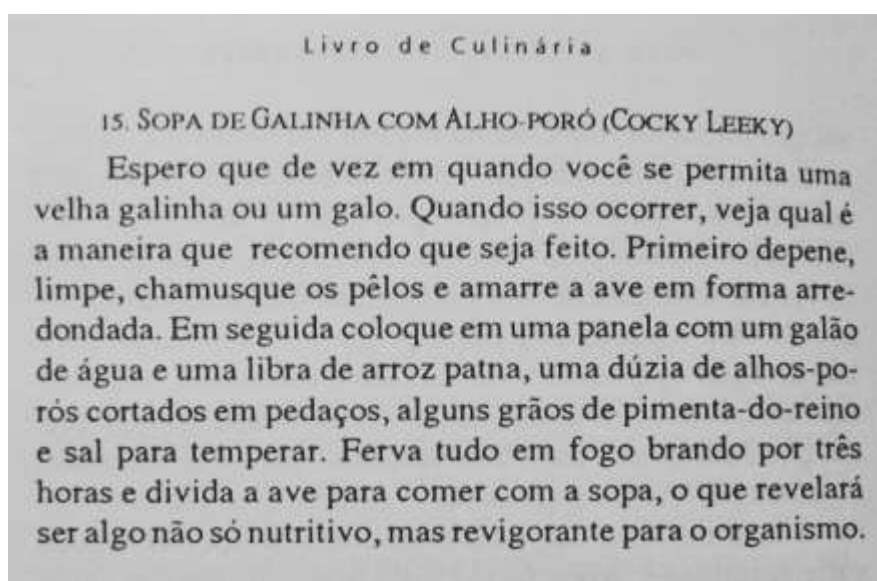
Para além desse horizonte da padronização do texto culinário, em ambiente virtual, as receitas incorporam ao texto propriamente dito, recursos icônicos dos mais variados tipos. As fotos dos alimentos ou do prato e as imagens explicativas para cada etapa do preparo são apenas alguns exemplos. São inseridas informações adicionais para adaptações ou substituições de ingredientes possíveis, menciona-se o valor calórico do prato, o rendimento, algumas vezes encontramos a proveniência da receita e aplicativos conversores de padrão de medidas. Alguns *websites* especializados incluem pequenos glossários ou referências regionais do prato descrito.

Esses recursos icônicos parecem ser um incremento tecnológico das receitas disponibilizadas pela rede, mas notamos, a bem da verdade, que esses dispositivos têm a função de incluir informações satélites a cada receita e proporcionar interatividade ao usuário do *website* ou do aplicativo. Encontramos informações sobre propriedades dos alimentos, cuidados ou dicas sobre o preparo do prato, que também existem em textos culinários impressos antigos, e que mimetizam um diálogo com o leitor acabando por revelar o dialogismo daquele tempo entre leitor e autor da receita.

Em uma obra de 1852, destinada à classe trabalhadora da época vitoriana, um cozinheiro-chefe da rainha, Charles Elme Francatelli⁵, escreve receitas com sugestões sobre como melhor escolher os cortes de carne baratos e como prepará-los para se obter o sabor e a qualidade à qual essa classe social não estava habituada em sua alimentação do dia a dia.

⁵ Essa obra foi traduzida por Renata Lúcia Bottini e teve sua primeira edição em português em 2001.

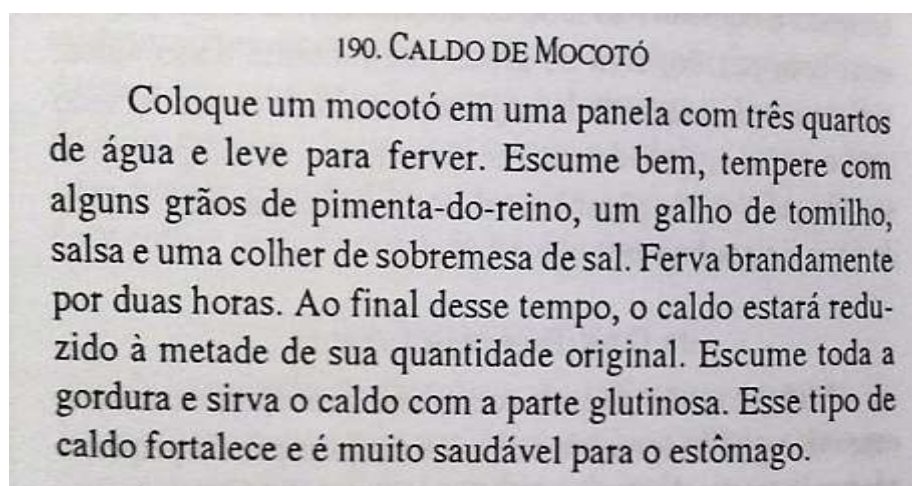
Figura 1: Receita de um cozinheiro-chef dos palácios vitorianos com sugestões.



Esse texto poderia compor nosso corpus de Culinária, pois seus aspectos interativos, de diálogo, em muito se assemelham à conversa dos chefs nos vídeos televisionados, cuja função é prender a atenção do público por meio de uma linguagem informal, desprovida de um nível de tecnicidade inacessível para quem não está familiarizado à linguagem técnica da alta Gastronomia.

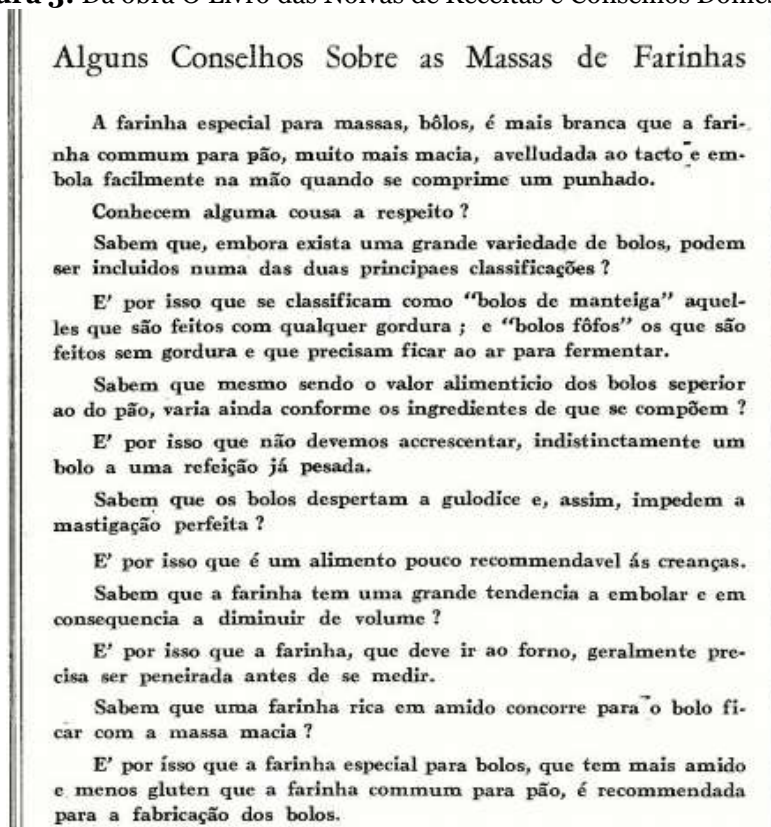
Constatamos que, com relação à padronização, não há separação por tópicos como *Ingredientes* e *Modo de Preparo*, nem marcas gráficas indicando quantidades em proporção, nem números e as sugestões estão incluídas no corpo do texto. As receitas atuais geralmente apresentam todos os tópicos separadamente.

