XV ENCONTRO DE LINGUÍSTICA DE CORPUS - 2021

O alcance da Linguística de Corpus: do Léxico ao Discurso

Organização:
Regiani A. S. Zacarias e
Stella E. O. Tagnin

Corpus

O presente trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Comissão Organizadora do evento:

Elaine Alves Trindade (USP)
Giovana de Castro Marchese Rampini (USP)
Jamilly Brandão Alvino (USP)
Marina Leivas Waquil (USP)
Regiani Aparecida Santos Zacarias (UNESP)
Stella E. O. Tagnin (USP)

Apoio: Programa de Pós-Graduação Estudos Linguísticos e Literários em Inglês (PPGELLI/USP) e Programa de Pós-Graduação Linguística e Língua Portuguesa (UNESP/FCLAr)

Projeto Gráfico e Editoração: Milxtor Arte

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

XV Encontro de Linguística de Corpus [livro eletrônico] : o alcance da linguística de corpus : do léxico ao discurso / organização Regiani A. S. Zacarias, Stella E. O. Tagnin. -- São Paulo, SP : Ed. dos Autores, 2023.
PDF

Vários autores.
Bibliografia.
ISBN 978-65-00-84946-2

1. Lexicologia 2. Lingüística - Congressos 3. Terminologia 4. Tradução I. Zacarias, Regiani A. S. II. Tagnin, Stella E. O.

23-179091 CDD-418

Índices para catálogo sistemático:

1. Lingüística : Congressos 418

Tábata Alves da Silva - Bibliotecária - CRB-8/9253

Publicação segundo as regras do novo Acordo Ortográfico da Língua Portuguesa.

Regiani A. S. Zacarias
Stella E. O. Tagnin
organizadoras

XV Encontro de
Linguística de Corpus

O alcance da Linguística de Corpus: do Léxico ao Discurso

Corpus

São Paulo, 2023

Sumário

Homenagem

Neste livro, que reune trabalhos apresentados no XV Encontro de Linguística de Corpus (ELC), realizado em 2021, gostaríamos de prestar nossa sincera homenagem à Professora Doutora Stella Esther Ortweiler Tagnin, idealizadora e organizadora do ELC desde a sua primeira realização em 1999.

A primeira edição do ELC teve como protagonista o Prof. Göran Kjellmer, da Universidade de Gotemburgo que esteve no Brasil a convite da Profa. Stella. Em visita à USP, além da participação no ELC, o Prof. Kjellmer ministrou um curso de LC. O contato e a parceria com o professor Kjellmer tiveram início em 1995 por ocasião da viagem da professora ao *Third Language International Conference on Translator and Interpreter Training* na Dinamarca. Conta-nos a Profa. Stella que o Prof. Kjellmer prontamente a recebeu e disponibilizou o corpus *Brown*, que lhe proporcionou o primeiro encontro com a Linguística de Corpus.

O contato com o Prof. Kjellmer foi o grande marco na carreira da Profa. Stella como promotora da Linguística de Corpus no Brasil, compromisso assumido e que tem sido concretizado e fortalecido ao longo de todos esses anos. Podemos afirmar que o ELC somente chegou em sua XV edição devido à competência e determinação contagiantes da professora Stella.

Em uma breve busca por palavras no texto “Linguística de Corpus no Brasil, ontem e hoje: entrevista atualizada com Stella Tagnin”, concedida e magistralmente conduzida pela Profa. Dra. Elisa Duarte Teixeira para a revista **TradTerm**, Vol. 37 nº 2, publicada em 2021 – edição comemorativa aos 20 anos da LC na USP –, constatamos:

Quadro 1: Nuvem de palavras

Autora: Elaine Trindade

De fato, a convivência acadêmica e pessoal com a Profa. Stella traz para o concreto todos os conceitos e os 'abstratos' das palavras a ela associadas. Com seu jeito único e especial, a Profa. Stella acolhe, orienta, instrui e proporciona aprendizado e crescimento. É de sua natureza valorizar a pessoa e potencializar as suas competências e habilidades para o sucesso no campo acadêmico-científico e, consequentemente, na vida pessoal. Em universidades no Brasil e no exterior atuam alunos e alunas egressos/as que divulgam a Linguística de Corpus e espalham as mesmas sementes que aprenderam a cultivar com ela.

Quadro 2: LancsBox

Autora: Elaine Trindade

O sucesso daqueles/as que tiveram o privilégio de sua orientação atribui ainda mais luz ao brilhantismo de sua trajetória, inegavelmente um exemplo de dedicação à formação de cientistas da Linguística de Corpus.

As edições do ELC comprovam a repercussão e alcance de seu compromisso constante no ensino e divulgação da LC, pois todas as pesquisas apresentadas nos eventos, sejam aquelas embasadas nas teorias da LC e da Tradução ou no uso da abordagem da LC como etapas investigativas sobre o comportamento lexical, há evidências da contribuição da Profa. Stella, quer por meio de orientações (ou a elas relacionadas), participações em bancas ou leitura de suas publicações:

Quadro 3: Professora Stella em números

- **Orientações concluídas**:
  38 mestrados e 24 doutorados
  4 supervisões de pós-doutorado
  13 iniciações científicas

- **Produção bibliográfica**:
  28 artigos
  19 livros publicados/organizados
  23 capítulos de livros publicados
  28 trabalhos completos em anais de congressos
  90 apresentações de trabalhos
  28 publicações da "Coluna da Stella" no blog da Disal, entre 2011 e 2017

- **Participação em bancas**:
  47 mestrados
  37 doutorados

- **Qualificações**:
  19 mestrados
  26 doutorados

- **Eventos**:
  organização de 22 eventos
  participação em 114 eventos

- **Em andamento** (2023):
  3 dissertações de mestrado
  1 tese de doutorado
  2 supervisões de pós doutorado

Autora: Giovana de Castro Marchese Rampini

Os números quantificam e revelam a expressiva produtividade de sua atuação. Qualitativamente os resultados são imensuráveis. A sua presença inspira sucesso e competência em todo e qualquer tipo de atuação acadêmico-científica da qual participa.

A realização do **XV ELC**, em 2021, uniu uma equipe de profissionais, ex-orientandas e orientandas em andamento em uma força tarefa desafiadora para garantir a qualidade do evento em sua versão online em contexto de pandemia. O sucesso foi garantido, assim como a certeza de que a força, entusiasmo, competência e compromisso desta equipe são inabaláveis.

Por tudo isso é que queremos registrar a nossa homenagem à nossa querida Professora Doutora Stella Esther Ortweiler Tagnin. Parabéns e obrigada por levar conhecimento e oportunizar a troca de experiências científicas nos Encontros de Linguística Corpus. Você certamente fez e faz a diferença nas pesquisas de um grande número de pessoas.

Abraço querido,

Equipe XV ELC-2021: Elaine Alves Trindade
Giovana de Castro Marchese Rampini
Jamilly Brandão Alvino
Marina Leivas Waquil
Regiani Aparecida Santos Zacarias

Apresentação

O Encontro de Linguística de Corpus (ELC) é dedicado a todos os estudos que envolvem o uso de "corpora eletrônicos".

O tema da edição 2021 – o XV ELC - foi "O alcance da Linguística de Corpus: do Léxico ao Discurso", pois pretendíamos apresentar um panorama das possibilidades de aplicação da Linguística de Corpus, muitas das quais já abordadas em edições anteriores. Também esperávamos que fosse uma comemoração, embora um pouco tardia, dos 20 anos desse evento no Brasil, que teve sua primeira edição em 1999, quando da visita do Prof. Göran Kjellmer, da Universidade de Gotemburgo, a quem somos imensamente gratas por nos ter apresentado a essa disciplina investigativa que oportuniza pesquisas em campos tão diversificados.

Por meio da análise de grande quantidade de textos, denominados corpora, a Linguística de Corpus permite averiguar intuições e hipóteses sobre a língua em vários níveis da linguagem, do morfossintático ao pragmático, do lexical ao discursivo, do sociolinguístico ao cultural. É cada vez maior o número de campos do conhecimento que se valem da Linguística de Corpus, dentre os quais podemos citar a Lexicologia e a Lexicografia, a Terminologia, os Estudos de Tradução, a Linguística Computacional, a Análise Crítica do Discurso, a Linguística Forense, a Pragmática, a Análise de Sentimentos, a Literatura, dentre outros que dela vão se apropriando para suas pesquisas.

O presente volume reflete essa variedade de abordagens. No artigo de Carlos Henrique Kauffmann, intitulado "Dimensões de variação lexicogramatical na caracterização do estilo machadiano", o autor investiga o estilo da ficção de Machado de Assis na perspectiva da linguística de corpus e caracteriza o estilo machadiano por meio da variação linguística encontrada no corpus

CLIMA (859,5 mil palavras), composto pelos nove romances e 76 contos publicados em livro pelo escritor. Para tanto, empregou a técnica da análise multidimensional, nas vertentes funcional e lexical (Berber Sardinha; Veirano Pinto 2019). O uso correlacionado de características gramaticais e de lemas fez surgir dois conjuntos de dimensões funcionais e lexicais, que, em seguida, foram integrados em uma análise canônica (Mayer 2018). O estudo apresenta como resultado três dimensões estéticas de cunho lexicogramatical formadas por dimensões de ambos os conjuntos em proporções diversas que geram uma interpretação estilística da obra singular de Machado de Assis.

Ariel Novodvorski, em seu estudo "Relações metafóricas e fraseologia especializada em corpus de Humor Político", vale-se de um corpus com 406 artigos extraídos da coluna dominical de Humor Político do jornal argentino Clarin, cobrindo um período de dez anos. O autor explorou o corpus a partir de três perspectivas: a terminológica para identificar candidatos que possam ser considerados como ocorrências especializadas; a fraseológica para determinar se esses candidatos atendem aos requisitos para serem classificados como fraseologias e, por fim, a metafórica, para a identificação das metáforas conceptuais subjacentes às fraseologias identificadas.

Nunes e Mello apresentam os dados iniciais de um estudo inovador do português brasileiro (PB) falado, uma área ainda pouco explorada academicamente. O trabalho baseia-se nos corpora que compõem o C-Oral-Brasil, composto de quatro tipos distintos de fala: a) informal, b) formal, c) da mídia e d) telefônica. Valendo-se de diversas metodologias, buscam, em "Os corpora C-ORAL-BRASIL como uma fonte de informação lexical da fala", identificar o repertório e as variações lexicais que ocorrem nesses tipos.

Segue-se o artigo de Andressa Costa que aborda a temática "Os Discursos Feministas no Brasil e na Alemanha: um estudo contrastivo assistido por corpus sobre suas estruturas temáticas" em

que analisa e contrasta os discursos de feministas na Alemanha e no Brasil na perspectiva da linguística de corpus. A base de dados compõe-se dos corpora feminaDE (alemão) e feminaBR (português), compilados para o projeto. A autora esclarece que a modelagem de tópicos (*topic modeling*) com o modelo LDA foi o método usado para identificar temas latentes nos dois corpora considerando-se que as palavras são variáveis observáveis e a estrutura de tópicos, *cluster* de palavras que coocorrem nos textos, são as variáveis escondidas. As variáveis observadas compõem-se de lemas de substantivos, verbos, adjetivos e advérbios. Segundo a autora, os resultados mostram que há mais diferenças do que semelhanças em relação à composição dos tópicos. Mesmo havendo várias palavras-chave comuns nos dois corpora, elas coocorrem com diferentes palavras-chave criando diferentes estruturas temáticas. Por exemplo, apenas um tópico no corpus brasileiro apresenta estrutura temática similar a outros dois tópicos no corpus alemão.

O artigo intitulado *"We use "excuse me" when our body makes noises: A corpus-based contrastive study of pragmatic routines in undergraduate translation courses in China"*, de autoria de Malila Prado, Adriana Mendes Porcellato e Xiao Wang, apresenta um estudo contrastivo de fórmulas de rotina no inglês britânico e no chinês. Baseia-se especificamente no uso de *excuse me* em seis contextos situacionais diversos, conforme identificados pelos alunos a partir de um corpus de língua oral, o Spoken BNC2014. O uso de tarefas de preenchimento de situações (denominadas *Discourse Completion Task - DCT*) permitiu que os graduandos chineses da disciplina de Estudos de Corpus na Tradução identificassem as expressões que seriam usadas em sua cultura em situações equivalentes. Os resultados obtidos foram investigados num corpus oral de língua chinesa para sua validação. O estudo promoveu a conscientização pragmática dos alunos salientando a importância de se conhecer as diferenças culturais para um convívio intercultural satisfatório.

O artigo de Anna Catharina de Mendonça Paes traz "Um estudo da função conativa em um corpus paralelo através da pesquisa de verbos no modo imperativo por meio do Sketch Engine" a partir de um corpus paralelo composto de um conjunto de treze textos em inglês e de suas respectivas traduções para o português do antigo website da Escola Moderna de Mistérios (sede brasileira). A análise dos textos, utilizando o *software* Sketch Engine (KILGARRIFF et al., 2004), trouxe novas informações sobre o corpus, como o uso da função conativa que se evidencia pela alta frequência dos pronomes "*you*" e "você". Para investigar mais a fundo a função conativa no corpus paralelo, foram levantados os verbos de maior ocorrência e desses comparou-se a ocorrência de verbos no modo imperativo nos textos traduzidos e nos textos fonte.

Outro artigo que comprova o alcance de aplicações da Linguística de Corpus é assinado por Patrícia Tosqui-Lucks e Rafaela Araújo Jordão Rigaud Peixoto. Como o título "Análise da tradução de termos da área de Meteorologia da Fraseologia Padrão Aeronáutica" já adianta, as autoras debruçam-se sobre a fraseologia aeronáutica relativa à meteorologia, analisando como os termos em língua portuguesa são vertidos para a língua inglesa em material institucional destinado a controladores do tráfego aéreo e pilotos. Esses termos são então buscados num corpus constituído por Fraseologias originalmente usadas em língua inglesa por várias instituições aeronáuticas internacionais. O estudo revelou que várias delas não correspondem às fraseologias empregadas em situações não rotineiras, quando os profissionais têm de recorrer ao *plain English*. Em situações de emergência, é imprescindível uma comunicação clara e precisa, razão pela qual os equivalentes fraseológicos identificados podem ser submetidos ao Processo de Revisão Normativa (PRENOR), do Departamento de Controle do Espaço Aéreo (DECEA). Em suma, um artigo que pode ter um impacto positivo na comunicação aeronáutica.

O último artigo – *last but not least*, como se diz em inglês – traz a Linguística de Corpus à área comercial. Em "Transdisciplinaridade na Tecnologia da Linguagem", Ana Claudia Zandavalle e Livy Real preconizam o trabalho conjunto de profissionais de diversas áreas para alcançar melhores resultados sem ignorar os desafios inerentes a essa abordagem, tanto no campo profissional quanto no pessoal, exigindo poder de escuta, aprendizado colaborativo e estratégias de comunicação. A Linguística de Corpus se faz presente, como solução investigativa, para a analise de uma grande quantidade de *feedbacks* de usuários com o objetivo de, a partir dos achados, implantar melhorias na empresa.

A coletânea resultante do XV Encontro de Linguística de Corpus (ELC) revela que, após 20 anos de sua chegada ao Brasil, a LC consolidou-se como disciplina que, além de dialogar com outras disciplinas científicas e colaborar para a geração de conhecimento teórico fundamentado na exatidão de resultados automatizados, firma parceria com as novas tendências tecnológicas e mercadológicas. A LC projetou-se para além do ambiente acadêmico-científico e encontra-se presente em vários segmentos da sociedade, como no apoio ao profissional da aeronáutica, na interlocução entre consumidor e empresa, na compreensão de aspectos interculturais e na constatação e avaliação de questões linguísticas de natureza qualitativa. Para onde segue o alcance da LC? Esta é uma pergunta que somente os próximos Encontros de Linguística de Corpus poderão nos dizer!

Regiani A. S. Zacarias
Stella E. O. Tagnin

Dimensões de variação lexicogramatical na caracterização do estilo machadiano

Characterizing Machado de Assis's style through dimensions of lexicogrammatical variation

Palavra tem sexo.
— Mas, então, amam-se umas às outras?
Amam-se umas às outras. E casam-se.
O casamento delas é o que chamamos estilo.
(Machado de Assis, *O cônego* ou *Metafísica do estilo*)

Carlos Henrique Kauffmann[1]

1 GELC, LAEL/PUC-SP.

***Resumo**:* Este estudo investiga o estilo da ficção de Machado de Assis na perspectiva da linguística de corpus. Buscou-se caracterizar o estilo machadiano por meio da variação linguística encontrada no corpus CLIMA (859,5 mil palavras), composto pelos nove romances e 76 contos publicados em livro pelo escritor. A técnica principal empregada foi a análise multidimensional, nas vertentes funcional e lexical (Berber Sardinha; Veirano Pinto 2019). Do uso correlacionado de características gramaticais e lemas emergiram, respectivamente, dois conjuntos de dimensões funcionais e lexicais, que, em seguida, foram integrados em uma análise canônica (Mayer 2018). Resultaram três dimensões estéticas de cunho lexicogramatical formadas por dimensões de ambos os conjuntos em proporções diversas, capazes de gerar uma interpretação estilística nuançada da obra singular de Machado de Assis.