Figura 2: Receita destinada à classe operária (FRANCATELLI; 1852, p. 98).



As observações sobre indicações do prato, sugestões sobre hábitos alimentares ou propriedades nutricionais dos alimentos também não são uma novidade da linguagem culinária atual, assim como não é recente dirigir o conteúdo desse gênero de texto para um público específico. Um livro destinado a noivas, publicado em 1929 e disponível para consulta no acervo virtual da Biblioteca Brasileira⁶, traz em seu corpo informações das mais diversas, além de receitas. Há uma tabela de conversão de medidas sobre a qual o editor faz menção por se tratar de uma obra traduzida. Essa menção esclarece que as referências originais exigiram adaptação aos parâmetros do Brasil. Num desses itens suplementares, as qualidades nutritivas dos alimentos e seus efeitos no prato são descritos com certo detalhamento, e já na época o glúten chamava a atenção dos culinários por ser prejudicial à saúde, como se verifica ao final do texto.

Figura 3: Da obra O Livro das Noivas de Receitas e Conselhos Domésticos.



⁶ Acessível para download em <http://www.brasiliana.usp.br/handle/1918/06005600#page/6/mode/1up>. Acesso em: 05 ago. 2016.

Esse texto é muito similar aos que encontramos em blogs e *websites* espalhados pela rede, além de se assemelhar ao que se ouve em vídeos de celebridades que atuam nessa área, ou mesmo de pessoas comuns que produzem conteúdo em Culinária para ser divulgado pela Internet por interesse pessoal.

A comunicação em rede tem na sua essência a interatividade e a autonomia que disponibiliza ao usuário. É uma faculdade que os outros meios de comunicação não puderam incorporar ao seu veículo, ainda que alguns meios tenham se aproximado do universo cibernético. A televisão vem sendo absorvida pelo suporte digital e, embora haja um esforço para dinamizar a interação entre o telespectador e os canais, a maleabilidade que os recursos informatizados possuem é notadamente mais efetiva nesse sentido.

Esses facilitadores da interação via comunicação virtual imprimem dinamismo ao diálogo, às trocas, às novidades linguísticas num tempo muito curto. Isso não é possível entre o leitor e o livro de receitas, posto que o diálogo é parte de um simulacro, como já expusemos sobre o excerto da obra *O Livro das Noivas de Receitas e Conselhos Domésticos e Um Simples Livro de Culinária Para Classes Trabalhadoras* (Figura 3).

Também pesquisamos alguns dos livros de receitas mais populares no Brasil, que estão associados a marcas de produtos como *União*, *Dona Benta* e *Nestlé*. Nós os consideramos populares em razão da distribuição massiva das brochuras em supermercados e bancas de jornal, por conta da veiculação dos produtos em canais diversos e pela linguagem usada nas receitas. São marcas muito consolidadas pelo tempo e pela presença nas gôndolas de supermercados.

Antigamente as receitas eram publicadas nas embalagens desses produtos, como ainda o são. A evolução do material impresso para o material digital, acessível via Internet, resultou em coleções completa, até mesmo bibliotecas, de livros culinários disponíveis nos *websites* de tais marcas.

O percurso que um produto ou marca faz para alcançar seu público depende de canais pelos quais uma linguagem de especialidade é compartilhada dentro de uma comunidade. São esses meios comunicativos que oferecem diferentes registros linguísticos. Rádio, televisão, jornais e as embalagens aderiram à linguagem culinária. Rapidamente a Internet também se tornou um meio de propagação do cozinhar e da sua linguagem, traduzindo-a em formatos

que lhe são próprios. Na sequência inserimos algumas imagens capturadas do *website* da empresa União, dando ideia do percurso tecnológico e linguístico:

Figura 4: História da Companhia União e os meios de divulgação de seus produtos.



Antes dessa revolução tecnológica, a comunicação culinária escrita dos cadernos de receitas tinha a função de perpetuar o saber dos sabores que passava de uma geração a outra. Os cadernos passavam da mãe para filha ou da avó para a neta, e assim segredos e astúcias se conservaram e se transmitiam as tradições. O livro impresso e o acesso a livrarias também modificou essa realidade “familiar”. A expansão mercadológica traz novos aspectos para a Culinária, não mais reduzidos ao convívio das famílias.

O que desejamos mostrar com essas informações em digressão temporal está expresso nas ideias de Biber (1998) e de outros linguistas que discutem a classificação de gênero textual. O gênero é concebido como um dado extralinguístico e não se resume a um padrão textual uniforme, excludente de outros formatos a fim de validar a tipologia textual. A linguagem culinária guarda sua natureza instrucional e técnica de um saber que se prolonga em tradições e por veículos diversos, independentemente do modo como o texto culinário oral e escrito se materializa ou é transmitido.

1.2. O TEXTO TRANSCRITO: AS EXIGÊNCIAS ACERCA DA MATERIALIZAÇÃO DA MODALIDADE ORAL

Sobre os aspectos relacionados especificamente a corpus oral, há de se mencionar uma passagem em que Biderman (2001) discorre sobre o fato de um corpus de material oral transcrito não deixar de ser um conjunto de dados da língua escrita:

Na prática, um *corpus* textual é geralmente entendido como um conjunto de textos da língua escrita. Quanto ao *corpus* oral, ele pode consistir em transcrições ortográficas da língua falada e também gravações acompanhadas das transcrições correspondentes. Enquanto na *Linguística de Corpus* existe uma tendência a considerar as transcrições ortográficas da fala como *corpora* orais (*spoken corpora*), tanto em Fonética como em *Tecnologias da Fala*, dificilmente se concebe um *corpus* que não seja acompanhado do registro sonoro correspondente em formato digital (*speech corpus*). (Biderman, 2001: 79)

Dister e Simon (2008), pesquisadoras do Valibel⁷, local em que se encontra o maior corpus oral francófono para pesquisa, são mais incisivas no que tange à questão da transcrição em estudos linguísticos da modalidade oral. Com o intuito de se estabelecer critérios para a produção de corpus de transcrição, são discutidas as interferências que o transcritor causa por falta de critérios mínimos, e esses efeitos são de ordem ortográfica, mas também pragmática.