***Palavras-chave**:* Análise Multidimensional; Variação Linguística; Machado de Assis.

***Abstract**:* This study employs a corpus linguistics approach to investigate the style of Machado de Assis's (1839-1908) fictional prose. A literary corpus of his published nine novels and 76 short stories, named CLIMA (859.521 words), was compiled to identify linguistic variation of selected part of speech / grammatical categories, and frequent lemmas. Multi-dimensional analysis (Berber Sardinha; Veirano Pinto 2019) was the main technique employed for identifying latent functional and lexical dimensions that emerged from co-ocurrent variables. A canonical correlation analysis (Mayer 2018) was then carried out in which both dimension sets revealed major relevant associations within and across the group members. Each of the resulting three new aesthetic dimensions comprised a different mix of functional and lexical dimensions, to create a wide-ranging stylistic analysis of Machado de Assis's singular fiction taking lexicogrammar in account.

***Keywords**:* Multi-dimensional Analysis; Linguistic Variation; Machado de Assis.

1. Introdução

Embora a obra de Machado de Assis possua a mais vasta fortuna crítica da literatura brasileira (GUIMARÃES, 2017), na área linguística essa obra é surpreendentemente objeto de poucos estudos. Dois trabalhos conhecidos são Mattoso Câmara Júnior (1962), que realizou uma análise da prosa do escritor com foco na oralidade e no uso do discurso indireto livre, e Ferreira (2007 [c. 1940]), ao explorar, do ponto de vista lexicográfico, expressões coloquiais e brasileirismos presentes no vocabulário machadiano.

No âmbito da linguística de corpus, os estudos existentes priorizaram a pesquisa de palavras-chave, frequência e distribuição vocabular — seja de uma obra isolada, como "Dom Casmurro" (RECSKI, 2005), ou de um corpus abrangente (PRADERA, 2014; FREITAS, 2007), capaz de reunir boa parte da prosa ficcional do escritor. Com semelhante amplitude à desses últimos trabalhos, porém sob uma perspectiva diversa, este estudo tem como objetivo descrever o estilo de Machado de Assis a partir de correlações relevantes existentes entre funções comunicativas e tópicos ou temas que emergem do texto machadiano.

Postula-se aqui que o estilo poderia ser considerado o resultado de escolhas estéticas do falante/escritor no contexto do registro linguístico (BIBER; CONRAD, 2009) no terreno da morfossintaxe e do léxico. Nessa hipótese, do modelo que busca integrar léxico e gramática poderia derivar uma interpretação ao menos básica do discurso individual, de natureza estilística.

Para atingir tal meta, foram identificados dois conjuntos de dimensões de variação linguística, cada qual reunindo variáveis coocorrentes compostas por categorias gramaticais ou pelo léxico mais frequente, resultantes de análises multidimensionais funcionais e lexicais (BERBER SARDINHA; VEIRANO PINTO, 2019) efetuadas sobre o corpus de estudo. Uma análise de correlação canônica (MAYER, 2018; TABATCHNICK; FIDELL, 2013) pôde, por fim, produzir uma interpretação unificada que considera as relações entre ambas as análises multidimensionais precedentes.

2. Dados e instrumentos de análise

No intuito de representar o estilo literário de Machado, foi dada preferência a um recorte particular de sua vasta e variada obra. Denominado CLIMA, o corpus limitou-se apenas aos 76 contos e 9 romances que foram reunidos em vida pelo escritor no formato de livro. Foram gerados dois conjuntos de variáveis: o conjunto de variáveis funcionais com diversas características gramaticais foi etiquetado pelo *parser* PALAVRAS (BICK, 2014), enquanto o conjunto lexical foi realizado pelo *parser* TreeTagger (SCHMID, 2013) na versão para o português, que gerou diversos lemas. Ambos os conjuntos de dados, assim como o corpus CLIMA, estão disponíveis online (em https://osf.io/zktcq/).

A análise multidimensional baseia-se principalmente na análise fatorial. De natureza exploratória, essa técnica estatística possibilita identificar um número pequeno de fatores interpretáveis, ou dimensões, que exercem influência efetiva sobre a variação linguística encontrada no corpus. Assim, cada fator/dimensão define-se a partir da articulação de um grupo próprio de variáveis coocorrentes. Duas análises multidimensionais foram efetuadas separadamente sobre cada conjunto de variáveis do corpus CLIMA, cujos resultados são apresentados a seguir.

É necessário mencionar, porém, que as duas análises multidimensionais foram objeto de uma subsequente análise de correlação canônica, a qual verificou se existiam associações entre os componentes dos dois conjuntos de dimensões. Os pares canônicos gerados pela análise puderam sintetizar sentidos lexicogramaticais do corpus CLIMA de maneira articulada, interpretados em termos de dimensões estéticas capazes de expressar características definidoras do estilo de Machado de Assis.

3. Resultados e discussão

As análises multidimensionais possibilitaram a identificação de cinco dimensões de variação funcional e nove dimensões de variação lexical, descritas resumidamente a seguir – os dados completos podem ser consultados em Kauffmann (2020).

A primeira análise multidimensional, de natureza funcional (EGBERT, 2012), utilizou 29 variáveis iniciais relativas a categorias gramaticais e obteve uma medida de mensuração amostral (KMO) de 0,689, considerada aceitável (TABACHNICK; FIDELL, 2013, p. 619-620). A análise fatorial gerou cinco fatores, com variância total explicada de 15,8%.

A Dimensão F1 foi chamada de Discurso abstrato *versus* Oralidade, uma vez que agrupa, de um lado, um polo de expressão abstrata ligado a nominalizações (carga de 0,71), substantivos abstratos (0,70), particípio passado (0,47), artigos indefinidos (0,38), adjetivos em posição atributiva (0,36) e substantivos em posição de sujeito (0,35), enquanto no polo oposto encontram-se conjunções coordenadas oracionais (-0,73), conjunções coordenadas aditivas (-0,69), marcadores discursivos (-0,58) e verbos de comunicação (-0,36), que se ligam a registros orais. Tal diferença estilística pode ser observada comparando-se os Exemplos 1 e 2, associados aos respectivos polos da dimensão. Assim como nos demais exemplos, as características que compõem os polos das dimensões estão sublinhadas.

Exemplo 1: Discurso abstrato

> Nenhum dos criminosos, ao deixar a prisão, suspeitava o destino científico que o esperava. Saíam um por um; às vezes dois a dois, ou três a três. Muitos deles, estendidos e atados à mesa da operação, não chegavam a desconfiar nada; imaginavam que era um novo gênero de execução sumária. Só quando os anatomistas definiam o objeto do estudo do dia, alçavam os ferros e davam os primeiros talhos, é que os desgraçados adquiriam a consciência da situação. (“Conto alexandrino”, 1884).

Exemplo 2: Oralidade

> — Sim, Adão e Eva.
>
> — Duas belas criaturas que vimos andar há tempos, altas e direitas como palmeiras?
>
> — Justamente.
>
> — Oh! detesto-os. Adão e Eva? Não, não, manda-me a outro lugar. Detesto-os! Só a vista deles faz-me padecer muito. Não hás de querer que lhes faça mal...
>
> — É justamente para isso.
>
> — Deveras? Então vou; farei tudo o que quiseres, meu senhor e pai. ("Adão e Eva", 1896).

A Dimensão F2, por seu turno, foi intitulada Narração em virtude de sua semelhança com a dimensão narrativa descrita em Biber (1988) e Egbert (2012). É composta por terceira pessoa verbal (0,92), pronomes de terceira pessoa em posição de objeto (0,70), pronomes raros em posição de objeto (0,65), pretérito perfeito (0,53), pretérito imperfeito (0,37), verbos mentais (0,33) e verbos de ação (0,31). O Exemplo 3 ilustra o efeito narrativo.

Exemplo 3: Narração

> A corveta dele voltou de uma longa viagem de instrução, e Deolindo veio à terra tão depressa alcançou licença. Os companheiros disseram-lhe, rindo:
>
> — Ah! Venta-Grande! Que noite de almirante vai você passar! ceia, viola e os braços de Genoveva. Colozinho de Genoveva...
>
> Deolindo sorriu. Era assim mesmo, uma noite de almirante, como eles dizem, uma dessas grandes noites de almirante que o esperava em terra. ("Noite de almirante", 1884).

A Dimensão F3, denominada Discurso hipotético devido à presença de verbos sem ação concreta e verbos *dicendi*, combinada a uma ideia de incerteza atribuída aos modais, é formada por verbos «ser» e «estar» (0,85), verbos

de existência ou relação (0,76), adjetivos em posição predicativa (0,58), verbos de comunicação (0,38), verbos modais (0,37) e pronomes demonstrativos (0,35). O Exemplo 4 apresenta uma amostra dessa dimensão.

Exemplo 4: Discurso hipotético

> Todavia era conveniente obter o apoio de Augusta. Vasconcelos pensou em tratar disso o mais cedo que lhe fosse possível.
>
> Entretanto, urgia organizar os seus negócios, e Vasconcelos procurou um advogado a quem entregou todos os papéis e informações, encarregando-o de orientá-lo em todas as necessidades da situação, quais os meios que poderia opor em qualquer caso de reclamação por dívida ou hipoteca. ("O segredo de Augusta", 1870).

A Dimensão F4 foi interpretada como Informação contextual *versus* Referência dependente de situação, composta por um polo de concentração informativa que enfeixa locuções adverbiais (0,64), tempo verbal pretérito imperfeito (0,55), orações reduzidas de infinitivo precedidas de preposição (0,53), verbos de ação (0,36) e adjetivos em posição predicativa (0,35), em oposição a um polo de perfil endofórico e dêitico, onde estão agrupados pronomes possessivos (-0,55), substantivos em posição de sujeito (-0,44), verbos de comunicação (-0,31), marcadores discursivos (-0,31) e pronomes demonstrativos (-0,31). Os Exemplos 5 e 6 refletem, respectivamente, o discurso desses dois polos dimensionais.

Exemplo 5: Informação contextual

> Há meio século, os escravos fugiam com frequência. Eram muitos, e nem todos gostavam da escravidão. Sucedia ocasionalmente apanharem pancada, e nem todos gostavam de apanhar pancada. Grande parte era apenas repreendida; havia alguém de casa que servia de padrinho, e o mesmo dono não era mau; além disso, o sentimento da propriedade moderava a ação, porque dinheiro também dói. A fuga repetia-se, entretanto. ("Pai contra mãe", 1906).

Exemplo 6: Referência dependente de situação

> PROMETEU. — Prometeu é o meu nome.
>
> AHASVERUS. — Tu Prometeu?
>
> PROMETEU. — E qual foi o meu crime? Fiz de lodo e água os primeiros homens, e depois, compadecido, roubei para eles o fogo do céu. Tal foi o meu crime. Júpiter, que então regia o Olimpo, condenou-me ao mais cruel suplício. Anda, sobe comigo a este rochedo.
>
> AHASVERUS. — Contas-me uma fábula. Conheço esse sonho helênico. ("Viver!", 1896).

A Dimensão F5, com também dois polos, foi chamada de Apresentação do pensamento *versus* Descrição elaborada, uma vez que identifica o uso do discurso de personagens e narradores em detrimento do relato da cena narrativa. O polo positivo reúne verbos mentais (0,69), primeira pessoa verbal (0,66), conjunções subordinadas (0,53), advérbio "não" (0,46), pronomes possessivos (0,37), verbos de comunicação (0,35) e pretérito perfeito (0,35), enquanto o polo negativo agrupa adjetivos em posição atributiva (-0,52), artigos indefinidos (-0,39) e artigos definidos (-0,33). Duas amostras de textos representam os polos da dimensão, nos Exemplos 7 e 8.

Exemplo 7: Apresentação do pensamento

> Rita jantou comigo; disse-lhe que estou são como um pêro, e com forças para ir às bodas de prata. Ela, depois de me aconselhar prudência, concordou que, se não tiver mais nada, e for comedido ao jantar, posso ir; tanto mais que os meus olhos terão lá dieta absoluta.
>
> — Creio que Fidélia não vai, explicou. ("Memorial de Aires", 1908).

Exemplo 8: Descrição elaborada

> [Bernardino] particularmente encomendou uma genealogia a um grande doutor dessas matérias, que em pouco mais de uma hora o entroncou a um tal ou qual general romano do século IV, Bernardus Tanoarius; — nome que deu lugar à controvérsia, que ainda dura, querendo uns que o rei Bernardão tivesse sido tanoeiro, e outros que isto não passe de uma confusão deplorável com o nome do fundador da família. ("O dicionário", 1899).

A segunda análise multidimensional realizada no estudo analisou variáveis lexicais, especificamente palavras ou lemas que congregam formas semânticas comuns, mas que apresentam variação de gênero, número, pessoa etc. A realização da análise exigiu a segmentação em capítulos do corpus CLIMA, que elevou o número total de observações a 1.109 textos. Tal configuração permitiu a inclusão de 346 variáveis lexicais iniciais, número considerado mais abrangente para representar o léxico comumente encontrado no corpus machadiano. Listas dos 150 lemas mais frequentes por texto foram geradas e, em seguida, foram selecionadas palavras ou lemas que obtiveram alcance igual ou maior que 25% nos textos capitulados. Embora com uma amostra menos representativa (KMO = 0,459) que na primeira análise multidimensional, chegou-se a nove fatores, com variância total explicada de 21,8%.

A dimensão L1 foi denominada Aparência e sentimentos (Exemplo 9), composta pelos lemas sentimento (0,42), olho (0,36), pouco (0,32), sangue (0,31), ombro (0,30), jantar (0,29), fechar (0,29), braço (0,27), futuro (0,25) e cabeça (0,25).

Exemplo 9: Aparência e sentimentos

> Achou-a sentada na cama, com a cabeça sobre uma almofada, e soluçando. Luís Negreiros ajoelhou-se diante dela e pegou-lhe numa das mãos.
>
> — Clarinha – disse ele –, perdoa-me tudo. Já tenho a explicação do relógio; se teu pai não me fala em vir jantar amanhã, eu não era capaz de adivinhar que o relógio era um presente de anos que tu me fazias.
>
> Não me atrevo a descrever o soberbo gesto de indignação com que a moça se pôs de pé quando ouviu estas palavras do marido. Luís Negreiros olhou para ela sem compreender nada. ("O relógio de ouro", 1873).

A dimensão L2 foi interpretada como Referência romântica (Exemplo 10), composta pelos lemas coração (0,34), saber (0,29), amor (0,29), perder (0,29), amar (0,28) e situação (0,28).

Exemplo 10: Referência romântica

> Sua mãe, que morrera com trinta e oito anos, amou o marido até os últimos dias, e poucos meses lhe sobreviveu. Estêvão soube que fora ardente e entusiástico o amor de seus pais, na estação do noivado, durante a manhã conjugal; conheceu-o assim por tradição; mas na tarde conjugal a que ele assistiu viu o amor calmo, solícito e confiante, cheio de dedicação e respeito... ("A mulher de preto", 1870).

A dimensão lexical L3 foi intitulada Vicissitudes do homem (Exemplo 11) e é composta pelos lemas falar (0,43), vista (0,43), tanto (0,42), noivo (0,40), homem (0,32), volta (0,29), alegre (0,26), vontade (0,26) e antigo (0,26).

Exemplo 11: Vicissitudes do homem

> Mal acabava D. Beatriz de falar, e José Lemos de assentir mentalmente à opinião da mulher, ouviu-se na escada a voz do Tenente Porfírio. O dono da casa soltou um suspiro de

> alívio e satisfação. Entrou na sala o longamente esperado conviva. Pertencia o tenente a essa classe feliz de homens que não têm idade; uns lhe davam 30 anos, outros 35 e outros 40; alguns chegavam até os 45, e tanto esses como os outros podiam ter igualmente razão. ("As bodas de Luís Duarte", 1873).

A dimensão L4 foi chamada de Representação social (Exemplo 12), sendo composta dos lemas nome (0,76), cabeça (0,66), notar (0,47), corte (0,46), mesa (0,45), felicidade (0,38), dizer (0,33), realmente (0,30) e explicar (0,29).

Exemplo 12: Representação social

> Notei que a conversa dele fazia mais efeito no meio da viagem — arejando o espírito e deixando a gente em paz com Deus e os homens; mas devo dizer que o almoço pode ter prejudicado o resto. Realmente era magnífico; e seria impertinência histórica pôr a mesa de Lúculo na casa de Platão. Entre o café e o cognac, disse-me ele, apoiando o cotovelo na borda da mesa, e olhando para o charuto que ardia:
>
> — Na minha viagem agora, achei ocasião de ver como o senhor tem razão com aquela ideia do Brasil engatinhando. ("Evolução", 1906).