Primeiramente, vale ressaltar que os *corpora* orais têm, por natureza, dois níveis de dados; um nível primário relacionado ao áudio, portanto sujeito a uma exploração indireta. Os dados secundários para as pesquisadoras são as transcrições, cuja exploração é direta, pois não é possível realizar análise linguística de produções orais baseando-se exclusivamente na fonte sonora.

On ne peut pas étudier l'oral par l'oral, en se fiant à la mémoire qu'on en garde. On ne peut pas, sans le secours de la représentation visuelle, parcourir l'oral en tous sens et en comparer les morceaux. (Blanche-Benveniste, 2000: 24 *apud* Dister; Simon, 2008: 1).⁸

⁷ Sobre o laboratório: “Le centre Valibel – Discours et variation propose une approche intégrée du langage comme phénomène complexe où les différents niveaux d’analyse sont en interrelation, qu’il s’agisse des unités linguistiques les plus fondamentales (l’unité intonative, le lexème) ou les plus élaborées (le discours, l’interaction verbale).” Para saber mais, acesse: <http://www.uclouvain.be/valibel.html> Acesso em: 01 ago. 2016.

⁸ Dister e Simon (2007) – Acessível em http://icar.univ-lyon2.fr/ecole_thematique/idocora/documents/arena_romanistica_Dister-Simon.pdf Acesso

Portanto, eis o paradoxo insolúvel: para abordar o dito, deve-se obrigatoriamente passá-lo pelo filtro da escrita de um transcritor e, sendo ele humano ou mecânico, há um referencial de códigos que viabilizam a percepção do oral por sua materialização em texto escrito. A transcrição é em si, para Dister e Simon (2007: 02), uma análise. No momento em que se opta por transcrever um som a partir de uma representação gráfica única, uma concepção de língua é adotada e passa a ser representada pelo texto materializado (a transcrição é considerada um texto escrito intermediário, segundo as autoras).

Significa dizer que ao reproduzir exatamente o que se ouve por meio de supressão de vogais, trocas de consoantes próprias de certos grupos ou nichos sociais desprovidos de educação formal ou cacoetes típicos de uma linguagem marginalizada, cria-se a imagem estigmatizada dos falantes pela forma sonora pela qual se expressam. As autoras relatam que esse método em que se apagam ou se ofuscam vogais, como em francês com os exemplos de *p'etit*, *j'suis* e *chui*, no esforço de reproduzir “um jeito de falar” foi utilizado no século XIX por escritores, justamente para estigmatizar o dizer das personagens que representavam as pessoas de origem humilde. Por essas razões, as autoras afirmam que alguns pesquisadores defendem a ideia de que transcrever um áudio, por se tratar de um procedimento que envolve análise, crítica e a transposição para outro código linguístico, em muito se assemelha ao ato de traduzir.

Vale incluir a essa reflexão um dado histórico relatado por Biderman (2001: 62-63) que chama a atenção para o fato de a ortografia ser fruto de uma política de idioma, portanto, uma política de Estado. A constituição de um padrão linguístico nacional se realiza na medida em que uma política centralizadora elege modelos em meio a uma literatura que deveria ser reconhecida como patrimônio nacional e imitada. A força desse código escrito, definitivo e objetivo por natureza, se choca com a sutileza do som que precisa ser capturado por símbolos gráficos para se tornar viável como objeto de estudo.

em: 01 ago. 2016. Tradução nossa: Não é possível estudar o oral pelo oral, confiando na memória que se preserva. Não se pode, sem o recurso da representação visual, percorrer o oral em todos os sentidos e comparar seus recortes.

Existem momentos em que passar da forma oral para uma formalização muito ortodoxa nos parece interferir demasiadamente no “produto original” que está à mão do pesquisador. Optamos por uma ponderabilidade para alcançar um nível de respeito ao que é dito, do modo pelo qual se diz. E, assim como advogam Dister e Simon (2007), em posse dos áudios, cabe ao pesquisador recorrer a esse conteúdo toda vez que julgar necessário esclarecer ou orientar melhor sua análise.

Em uma primeira revisão do corpus, não alteramos a transcrição, senão quanto aos erros de digitação e formalização ortográfica. Aplicamos a pontuação formal – ponto final, exclamação, reticências e afins – porque esses elementos gráficos em nada interfeririam em nossas buscas por UFs. Outra revisão foi realizada posteriormente apenas para corrigir o cabeçalho dos *links* selecionados.

Com essas ponderações feitas sobre os critérios adotados no presente estudo, a nota conclusiva das autoras ganha relevância para nossa pesquisa e outras que tenham como objeto um corpus oral transcrito:

Não é fácil delimitar até onde se revê e que marcas do oral se devem manter, para manter a naturalidade e a fluidez da entrevista, mas parece óbvio que há diferentes níveis de autenticidade que se podem defender. O que é um texto (final)? Para tipos de estudos linguísticos diferentes, diferentes versões seriam preferíveis. Mantendo ambas, permitimos pelo menos dois tipos de estudos: o dos desvios à norma (escrita), e o do conteúdo e forma normalizados. (Taveira; Santos, 2015: 20)

Lemos nesse excerto a clara orientação: para diversos objetivos científicos, diferentes “versões de transcrição” que atendam a essas necessidades, assim como vemos a Linguística de Corpus promover quanto aos seus parâmetros de corpus. Portanto, é importante salientar que não vemos visões contraditórias entre os dois modelos de transcrição demonstrados. Ao contrário, há coerência em ambos os trabalhos quanto das regras e motivações para adotar tal posicionamento diante do objeto.

2. METODOLOGIA PARA A ELABORAÇÃO DO CORPUS: DA SELEÇÃO À COLETA DAS UNIDADES FRASEOLÓGICAS

A primeira etapa da produção do corpus exigiu bastante esforço e tempo. A seleção dos *websites* que respondessem às necessidades prévias dos dados que buscávamos exigiu esforço especial, pois o conteúdo do vídeo e o texto culinário deveriam ser sobre a mesma receita e, além disso, era imprescindível que a qualidade dessas fontes fosse testada para que não houvesse problemas no momento de transcrever os áudios, nem muita discrepância entre o conteúdo do vídeo e o texto da receita.