A dimensão L5 foi denominada Cena urbana (Exemplo 13) e é composta pelos lemas janela (0,59), trazer (0,42), cidade (0,39), depressa (0,37), decerto (0,34), correr (0,33), longe (0,33), novo (0,31), primeiro (0,30), opinião (0,29) e pronto (0,29).

Exemplo 13: Cena urbana

> — Diga, minha senhora.
>
> — É que nos toque agora aquela sua polca Não Bula Comigo, Nhonhô.
>
> Pestana fez uma careta, mas dissimulou depressa, inclinou-se calado, sem gentileza, e foi para o piano, sem

> entusiasmo. Ouvidos os primeiros compassos, derramou-se pela sala uma alegria nova, os cavalheiros correram às damas, e os pares entraram a saracotear a polca da moda. Da moda; tinha sido publicada vinte dias antes, e já não havia recanto da cidade em que não fosse conhecida. ("Um homem célebre", 1896).

A dimensão L6 foi interpretada como Autoridade patriarcal (Exemplo 14), composta pelos lemas afinal (0,63), noite (0,61), conselheiro (0,60), nunca (0,58), senhor (0,55) e homem (0,27).

Exemplo 14: Autoridade patriarcal

> O negócio ainda não estava composto; o pai ficou furioso e quis quebrar tudo; bradou que não, senhor, que o peralta havia de ir para o seminário, ou então metia-o no Aljube ou na presiganga. João Carneiro lutou muito para conseguir que o compadre não resolvesse logo, que dormisse a noite, e meditasse bem se era conveniente dar à religião um sujeito tão rebelde e vicioso. Explicava na carta que falou assim para melhor ganhar a causa. Não a tinha por ganha; mas no dia seguinte lá iria ver o homem, e teimar de novo. ("O caso da vara", 1899).

A dimensão L7 foi intitulada Sabedoria (Exemplo 15) e é composta pelos lemas entretanto (0,54), calar (0,48), princípio (0,48), abrir (0,47), boca (0,43), trabalho (0,34) descer (0,33) e terra (0,29).

Exemplo 15: Sabedoria

> — Justamente. Conheço agora tudo, a origem das coisas e o enigma da vida. Anda, come e terás um grande poder na terra.
>
> — Não, pérfida!
>
> — Néscia! Para que recusas o resplendor dos tempos? Escuta-me, faze o que te digo, e serás legião, fundarás cidades, e chamar-te-ás Cleópatra, Dido, Semíramis; darás heróis do teu ventre, e serás Cornélia; ouvirás a voz do

> céu, e serás Débora; cantarás e serás Safo. E um dia, se Deus quiser descer à terra, escolherá as tuas entranhas, e chamar-te-ás Maria de Nazaré. ("Adão e Eva", 1896).

A dimensão lexical L8 foi intitulada Metalinguagem (Exemplo 16) e compõe-se dos lemas capítulo (0,60), segundo (0,58), contrário (0,57), sair (0,42), estar (0,41), político (0,33), escrever (0,28) e acabar (0,27).

Exemplo 16: Metalinguagem

> O baile acabou. O capítulo é que não acaba sem que deixe um pouco de espaço a quem quiser pensar naquela criatura. Pai nem mãe podiam entendê-la, os rapazes também não, e provavelmente Santos e Natividade menos que ninguém. Tu, mestra de amores ou aluna deles, tu, que escutas a diversos, concluis que ela era...
>
> Custa pôr o nome do ofício. Se não fosse a obrigação de contar a história com as próprias palavras, preferia calá-lo, mas tu sabes qual é ele, e aqui fica. ("Esaú e Jacó», 1904).

A dimensão L9, denominada Incerteza *versus* Correspondência (Exemplos 17 e 18, respectivamente), foi a única na análise lexical em que ocorreram dois polos de variáveis lexicais. No polo positivo, estão os lemas pequeno (0,33), claro (0,29), mundo (0,27) e talvez (0,25), enquanto no polo negativo encontram-se os lemas carta (-0,30), pronto (-0,30) e maneira (-0,26).

Exemplo 17: Incerteza

> Gostava de falar de todas as coisas, menos de dinheiro, e principalmente de dinheiro achado; todavia não era crime achar dinheiro, era uma felicidade, um bom acaso, era talvez um lance da Providência. Não podia ser outra coisa. Não se perdem cinco contos, como se perde um lenço de tabaco. ("Memórias póstumas de Brás Cubas", 1881).

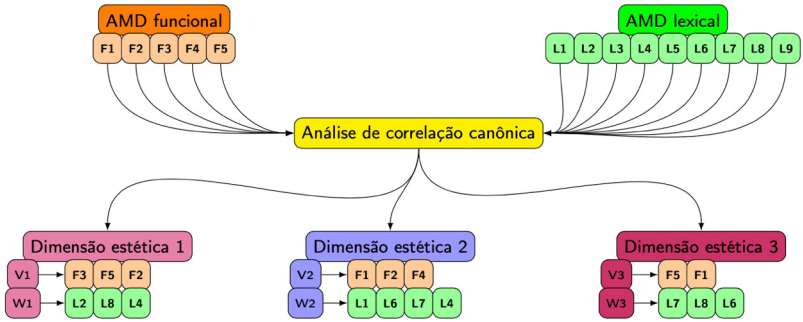
Exemplo 18: Correspondência

> Finalmente raiou aquele em que devia partir o brigue. Helena saiu de seu quarto com os olhos vermelhos de chorar. Interrogada bruscamente pela tia, disse que era uma inflamação adquirida pelo muito que lera na noite anterior. A tia prescreveu-lhe abstenção da leitura e banhos de água de malvas.
>
> Quanto ao tio, tendo chamado Simão, entregou-lhe uma carta para o correspondente, e abraçou-o. A mala e um criado estavam prontos. A despedida foi triste. Os dois pais sempre choraram alguma coisa, a rapariga muito. ("Frei Simão", 1870).

Posteriormente, foi efetuada uma análise de correlação canônica (MAYER, 2018), análise estatística multivariada que avaliou a associação entre as dimensões de um e de outro conjunto, expressa em pares canônicos que guardam independência em relação aos demais pares canônicos da extração. A extração canônica foi efetuada no programa CANCORR do SAS University Edition, tomando como base os escores das dimensões funcional e lexical atribuídos a cada texto do corpus CLIMA.

A partir do *input* das dimensões das análises multidimensionais, cinco pares canônicos foram gerados, dos quais quatro pares foram significativos (F = 3,65), e os três primeiros pares foram interpretados em termos de dimensões estéticas de variação lexicogramatical (KAUFFMANN, 2020; KAUFFMANN; BERBER SARDINHA, 2021), conforme mostra o esquema da Figura 1. Para uma apresentação mais detalhada da análise canônica, v. Kauffmann (2020).

Figura 1: Esquema da análise canônica



Fonte: o autor.

A dimensão estética 1 foi intitulada Romantismo introspectivo formal e é composta pelas dimensões funcionais Discurso hipotético (0,84), Apresentação do pensamento (0,69) e Narração (0,36), em conjunto com as dimensões lexicais de Referência romântica (0,70), Metalinguagem (0,62) e Representação social (0,38). Algumas das obras que obtiveram maiores coeficientes na dimensão Romantismo introspectivo formal pertencem à fase inicial de Machado, com predominante influência romântica, como por exemplo "Ponto de vista" (1873) e "Confissões de uma viúva moça" (1870). No Exemplo 19, abaixo, esta dimensão está representada por traços como primeira pessoa do singular, verbos modais e de existência, pronomes demonstrativos e conjunções subordinativas, entre outros, em associação com o lema coração.

Exemplo 19: Romantismo introspectivo formal

> Dizem que é um bandoleiro dos quatro costados; mas você sabe que eu não creio em bandoleiros. Quando uma pessoa quer, vence o coração mais versátil deste mundo.
>
> O casamento parece que será daqui a dois meses. Irei naturalmente às exéquias, quero dizer às bodas. Pobre Mariquinhas! Lembra-se das nossas tardes no colégio? Ela

> era a mais quieta de todas, e a mais cheia de melancolia. Parece que adivinhava este destino. (“Ponto de vista”, 1873).

A segunda dimensão estética foi rotulada de Narrativa oralizada contextual, composta pelas dimensões funcionais de Oralidade (-0,80), Narração (0,78) e Informação contextual (0,57), acompanhadas das dimensões lexicais de Aparência e sentimentos (0,64), Autoridade patriarcal (0,63) e Sabedoria (0,61), e pela ausência da dimensão L4, relativa à Representação social (-0,30). Os contos “A cartomante” (1896) e “O caso da vara” (1899) obtiveram escores representativos da dimensão Narrativa oralizada contextual, que se distribui mais amplamente pelo corpus. O Exemplo 20, excerto do conto “Uns braços”, apresenta elementos da dimensão tais como o uso de verbos de ação, tempo pretérito, marcas de terceira pessoa, e lemas como cabeça, braço e pouco.

Exemplo 20: Narrativa oralizada contextual

> Saiu da sala, atravessou rasgadamente o corredor e foi até o quarto do mocinho, cuja porta achou escancarada. Dona Severina parou, espiou, deu com ele na rede, dormindo, com o braço para fora e o folheto caído no chão. A cabeça inclinava-se um pouco do lado da porta, deixando ver os olhos fechados, os cabelos revoltos e um grande ar de riso e de beatitude. Dona Severina sentiu bater-lhe o coração com veemência e recuou. Sonhara de noite com ele; pode ser que ele estivesse sonhando com ela. (“Uns braços”, 1896).

Por fim, a dimensão estética 3 foi denominada Representação dramática, sendo composta pelas dimensões funcionais Apresentação do pensamento (0,48) e Oralidade (-0,31), associadas às dimensões lexicais Sabedoria (0,56) e Metalinguagem (0,54), e, ainda, pela ausência de léxico ligado à dimensão de Autoridade patriarcal (-0,39). São exemplos de obras que pontuaram nessa dimensão estética os contos “Umas férias” (1906) e “Viver!” (1896). A dimensão Representação dramática está mais associada com a produção final do escritor e é a que veicula boa parte da ironia do texto machadiano, conforme ilustra o Exemplo 21, que traz ocorrências de verbos mentais, advérbio “não”, conjunções subordinativas e coordenativas, além do lema calar.

Exemplo 21: Representação dramática

> — Meus filhos, vosso pai morreu! [...]
>
> Não se tratava de um dia santo, com a sua folga e recreio, não era festa, não eram as horas breves ou longas, para a gente desfiar em casa, arredada dos castigos da escola. Que essa queda de um sonho tão bonito fizesse crescer a minha dor de filho não é coisa que possa afirmar ou negar; melhor é calar. O pai ali estava defunto, sem pulos, nem danças, nem risadas, nem bandas de música, coisas todas também defuntas. ("Umas férias", 1906).

4. Considerações finais

Este estudo buscou identificar dimensões estéticas de variação lexicogramatical do corpus machadiano sob o prisma das relações que se estabelecem entre suas dimensões funcionais e lexicais, a fim de compor uma síntese conceitual de conjuntos de variáveis lexicais e gramaticais coocorrentes. A análise de correlação canônica promoveu uma redução efetiva do número de variáveis analisadas. Percebeu-se no processo uma camada adicional de complexidade durante a fase de atribuição de rótulos às dimensões estéticas, por existir uma prévia rede de construtos já elaborados na construção das dimensões, seja de natureza funcional ou lexical. A metodologia utilizada conseguiu identificar com algum grau de detalhe aspectos estilísticos de Machado de Assis que poderão contribuir para a compreensão da sua obra ficcional. Por fim, espera-se que tenha sido devidamente destacado o uso da evidência linguística como meio para enriquecer a fruição artística de uma obra singular como a de Machado de Assis.

Referências bibliográficas

Berber sardinha, t.; veirano pinto, m. (Orgs.). *Multi-Dimensional Analysis: Research Methods and Current Issues*. London; New York: Bloomsbury, 2019.

Biber, D.; Conrad, S. *Register, Genre, and Style*. Cambridge: Cambridge University Press, 2009.

Bick, E. PALAVRAS, a constraint grammar-based parsing system for Portuguese. In: Berber Sardinha, T.; São Bento Ferreira, T. *Working with Portuguese corpora*. London; New York: Bloomsbury; Continuum, 2014: 279–302.

Egbert, J. Style in nineteenth century fiction: A Multi-Dimensional analysis. Scientific Study of Literature, v. 2, n. 2, pp. 167–198, jan., 2012.

Ferreira, A. B. de H. *Linguagem e estilo de Machado de Assis, Eça de Queirós e Simões Lopes Neto*. Rio de Janeiro: Academia Brasileira de Letras, 2007 [c. 1940].

Freitas, D. J. T. *A composição do estilo do contista Machado de Assis*, 2007. 211 f. Tese (Doutorado em Literatura) – Programa de Pós-Graduação em Literatura, Universidade Federal de Santa Catarina, Florianópolis, 2007.

Guimarães, H. de S. *Machado de Assis, o escritor que nos lê*: As figuras machadianas através da crítica e das polêmicas. São Paulo: Editora Unesp, 2017.

Kauffmann, C.; Berber Sardinha, T. Brazilian Portuguese literary style. In: Friginal, E.; Hardy, J. A. *The Routledge handbook of corpus approaches to discourse analysis*. New York; London: Routledge, 2021: 354–375.

Kauffmann, C. H. *Linguística de corpus e estilo*: análises multidimensional e canônica na ficção de Machado de Assis, 2020. 276 f. Tese (Doutorado em Linguística Aplicada e Estudos da Linguagem) – Programa de Pós-Graduação em Linguística Aplicada e Estudos da Linguagem, Pontifícia Universidade Católica de São Paulo, São Paulo, 2020. Disponível em: <https://tede2.pucsp.br/bitstream/handle/23128/2/Carlos%20Henrique%20Kauffmann.pdf>. Acesso em: 28 mar. 2022.

Mattoso Câmara Júnior, J. *Ensaios machadianos*: língua e estilo. Rio de Janeiro: Livraria Acadêmica, 1962.

Mayer, C. *O que e como escrevemos na web*: um estudo multidimensional de variação de registro em língua inglesa, 2018. 126 f. Tese (Doutorado em Linguística Aplicada e Estudos da Linguagem) – Programa de Pós-Graduação em Linguística Aplicada e Estudos da Linguagem, Pontifícia Universidade Católica de São Paulo, São Paulo, 2018.

Pradera, L. C. *Machado de Assis*: Uma nova leitura através das lentes do corpus linguístico, 2014. 80 f. Dissertação (Master of Arts) – Departament of Spanish and Portuguese, Brigham Young University, EUA, Provo, 2014. Disponível em: <https://scholarsarchive.byu.edu/etd/5543/>. Acesso em: 23 jan. 2022.

Recski, L. J. Concordâncias, listas de palavras e palavras-chave: o que elas podem nos dizer sobre a linguagem? *Literatura y Lingüística*, v. 16, 2005. Santiago. Disponível em: <http://dx.doi.org/10.4067/S0716-58112005000100014>. Acesso em: 23 jul. 2021.

Schmid, H. Probabilistic part-of-speech tagging using decision trees. New methods in language processing. Anais... . pp. 154, 2013.

Tabachnick, B. G.; Fidell, L. S. *Using multivariate statistics*. 6ª ed. Upper Saddle River, NJ: Pearson Education, 2013.