Para isso, realizamos uma lista prévia de 460 *links* de *websites*, dos quais selecionamos 280 que integraram o corpus. Desse modo, foram 280 receitas com vídeos que foram testadas e avaliadas quanto à adequação. Finalmente 280 vídeos foram transcritos manualmente e 280 receitas compiladas também manualmente.

Foi produzida uma sitografia – relação de *websites* – com cabeçalho discriminando o endereço eletrônico e a receita por um título numerado. Seguem alguns exemplos dos itens selecionados e elencados em uma lista sitográfica, conforme Quadro 1.

Quadro 1: Alguns *websites* selecionados para a compilação do Subcorpus CT.

| |
|--|
| 12. Bolo de cenoura 2 [tudogostoso.bolodecenoura.2] http://www.tudogostoso.com.br/receita/23-bolo-de-cenoura.html |
| 13. Caribéu Pantaneiro 1 [tudogostoso.caribeupantaneiro.1] http://www.tudogostoso.com.br/receita/76055-caribeupantaneiro.html |
| 14. Bolo gelado 3 [tudogostoso.bologelado.3] http://www.tudogostoso.com.br/receita/2313-bolo-gelado.html |
| 15. R. Frango Assado Com Farofa 1 [cyberbook.frangoassadocomfarofa.1] http://cybercook.com.br/receita-de-frango-assado-com-farofa-r-4-58854.html |
| 15. V. Frango Assado Com Farofa 1 [cyberbook.frangoassadocomfarofa.1] http://culinariadativ.blogspot.com.br/2014/10/edu-guedes-hoje-em-dia-frango-assado.html |
| 16. Bolo Salgado de Queijo Coalho 4 [receitasdeminuto.com.bolosalgado.4] http://receitasdeminuto.com/bolo-salgado-de-queijo-coalho/#more-11483 |
| 17. Molho de Mostarda e Mel 1 [receitasdeminuto.molhodemostarda.1] http://receitasdeminuto.com/molho-de-mostarda-e-mel/#more-11475 |

O número 2 da receita de Bolo de Cenoura (12) indica que esse é o segundo bolo inserido na lista. Observa-se que o índice 14 corresponde ao terceiro bolo e o índice 16 traz o quarto tipo de bolo. Utilizamos uma numeração

crecente para itens gerais de identificação de receitas para evitar repetição do mesmo prato ou para assegurar uma variação dos tipos de receitas. Isso significa que a cada bolo inserido, precisávamos a quantidade de bolos que já constava no corpus. Buscamos não passar de 20 receitas de um mesmo tipo de prato, variando a seleção entre pratos doces, salgados, referentes à culinária brasileira e à culinária estrangeira, sobremesas, molhos, pratos principais, acompanhamentos, saladas, sucos etc. Essa preocupação se justifica pela questão da diversidade de termos relevantes que se busca alcançar.

Em alguns casos, como o do item 15, foi necessário criar duas entradas na sitografia por haver um *link* no *website* da receita que remetia o usuário ao Youtube, para a visualização do vídeo. Acreditamos que isso ocorreu por razões técnicas, ou seja, pelo fato de não ser possível inserir o conteúdo todo do vídeo no blog ou na página de acesso da receita. Então, resolvemos o problema com as duas entradas, fazendo referência ao conteúdo da seguinte forma: V para a entrada do vídeo e R para a entrada da receita.

A título de ilustração dos conteúdos gerais, expomos algumas imagens da página virtual da Rita Lobo como um tipo padrão de *website* que buscamos para a pesquisa:

Figura 5: Website de Culinária - GNT - Rita Lobo.



Esse *website* é do programa do canal da GNT em que a apresentadora realiza um prato sem a presença de outros participantes. Os vídeos que são

exibidos pelo canal ficam disponibilizados pelo *website* do canal de televisão. Logo abaixo do vídeo retratado pela imagem inserida na sequência, encontramos o texto da receita da Sangria, completando a Figura 5.

Figura 6: Receita preparada em programa da GNT.

Ingredientes

- 1 garrafa de vinho tinto (750 ml)
- 1 lata de água tônica
- 1 xícara (chá) de caldo de laranja (casca de 1 laranja/bela)
- 1 maçã (fuj)
- 15 abacaxi em cubos médios congelados
- 4 colheres (sopa) de açúcar
- 15 de xícara (chá) de grm (opcional)
- cubos de gelo (a gosto)

Modo de preparo

- 1 Coloque o caldo de laranja em uma jarra grande e reserve.
- 2 Lave, seque e mantenha a casca da maçã. Com um fatiador de legumes, corte a maçã em fatias finas, no sentido da largura (se preferir fatie Fino com a Faca), e transfira para a jarra com o caldo de laranja.
- 3 Junte os cubos de abacaxi congelado, o açúcar e misture com cuidado para não partir as rodela de maçã.
- 4 Leve para a geladeira e deixe macerar por, no mínimo, 30 minutos ou até a hora de servir. Enquanto isso, aproveite para gelar o vinho e a água tônica.
- 5 Na hora de servir, adicione às frutas o vinho, a água tônica e misture. Coloque os cubos de gelo no copo para servir - esse é o segredo para a sangria não ficar aguada.

Cubos de abacaxi congelados para drinque

- 1 Descasque e corte o abacaxi em cubos. Distribua os cubos em uma assadeira sem enfiar ou sobrepor um no outro e leve ao congelador por, pelo menos, 40 minutos. Depois de congelados, transfira os cubos para um saco plástico e volte ao congelador. Assim eles ficam saltonhos e prontos para usar em drinques.

Depois de discutirmos sobre as etapas e desafios da produção do corpus, tratamos da coleta de dados do corpus CT como ilustração do levantamento de UFs que executamos no subcorpus de transcrição.

2.1 ALGUMAS ANÁLISES DO CORPUS DE TRANSCRIÇÃO (CT)

Na primeira triagem, o critério de chavicidade foi aplicado ao corpus para a obtenção de uma primeira amostra de 30 palavras-chave com maior frequência no corpus. Via *playlist* produzida pelo concordanciador UNITEX e transferida para o Excel, constatamos que, das palavras-chave de maior frequência, mais de 75% ocorrem em ambas as modalidades, ou seja, a maioria das UFs está presente simultaneamente no subcorpus de receita (CR) e no subcorpus transcrito (CT).

Foram empregados, portanto, os dois critérios para essa amostra: o de frequência e o de chavicidade para UFs da modalidade oral e da modalidade escrita da linguagem de especialidade. Na Tabela 1, a seguir, temos os dados do

corpus e, na sequência, expomos as informações obtidas a partir das palavras-chave *água*, *alho* e *azeite*. Essa seleção se deu meramente por ordem alfabética das palavras-chave de maior frequência no corpus.