Relações metafóricas e fraseologia especializada em *corpus* de *Humor Político*

Metaphorical relations and specialized phraseology in a *corpus* of *Political Humour*

Ariel Novodvorski[1]

1 Docente do Instituto de Letras e Linguística da Universidade Federal de Uberlândia.

***Resumo**:* Abordamos neste trabalho a metaforização da política a partir de Unidades Fraseológicas Especializadas (UFEs) do domínio do futebol, num *corpus* jornalístico em espanhol rio-platense, que compilamos da coluna dominical de opinião intitulada *Humor Político*, de Alejandro Borensztein, publicada no jornal argentino *Clarín*. Nosso *corpus* de estudo possui 406 textos, publicados entre 2009 e 2019, com mais de 460 mil palavras. Os pressupostos teóricos contemplam Terminologia, Fraseologia Especializada, Metáfora Conceptual e Linguística Descritiva. A Linguística de Corpus recobre todo o trabalho, tanto pela abordagem quanto pelos procedimentos implicados, oportunizando a percepção dos fatos linguísticos pelo uso de ferramentas aliadas à introspecção e à observação das ocorrências, como um modo diferente de olhar e abordar os dados. Nesse sentido, exploramos seu potencial teorizador com relação à formulação de hipóteses. Aplicamos as ferramentas *WordList, KeyWords* e *Concord*, do programa para análises lexicais *WordSmith Tools* (WST), versão 7,0 (Scott, 2016), em suas diferentes funcionalidades, assim como alguns recursos do *Corpus del Español* (Davies, 2016), utilizado como consulta. Por meio da identificação e análise descritiva das UFEs características do futebol, estabelecemos as relações metafóricas na representação dos emaranhados políticos na Argentina. A partir dos resultados, destacamos a convergência de aspectos cognitivos, linguísticos e pragmáticos, permeados por uma dimensão cultural mais ampla. Do domínio fonte do futebol são transferidas características mais concretas, que passam a ser assimiladas para a compreensão do domínio alvo mais abstrato do campo da política. Tanto a identificação das fraseologias presentes no *corpus* quanto o conhecimento necessário sobre as áreas de especialidade presentes nos textos são fundamentais para a compreensão dos usos especializados de determinadas unidades fraseológicas do âmbito futebolístico no contexto da trama política. Também o contexto sócio-histórico e cultural desempenha um papel decisivo na percepção e entendimento das relações metafóricas linguísticas e conceptuais, dada a referência a fatos que deveriam acionar relações ou lembranças na memória dos leitores, assim como chamar a atenção para aspectos pragmáticos envolvidos, com valores humorísticos subjacentes. Ilustraremos diversos resultados alcançados na pesquisa, com o detalhamento dos procedimentos metodológicos adotados na identificação das UFEs e no estudo interpretativo das relações metafóricas encontradas.

Palavras-chave: Unidades Fraseológicas Especializadas; Metáfora Conceptual; Linguística de *Corpus*; futebol; política.

***Abstract**:* In this paper we address the metaphorization of politics by Specialized Phraseological Units (UFEs) of the soccer domain, in a journalistic *corpus* in the Buenos Aires variety of Spanish which we compiled from the Sunday opinion column entitled *Humor Político*, written by Alejandro Borensztein, published in the Argentine newspaper *Clarín*. Our study *corpus* contains 406 texts, published between 2009 and 2019, with more than 460,000 words. The theoretical assumptions contemplate Terminology, Specialized Phraseology, Conceptual Metaphor, and Descriptive Linguistics. Corpus Linguistics covers the whole work, both for its approach and the procedures involved, providing an opportunity for the perception of linguistic facts using tools allied to introspection and observation of occurrences, as a different way of looking at and approaching the data. In this sense, we explore its theorizing potential relating to hypothesis formulation. We used the tools WordList, KeyWords, and Concord, of the program for lexical analysis *WordSmith Tools* (WST), version 7.0 (Scott, 2016), in their different functionalities, as well as some resources of the *Corpus del Español* (Davies, 2016), used as a reference. Through the identification and descriptive analysis of the UFEs characteristic of soccer, we establish the metaphorical relations in the representation of political entanglements in Argentina. From the results, we highlight the convergence of cognitive, linguistic, and pragmatic aspects, permeated by a broader cultural dimension. From the source domain of soccer, more concrete characteristics are transferred, which come to be assimilated for the understanding of the more abstract target domain of the field of politics. Both the identification of the phraseologies present in the *corpus* and the necessary knowledge about the specialized areas present in the texts are fundamental for the understanding of the uses of certain phraseological units of the soccer field in the context of the political plot. Also, the socio-historical and cultural context play a decisive role in the perception and understanding of the linguistic and conceptual metaphorical relations, given the reference to facts that should trigger relations or memories in the readers' memory, as well as draw attention to pragmatic aspects involved, with underlying humorous values. We will illustrate several results achieved in the research, detailing the methodological procedures adopted in the identification of the UFEs and in the interpretative study of the metaphorical relations found.

***Keywords**:* Specialized Phraseological Units; Conceptual Metaphor; Corpus Linguistics; Football; Politics.

1. Introdução

Este trabalho deriva de nossa pesquisa de pós-doutoramento (UFRGS, 2020), realizada sob a supervisão da Profa. Dra. Cleci Bevilacqua, com vínculo junto ao Projeto Terminológico Cone Sul – Termisul. Abordamos a metaforização da política a partir de Unidades Fraseológicas Especializadas (UFEs) do campo do futebol, num *corpus* jornalístico composto por artigos escritos em espanhol rio-platense por Alejandro Borensztein. O *corpus*, com mais de 400 textos cobrindo um período de 10 anos de publicações (2010-2019), foi compilado da coluna dominical de opinião *Humor Político*, publicada no jornal argentino *Clarín*.

As seguintes áreas englobam o quadro teórico deste trabalho: Terminologia, Fraseologia Especializada e Metáfora Conceptual. Desde uma perspectiva descritivista da linguagem, nossa metodologia contempla a abordagem, os procedimentos e a utilização das ferramentas do programa computacional para análises lexicais *WordSmith Tools* (WST), versão 7,0 (Scott, 2016), em suas diferentes funcionalidades: *WordList, KeyWords* e *Concord*. Também ampliamos a pesquisa por meio de consultas ao *Corpus del Español* (Davies, 2016), em sua versão dialetal.

Desse modo, partindo da identificação e análise descritiva das UFEs características do futebol presentes no *corpus* de estudo, estabelecemos relações metafóricas na representação dos entramados políticos na Argentina. A hipótese que sustenta o trabalho é que aspectos cognitivos, linguísticos e pragmáticos estão materializados nos textos, realizando a dimensão cultural mais ampla que os engloba. Nesse sentido, são transferidas características mais concretas do domínio fonte do futebol, que passam a ser assimiladas para a compreensão do domínio alvo mais abstrato do âmbito da política.

A partir das observações anteriores, formulamos a seguinte indagação: identificados os candidatos a termos e as UFEs, como ocorre a metaforização entre duas áreas de conhecimento, a política e o futebol, num *corpus* de textos de opinião? Nossa hipótese, pela perspectiva da comunicação, é que o autor dos textos promove a representação de um domínio por meio das características do outro, proporcionando uma percepção mais acessível das questões

políticas. O objetivo específico é, portanto, refletir acerca dos aspectos relacionados à metaforização entre os domínios da política e do futebol, temáticas presentes no *corpus* de estudo.

Com o seguinte fragmento, ilustramos os procedimentos implicados na identificação, análise e descrição de candidatos a termo e a UFE, na metaforização da política pelo domínio do futebol: "Luis Juez, sobre los dichos de Alberto Fernández: 'Hay que ser nabo para dejársela picando a Bolsonaro'". Em destaque, temos ***picando***, gerúndio de ***picar*** (quicar uma bola, dentre outras acepções como coçar, correr, gerar dúvida...), precedido pelo verbo ***dejar*** (deixar) no infinitivo, acompanhado de dois pronomes complemento átonos em posição enclítica, ***dejársela***. Trata-se de duas orações: na primeira, alguém (Alberto Fernández, presidente da Argentina) deixa algo (referido pelo pronome feminino singular ***la***) para alguém (referido pelo pronome ***se***, Bolsonaro, presidente do Brasil); na segunda, algo, aquilo que foi deixado (referido pelo pronome ***la***) está ***picando*** (quicando).

Pelo conhecimento da área do futebol, inferimos que a referência corresponde a uma bola (***pelota***, substituída pelo pronome ***la***). A frase ***dejársela picando*** alude a uma jogada de futebol, em que um jogador deixa a bola quicando, para que seu companheiro de time receba a bola em condições propícias para chutar e fazer o gol, o que corresponderia a uma ***bola servida de bandeja***. Em consulta ao *Corpus del Español* (Davies, 2016; 2018) em sua versão dialetal, com ***picando*** como item de busca e ***DEJAR*** (lematizado, com todas as flexões) como colocado à esquerda, obtivemos 165 resultados, dos quais 79 correspondem à Argentina, sendo México o segundo país mais frequente, com 12 ocorrências.

Os resultados encontrados atestam pela frequência que se trata de uma UFE própria da variante argentina, em correspondência com a jogada de futebol mencionada acima, tanto no sentido literal do jogo quanto metafórico, em alusão ao mundo da política. Como exemplos, podemos citar: "Pero el equipo de investigadores prefiere dejar la pelota picando", "dejo la pelota picando, se les viene una lluvia de juicios", "te la dejo picando para que investigues" e "No menos curiosas resultaron algunas frases presidenciales que dejaron

picando sus sentimientos personales". Como se observa no último exemplo, o complemento do verbo não é uma ***pelota***, mas os ***sentimientos personales*** são referidos como se fossem uma bola que deixaram quicando.