Tabela 1: Dados extraídos do Corpus de Culinária em Português.

| | <i>Tokens</i> | <i>Types</i> | <i>Palavras-chave</i> |
|--------------------------------|---------------|--------------|-----------------------|
| Subcorpus Receitas – CR | 138.843 | 57.007 | 660 |
| Subcorpus Transc. – CT | 319.286 | 133.308 | 685 |
| Total | 458.129 | 190.315 | 1.345 |

2. 2. *ÁGUA*

No Quadro 2, a seguir, foram reunidas as UFs coletadas a partir da palavra-chave *água*. Em negrito, estão as UFs presentes em ambos os subcorpora, como *água fervente*.

Quadro 2: UFs de *água* presentes em CT e em CR.

| Subcorpus CT | Subcorpus CR |
|------------------------------------|------------------------------|
| Água bem quente | Água de coco |
| Água com gás | Água do funghi |
| Água corrente | Água do tomate |
| Água da lavagem | Água fervente |
| Água da massa | Água fria |
| Água de coco | Água gelada |
| Água de flor de laranjeira | Água morna |
| Água do arroz | Colher de sopa de água |
| Água do cozimento | Colher de sopa de água fria |
| Água do polvo | Divisor de águas |
| Água do seu feijão | Latas de água |
| Água em temperatura ambiente | Litros de água |
| Água fervente | ml de água |
| Água gelada | Panela de água fervendo |
| Água na boca | Xícara de água |
| Água quente | Xícara de água em fogo médio |
| Água saborizada com bacalhau | Xícara de água filtrada |
| Aipim cozido com água | Xícara de água fria |
| Amido de milho diluído em água | Xícara de água quente |
| Bowl de gelo com água | |
| Caldo de galinha hidratado na água | |
| Fundinho de água fervente | |
| Gema misturada com água | |
| Litro de água | |
| Panelinha com água fervendo | |
| Parte de água | |
| Polido com jatos de água | |
| Um dedo de água de altura | |
| Um quarto de xícara de água | |
| Vasilha com água | |

Os dados sugerem um uso muito similar do termo-chave tanto ao nível oral como ao nível escrito. Optamos por analisar a UF *água com gás*, *água de coco*, *água fria*, *água do cozimento*, extraindo-se as UFs referentes à medida e à quantidade, ou seja, descartamos os itens riscados como ~~2 copos de água de coco~~, ~~250 ml de leite integral~~, ~~1 colher de canela em pó~~, e assim por diante.

Pelos dados, observamos uma co-ocorrência que não se refere à Culinária, mas à língua geral, que é *água na boca* (07 ocorrências). Outro caso similar foi encontrado em “Essa receita de tomatada é um *divisor de águas* na vida na cozinha”.

Para simplificar nossas exposições neste artigo, fazemos menção apenas às UFs em que *água* é termo-núcleo. As UFs em que a palavra-chave é um colocado (termo secundário para a UF), como em *gema misturada com água*, foram desconsideradas nesse momento.

Uma particularidade que é notada nas frases concordanciadas do CT é o uso da estrutura preposição + artigo definido, em oposição ao uso apenas de preposição em casos em que há dois valores semânticos para *água*, o que se refere à água resultante do cozimento ou a água resultante de uma etapa de preparo de um alimento – *água do arroz*, *água do cozimento da batata* – e outro referente à água própria do alimento, como em *água de coco* e em *água de flor de laranjeira*. Nestes últimos casos, trata-se de ingredientes já prontos ou *in natura*. Inserimos, em seguida, alguns contextos frasais em que as UFs ocorrem.

[1] “Enquanto a clara bate a gente coloca uma chaleira com água para ferver. Ela vai servir para o banho-maria. Assim como o forno deve *tá* preaquecido, quando a gente coloca o pudim para assar, a água do banho-maria também tem que tá quente.”

[2] “Agora você pode colocar a água do seu feijão. Não tem uma regra clara aqui e você não precisa de uma quantidade certa. É uma receita diferente do arroz, que tem uma regrinha. Então aqui a gente recomenda você colocar de dois a três dedos acima do seu feijão.”

[3] “Você vai colocar seu arroz, tá... Pra cozinhar. Tem gente que já coloca uma ou duas colheres de açúcar aqui nessa água do arroz. Se quiser pode fazer isso. Se quiser diminuir as calorias, duas colheres de adoçante culinário.”

[4] “Você vai precisar de arroz já cozido, calda de laranja, coco ralado, açúcar, leite, leite de coco e água de flor de laranjeira. Numa panela, a gente mistura o leite, o leite de coco e o açúcar. Deixa ferver. Põe um pouquinho de água de flor de laranjeira, põe o arroz... Deixa apurar um pouquinho.”.

2.3. ALHO

Pelo concordanciador UNITEX observamos alguns contextos para o lema *alho* que, no CT, fornecem informações que nos fazem aferir que *alho* é empregado essencialmente como tempero – de uso muito frequente na cozinha brasileira. Selecionamos alguns exemplos que comprovam nossa percepção:

[5] “Tempere os filés com *alho*, sal e noz moscada.”.

[6] “Para dar um perfuminho eu vou colocar um *alho*.”

[7] “Tempero com o *amassadinho de alho* dos dois *alhos*.”.

[8] “Para o tempero, amasse num almofariz três dentes de *alhos picados* com uma colher e meia de sal e bastante pimenta branca.”

[9] “Aqui, eu tenho cebola picada pra você. Eu tenho *alho picado* pra você. *Cebola e alho*: a base da culinária brasileira.”.

E há uma ênfase dada ao fato de amargar o gosto do alho, caso queime ou fique muito dourado:

[10] “Quando começar a soltar o aroma do alho, cuidado para não queimar, que ele fica amargo.”

[11] “E por que que eu boto primeiro a cebola, antes do *alho*? Porque a cebola... Ela queima mais devagar que o *alho*. Se eu colocar o *alho* antes, ele vai dourar... E quando eu colocar a cebola, ele vai queimar.”

No Quadro 3, a seguir, estão inseridas as UFs de *alho*:

Quadro 3: UFs extraídas a partir da palavra-chave *alho*.

| Subcorpus CT | Subcorpus CR |
|---|-------------------|
| Alho assado espremido sem pele | Alho amassado |
| Alho em pó | Alho comum |
| Alho fatiado | Alho em pó |
| Alho Frito | Alho picadinho |
| Alho picado | Alho picado |
| Alho Poró | Alho poró |
| Alho torrado | Alho poró fatiado |
| Alho Vermelho da região de Campos Altos | Cabeça de alho |

| | |
|---------------------------|--|
| Crouton de alho | Cabeça de alho assada |
| Dente de alho | Dente de alho em cubinhos |
| Dentes de alho sem casca | Dente de alho picadinho |
| Farofinha de pão com alho | Dente de alho picado e esmagado |
| Pasta de alho | Dente de alho sem casca |
| Talo de alho Poró | Dentes de alho com pele |
| | Dentes de alho cortados grosseiramente |
| | Dentes de alho fatiados |
| | Dentes de alho inteiros descascados |
| | Dentes de alho inteiros e sem casca |
| | Dentes de alho ralados |
| | Maço de alho poró |
| | Talo de alho poró |

Alho Poró aparece em UFs com colocados *talho* e *maço*, enquanto *alho* aparece associado a *dente* e *cabeça*. Em CR a ocorrência de UFs com uma variedade maior de unidades lexicais provavelmente se deve ao fato de haver restrição de espaço para a redação do modo de preparo e por haver uma tendência em incluir toda a descrição do ingrediente na lista inicial dos itens utilizados na receita. Isso pode explicar o que apresentamos nos seguintes excertos do CR:

[12] “Em uma panela de fondue, passe o dente de *alho* sem casca cortado ao meio.”.

[13] “Quinze dentes de *alho* inteiros descascados”

[14] “Duas colheres de sopa de *alho* poro fatiado”

No entanto, nos vídeos, o interlocutor explica e demonstra como esse ingrediente deve ser preparado, e esse gesto é “conduzido” geralmente por frases mais longas – uma conversa contínua – porém com UFs mais reduzidas – com menos elementos lexicais. Com trechos do CT exemplificamos o que expomos agora:

[15] “Para fazer o camarão, refogue o *alho* e a cebola no azeite por três minutos. Coloque o camarão, o molho de tomate e tempere com sal a gosto.”

[16] “Olá... Eu vou fazer agora o salmão ao champignon. Pra marinar o salmão eu vou usar limão. Vou temperar aqui... Um pouco de *alho* e sal. Vou virar aqui o salmão e vou colocar um pouco mais de sal desse lado.”

[17] “Enquanto isso refogue a cebola e o *alho* já picados, até ficarem bem douradinhos... E depois reserve.”

[18] “Então é só adicionar um *alho* picado, refogar por mais uns 30 segundos e agora você vai voltar sua carne pra panela.”

Na frase [17], não há especificidade sobre como devem ser picados o *alho* e a cebola, e no CR o detalhamento é maior, com uso de advérbios ou outras UFs, como nesse outro exemplo do subcorpus de receitas: dois dentes de *alho* cortados grosseiramente.

A frase [18] não traz a medida exata de *alho* – uma cabeça ou um dente?

Pelos contextos frasais é possível identificar palavras-chave que não são encontradas em receitas, formas sintáticas mais recorrentes na modalidade oral que empregam UFs que não são encontradas no outro subcorpus CR.

2.4. AZEITE

Pelos contextos das ocorrências de *azeite* no CT é correto inferir que esse ingrediente é usado para untar, refogar, temperar e preparar molhos:

[19] Tempere o iogurte com sal, pimenta e *azeite*.

[20] Pra fazer o molho a gente precisa de *azeite*, mostarda de Dijon, raspas da casca e suco de limão siciliano, sal, pimenta do reino e um pouquinho do iogurte natural.

[21] Se precisar unte com *azeite* ou com manteiga para não grudar.

[22] Em uma panela a gente vai refogar o tofu com *azeite*.

Também há uma referência de medida para uma UF muito usada, sem precisão – *fio de azeite*:

[23] Na frigideira eu vou colocar um *fio de azeite*. Mais ou menos duas colheres de sopa.

No CR encontramos os mesmos valores de uso desse ingrediente, com emprego dos verbos no infinitivo e no particípio – diferentemente do que foi extraído no CT:

[24] Em uma travessa untada com *azeite*, coloque metade da massa, distribua o recheio [...].

[25] Refogar a cebola com *azeite* numa panela média.

[26] Refogue o tofu com gergelim no *azeite* até dourar.

[27] [...] a versão que eu preparava era particular, com folhas variadas, nozes pecã, shiitake refogado no *azeite* e temperado com sal e pimenta e bastante dill [...].

Há apenas uma ocorrência de *azeite de ervas* no CR – nenhuma no CT – e essa UF aparece inserida num contexto de texto escrito bem interessante:

[28] Ingredientes: Laranjas (uma laranja por pessoa); Cenouras (usei 5 cenouras médias); *Azeite de oliva* (usei o *azeite de ervas* que o link da receita está no final da página) [...].

As inserções feitas na receita, entre parênteses e em primeira pessoa, nos remetem às receitas do livro de Culinária para as classes trabalhadoras (p. 41 a 43) quando o chef “conversa” com seu leitor. É importante perceber que a Internet é um terreno fértil para que esse dialogismo ocorra dando fluidez ao texto, assim como as interações dos interlocutores nos vídeos dinamizam a apresentação e entretêm o espectador. Sobre as UFs coletadas do corpus, nós as inserimos no Quadro 4:

Quadro 4: UFs obtidas pelas buscas com a palavra-chave *azeite*.

| Subcorpus CT | Subcorpus CR |
|------------------------------|------------------------------|
| Azeite de dendê | Azeite (a gosto) |
| Azeite de oliva | Azeite a gosto |
| Azeite de oliva extra virgem | Azeite bem quente |
| Duas colheres de azeite | Azeite de dendê |
| Um fiozinho de azeite | Azeite de ervas |
| Um golinho de azeite | Azeite de oliva |
| Um pingo de azeite | Azeite de oliva com ervas |
| Um toque de azeite de dendê | Azeite de oliva extra virgem |
| Um toquezinho de azeite | Azeite do bacalhau |
| Uma boa quantidade de azeite | Azeite extra virgem |
| | Azeite extra virgem a gosto |
| | Azeite extravirgem |
| | Colher de azeite |
| | Fio de azeite |

Quanto a essa unidade lexical, *azeite*, as informações do corpus evidenciam alguns aspectos curiosos: a palavra-chave ocorre sozinha no corpus quando sua distinção semântica já foi mencionada e, portanto, já é evidente. Isto é, *azeite* sem nenhum outro elemento colocado é um termo equivalente a *óleo de oliva* ou *azeite de oliva* e, geralmente do tipo extravirgem. Os outros tipos de azeite reaparecem no corpus acompanhados dos seus elementos definitórios – *azeite de dendê*, *azeite de oliva com ervas* ou *azeite de ervas*.

Sobre a ortografia, encontram-se duas formas nas receitas compiladas: *extra virgem* e *extravirgem*, como já apresentadas no Quadro 4.

Depois de realizado o levantamento das UFs de cada subcorpus, exibimos os dados em tabelas divididas em UFs comuns, UFs do CR e UFs do CT, e

também inserimos as informações quantitativas obtidas pelo concordanciador pelas buscas com as palavras-chave, como podemos observar na sequência, nos Quadros 5, 6 e 7.

Quadro 5: UFs e dados referentes à palavra-chave *água*.

| Água | CR: 207 ocorrências – 0,194% | CT: 404 ocorrências – 0,126% |
|----------------------------|-------------------------------------|-------------------------------------|
| 11 UFs comuns | 6 UFs | 14 UFs |
| Água com sal | Água do funghi | Água transparente |
| Água de coco | Água do macarrão | Água turva |
| Água de flor de laranjeira | Água do tomate | Jatos de água |
| Água do banho maria | Água filtrada | Dedo de água quente* |
| Água do cozimento | Água limpa | Água ardente de pera |
| Água fervendo | Água sem óleo | Água branca |
| Água fervente | | Água da alcaparra |
| Água fria | | Água da lavagem |
| Água gelada | | Água da massa |
| Água morna | | Água em temperatura ambiente |
| Água quente | | Água do arroz |
| | | Água do polvo |
| | | Água rósea |
| | | Água saborizada com bacalhau |

Um termo coloquial usado nas cozinhas doméstica é *dedo* como unidade de medida, como consta no corpus em estudo. Pode-se pensar que se trata de um caso de idioleto, por exemplo. Essa dúvida pode ser respondida pelo próprio corpus, analisando a frequência desse termo específico, assim como sua procedência, segundo o cabeçalho das receitas e dos vídeos que guardam essa referência. Por essa razão, nos atentamos para a informação de cada nova receita inserida e cada nova transcrição finalizada. Não se trata de uma questão para nossos estudos, além de não haver um peso significativo com relação às ocorrências totais obtidas.

Em seguida, expomos as UFs encontradas a partir das buscas com *alho*:

Quadro 6: UFs e dados referentes à palavra-chave *alho*.

| Alho | CR: 176 ocorrências – 0,126% | CT: 209 ocorrências – 0,065% |
|---------------------|-------------------------------------|-------------------------------------|
| 4 UFs Comuns | 12 UFs | 18 UFs |
| Alho em pó | Alho amassado | Alho assado espremido sem pele |
| Alho picado | Alho assado | Alho descascado |
| Alho Poró | Alho assado espremido | Alho dourado |
| Dente de alho | Alho comum | Alho em lâminas |
| | Alho cortado | Alho espremido |
| | Alho picado e esmagado | Alho fatiado |
| | Alho Poró fatiado | Alho frito |

| | | |
|--|---------------------------|---|
| | Cabeça de alho | Alho ralado |
| | Dente de alho em cubinhos | Alho refogado |
| | Dente de alho fatiado | Alho vermelho |
| | Dente de alho picado | Aroma do alho |
| | Maço de alho Poró | Cheiro do alho |
| | | Dente de alho amassado |
| | | Dente de alho sem casca cortado no meio |
| | | Pasta de alho |
| | | Sabor do alho |
| | | Talo de alho Poró |
| | | Tempero alho e sal |

As buscas realizadas com *alho* revelam uma particularidade: há dois valores para esse termo-chave, um, que atrai elementos lexicais como *cabeça* e *dente*, e certos qualificadores como amassado, descascado e assado, além de formar a UF *alho e sal* e outro que compõe a UF *alho Poró* e co-ocorre com *maço* no corpus. Notamos que por esse confronto de UFs distinguimos elementos lexicais concorrentes, a partir de dados quantificados pelo concordanciador, extraídos do corpus. Esse procedimento é aplicado ao longo de todo o estudo, a fim de garantir precisão aos resultados, além da autonomia com a qual o linguista opera suas análises, independentemente de haver ou não um especialista para validar os resultados.

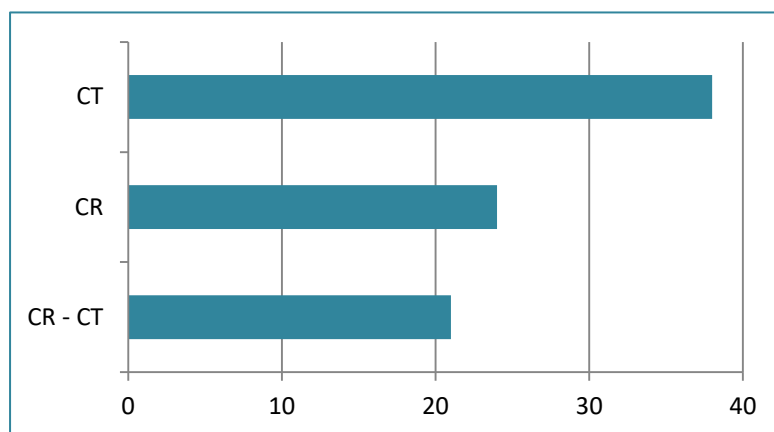
Analiseemos, então, os dados de *alho*, que incluem expressões familiares da Culinária, destacadas com asterisco, e outras UFs padronizadas como *azeite a gosto* ou *colher de azeite*.

Quadro 7: UFs e dados referentes à palavra-chave *azeite*.

| Azeite | CR: 213 ocorrências – 0,153% | CT: 244 ocorrências – 0,076% |
|-----------------------------|-------------------------------------|-------------------------------------|
| 6 UFs comuns | 6 UFs | 6 UFs |
| Azeite de dendê | Azeite a gosto | Azeite de oliva extra virgem |
| Azeite de oliva | Azeite (bem) quente | Boa quantidade de azeite* |
| Azeite de oliva extravirgem | Azeite de ervas | Golinho de azeite* |
| Azeite extravirgem | Azeite de oliva com ervas | Pingo de azeite* |
| Colher de azeite | Azeite do bacalhau | Toque de azeite de dendê* |
| Fio de azeite* | Azeite extravirgem a gosto | Toquezinho de azeite* |

Para concluir, inserimos os dados no Gráfico 1, a seguir, para melhor visualização, onde notamos que, para os exemplos apresentados aqui, o maior número de UFs foi obtido no CT das palavras-chave, havendo certo equilíbrio entre UFs comuns e UFs do CR:

Gráfico 1: UFs coletadas no Corpus a partir de três palavras-chave.



Esse breve levantamento realizado para o artigo aponta para a pertinência do estudo em desenvolvimento, sabendo-se que uma parte significativa dos trabalhos apoiados em Linguística de Corpus e Terminologia empregam exclusivamente documentos escritos em suas análises devido ao fato de haver corpora mais numerosos e mais volumosos exclusivamente formados por compilação de documentos escritos. Isso ocorre também por razão de praticidade ou pela inacessibilidade a meios tecnológicos que possibilitem a produção e compilação automática de transcrições. Dentre os corpora para pesquisa em que há predominância de conteúdo escrito, podemos citar o COMET, o Linguateca e o CORPOBRAS, disponíveis em rede.

CONSIDERAÇÕES FINAIS

Neste artigo, mostramos as principais etapas da pesquisa sobre Fraseologia da Culinária em rede, descrevendo a trajetória da sua concepção, pontos relevantes do desenvolvimento do estudo com exemplos ilustrativos, preservando a questão norteadora que assumimos ao elaborar o projeto de pesquisa: por meio de unidades fraseológicas é possível detectar o grau de complementariedade entre essas esferas de um mesmo recorte linguístico?

As inferências iniciais que foram apresentadas nesse artigo corroboram a ideia de que as linguagens de especialidade ganham com a inclusão da modalidade como objeto de estudo definitivo em suas pesquisas, e o estudioso

da Língua e de suas subáreas amplia seu horizonte de análise e sua autonomia com relação à validação dos resultados alcançados.

Nosso trabalho se propõe também a trazer ao debate a linguagem de especialidade disponível em rede, propondo com isso desvencilhar a investigação ao conceito de padrão textual, como o do gênero receita culinária. Demonstramos como o corpus foi concebido, as particularidades do trabalho de transcrição e os critérios de seleção de conteúdo do banco de dados a fim de atender a contento as demandas estabelecidas pelo estudo observacional comparativo.

Do ponto de vista linguístico e teórico, há muito a ser feito ainda. Contudo, foi possível apresentar nossos procedimentos e resumir a pesquisa nas principais etapas: produção do corpus, elaboração da metodologia relacionada à coleta dos dados e seleção das UFs representativas ao domínio de linguagem estudado, e, finalmente, a análise dos primeiros dados coletados.

REFERÊNCIAS BIBLIOGRÁFICAS

1. BEVILACQUA, Cleci Regina; SALGADO, Ana Rachel; DA SILVEIRA, Diônifer Alan. Unidades fraseológicas especializadas eventivas: novas questões sobre seu reconhecimento em corpora textuais. Intercâmbio. Revista do Programa de Estudos Pós-Graduados em Linguística Aplicada e Estudos da Linguagem. ISSN 2237-759X, v. 15, 2006.
2. BERBER SARDINHA, Tony. Linguística de corpus. Editora Manole Ltda, 2004.
3. BIBER, Douglas; CONRAD, Susan; REPPEN, Randi. Corpus linguistics: Investigating language structure and use. Cambridge University Press, 1998.
4. BIDERMAN, Maria Tereza Camargo. Teoria lingüística. Teoria lexical e, 2001.
5. DISTER, Anne; SIMON, Anne Catherine. La transcription synchronisée des corpus oraux. Un aller-retour entre théorie, méthodologie et traitement informatisé. *Arena Romanistica*, v. 1, n. 1, p. 54-79, 2008.

6. FRANCATELLI, Charles. Plain Cookery Book for the Working Classes. Tradução de Renata Lúcia Bottini, 1a. Ed. 2001. BoD–Books on Demand, 2009.
7. GARABATOS, Jesús Lago; REY, María Isabel González. Analyse des caractéristiques des unités phraséologiques. In: Análisis del discurso. Universidad de Granada, 2000. p. 119-137.
8. GROSS, G. Classes d’objet et descriptions des verbes. *Langages* n. 115 p. 15-30, 1994.
9. LAPORTE, E. Exemples attestés et exemples construits dans la pratique du lexique-grammaire. Paris: França, 2007. Disponível em: < <http://hal.archives-ouvertes.fr/docs/00/35/26/95/PDF/ObsManipLexGram.pdf> >. Acesso em: 05 jun. 2017.
10. ODORISSIO, R. M. As classes de objeto e a linguística de corpus na construção de equivalências para glossário francês português: o verbo na cozinha. 2011.
11. PAUMIER, Sébastien. "Unitex–Manual de utilização." Université Marne-la-Valée (2002).
12. REBECHI, Rozane Rodrigues. A tradução da culinária típica brasileira para o inglês: um estudo sob o enfoque da Linguística de Corpus. Tese de Doutorado. Universidade de São Paulo, 2015. Acessível em <http://www.teses.usp.br/teses/disponiveis/8/8147/tde-29092015-162829/en.php> Acesso em: 13 mar. 2017.
13. SANTOS, D. Comparando corpos orais (transcritos) e escritos na Gramateca. In Camilla Bardel (ed.), *Proceedings from the conference Parler les langues romanes/Parlare le lingue romanze/Hablar las lenguas romances/Falando línguas românicas GSCP 2014*, University Press Università di Napoli L'Orientale, 2014. Disponível em <http://www.linguateca.pt/Diana/download/SantosGSCP2014.pdf>. Acesso em: 23 maio 2017.
14. SINCLAIR, J. Corpus, concordance, collocation. Oxford: Oxford University Press, 1991.
15. TAGNIN, S. E. O. Linguística de Corpus e Fraseologia. In: Maria Luisa Ortiz Alvarez & Enrique Huelva Unternbäumen. (Org.). Uma (Re)visão

- da Teoria e da Pesquisa Fraseológicas. Campinas: Pontes Editores, 2011, p. 277 – 302.
16. TAVEIRA, Paula; SANTOS, Diana. Ensaio sobre a revisão da oralidade. 2015.
 17. TEIXEIRA, E. D. A linguística de corpus a serviço do tradutor: proposta de um dicionário de culinária voltado para a produção textual. Tese (Doutorado em Linguística). Faculdade de Filosofia, Letras e Ciências Humanas (FFLCH). São Paulo, 2008. Disponível em: http://www.teses.usp.br/teses/disponiveis/8/8147/tde-16022009-141747/publico/TESE_ELISA_DUARTE_TEIXEIRA.pdf Acesso em: 10 jan. 2017.
 18. XATARA, C.; PARREIRA, M. C. A elaboração de um dicionário fraseológico. In: Maria Luisa Ortiz Alvarez & Enrique Huelva Unternbäumen. (Org.). Uma (Re)visão da Teoria e da Pesquisa Fraseológicas. Campinas: Pontes Editores, 2011, p. 69-75.

Recebido em: 05/06/2017

Aceito em: 1º/08/2017