Assim, podemos concluir que a frase irônica proferida pelo deputado argentino funciona textualmente como metáfora linguística, apontando para a metáfora mais genérica POLÍTICA É FUTEBOL. A crítica ao presidente de seu país reside em que, embora sejam países cuja rivalidade no futebol é de conhecimento geral, Fernández teria atuado como um jogador do mesmo time do presidente brasileiro, a partir das frases polêmicas em que comparou o México, o Brasil e a Argentina[2]. Essas frases, proferidas em discurso público, provocaram inúmeros memes e manifestações. Metaforicamente, constatamos que os presidentes são representados como jogadores; a UFE revela a subjacência da metáfora conceptual POLÍTICOS SÃO JOGADORES DE FUTEBOL.

A percepção tanto de candidatos a termo e a eventuais fraseologias quanto de domínios que concorrem nos textos é fundamental, por um lado, para a compreensão leitora dos usos especializados de determinadas UFEs do campo futebolístico no âmbito da política. Por outro lado, o conhecimento sócio-histórico e cultural é também essencial para a compreensão que se alcança, por meio da extração de dados com suporte das ferramentas da LC, aliado a uma análise introspectiva, em que estão envolvidos aspectos factuais, pragmáticos, valores irônicos subjacentes, implicados na construção dos sentidos. Nesta publicação, apresentamos os procedimentos para a identificação das UFEs e para a interpretação de algumas relações metafóricas encontradas, a partir de diversos exemplos extraídos do *corpus*.

---

2 Fernández teria citado, por engano, o poeta e ensaísta mexicano Octavio Paz, afirmando "os argentinos viemos dos barcos, enquanto os mexicanos vieram dos índios e os brasileiros vieram da selva". Disponível em: https://www.perfil.com/noticias/politica/luis-juez-sobre-los-dichos-de-alberto-fernandez-hay-que-ser-nabo-para-dejarsela-picando-a-bolsonaro.phtml. Acesso em: 27 set. 2022.

2. Quadro teórico

O caráter de especialização de um dado texto é conferido pela instância comunicativa. A partir dessa perspectiva, atividades como o esporte, dentre outros, também produzem tipos de textos que se distinguem dos textos considerados gerais. Isto é, os âmbitos especializados não seriam exclusividade das matérias científicas ou técnicas, conforme Cabré (2005). Com isso, lexicalmente, os textos se especializam pela especificidade da terminologia e da fraseologia empregadas, fundamentalmente pela função semântica das unidades terminológicas que os integram.

Cabré (1993) aponta que as linguagens especializadas se caracterizam em função da temática, dos falantes e das situações comunicativas. Complementa afirmando que também fazem parte da especialização áreas como o comércio e o esporte. Por outro lado, com base na autora, corroboramos que os aspectos pragmáticos, se comparados aos aspectos linguísticos ou funcionais, são os que possibilitam uma distinção mais nítida entre os usos mais gerais ou especializados da linguagem, pelo fato de possibilitarem diferenciar termos de palavras. A distinção entre termos e palavras ocorre, portanto, levando em consideração aspectos pragmáticos, a saber: pelos usuários, pelas situações de uso, pela temática e pelo tipo de discurso em que costumam aparecer.

Para o reconhecimento das UFEs no alcance do *corpus* textual da pesquisa, o núcleo terminológico "deveria ser uma unidade nominal que corresponda a um nó cognitivo do âmbito tratado", conforme Bevilacqua (2004, p. 43). A relevância dessa consideração aponta para a possibilidade de tratamento das UFEs como unidades de representação e transmissão de conhecimento especializado. Considerando a possibilidade de investigar o conhecimento especializado em instâncias comunicativas, entendemos que o reconhecimento de unidades terminológicas e UFEs no *corpus*, em especial do domínio mais concreto, seja um dos caminhos viáveis para a análise de casos de metaforização. As características transferidas desse domínio fonte são assimiladas para a compreensão do domínio alvo, que é metaforizado.

Um dos princípios básicos na área da metáfora conceptual é que metáforas linguísticas realizam metáforas conceptuais (Deignan, 2005). Isto é, as relações metafóricas subjacentes são materializadas no plano textual. A identificação das metáforas linguísticas possibilita o alcance interpretativo das metáforas conceptuais que, de fato, são as que licenciam as diferentes ocorrências que funcionam como indícios metafóricos. Assim, uma metáfora linguística veicula o significado literal do domínio fonte (concreto), que passa a significar no domínio alvo (abstrato) como tópico.

Nesse ponto, é pertinente ressaltar o importante papel que cumpre o uso das ferramentas da LC, como aliadas na inferência do processamento mental metafórico, a partir de análises empíricas de instâncias concretas de uso em *corpus*. A identificação de metáforas linguísticas remete a uma análise introspectiva quanto à existência de metáforas conceptuais no plano cognitivo. Conforme Berber Sardinha (2007; 2009), esse caminho tem se tornado promissor, como um relevante desafio e uma especial oportunidade quanto ao potencial teorizador da LC.

Desse modo, para a compreensão e análise metafórica dos fragmentos "Hay que ser nabo para dejársela picando a Bolsonaro", da Introdução deste trabalho, ou de "Se la dejaron abajo del arco y la tiró por arriba del travesaño", como apontado em Novodvorski e Bevilacqua (2021), concluímos que ***dejársela picando***, ***dejársela abajo*** e ***tirarla por arriba*** são verbos sintagmáticos que formam frasemas (Pamies, 2019), haja vista que cumprem as condições necessárias para serem considerados UFs: estão formados por mais de um lexema, ainda que reunidos sob uma forma única, passível de separabilidade (***dejársela, tirarla***); apresentam fixação atestada pela frequência de uso; e idiomaticidade, pois o sentido resultante não decorre da somatória dos significados de cada uma das partes. Registram caráter especializado, portanto UFEs, uma vez que o núcleo se faz termo, no contexto de uso, ainda que esteja na forma pronominal (***dejársela, tirarla*** > a bola).

Por outro lado, funcionam como veículos da metáfora linguística, por fazerem parte do domínio fonte do futebol, especificamente uma bola (***pelota***) que é deixada quicando, na porta do gol (***abajo del arco***) ou chutada por cima do travessão (***la tiró por arriba del travesaño***). Já os elementos que funcionam como tópico, por estarem no campo do domínio alvo da política, fazem

referência aos nomes dos presidentes ou candidatos à presidência envolvidos nas jogadas ilustradas pelos fraseologismos, que realizam as metáforas conceptuais subjacentes: POLÍTICA É FUTEBOL, POLÍTICOS SÃO JOGADORES DE FUTEBOL. A seguir, descrevemos o *corpus* e os procedimentos metodológicos pertinentes à pesquisa.

3. Corpus e Metodologia

Compõem o *corpus* deste trabalho textos publicados na coluna de opinião *Humor Político*, do jornal argentino *Clarín*, que foram escritos em espanhol rio-platense pelo colunista Alejandro Borensztein. Quanto à tipologia do *corpus*, destacamos: (a) modo escrito, em formato eletrônico; (b) contemporâneo e diacrônico, de 2009 a 2019; (c) de seleção definida pelo gênero (seção de opinião: humor político) e extensão mensurada em textos (406), com 466.800 itens ou palavras totais (*tokens*) e 31.759 formas diferentes (*types*); (d) de conteúdo especializado, marcado pelo campo socioprofissional (política); (e) monolíngue, na língua espanhola, em sua variedade rio-platense; (f) de autoria única, em língua nativa; e (g) para finalidade de pesquisa. A razão forma/item (*type/token ratio*) corresponde a 6,8% de formas diferentes, com relação ao total de itens registrados no *corpus*, segundo os dados estatísticos reportados pela ferramenta *WordList*.

Resumidamente, alguns dos procedimentos implicados na pesquisa foram: (1) planejamento do *corpus*; (2) compilação, preparação e armazenamento do *corpus*; (3) tratamento do *corpus* com as ferramentas e utilitários do programa *WST*; (4) extração de listas de palavras e listas de palavras-chave do *corpus* de estudo, relacionadas aos diferentes campos semânticos do futebol e da política, utilizando como ponto de contraste um *corpus* de referência de escrita acadêmica, cuja extensão é de 813 textos, 2.834.385 itens e 95.649 formas, reunindo as publicações de seis Congressos internacionais da língua espanhola, realizados entre 1992 e 2010 (Alves, 2013); (5) identificação e extração dos candidatos a termo, no domínio do futebol; (6) identificação, extração e descrição das UFEs metafóricas do âmbito do futebol, com relação

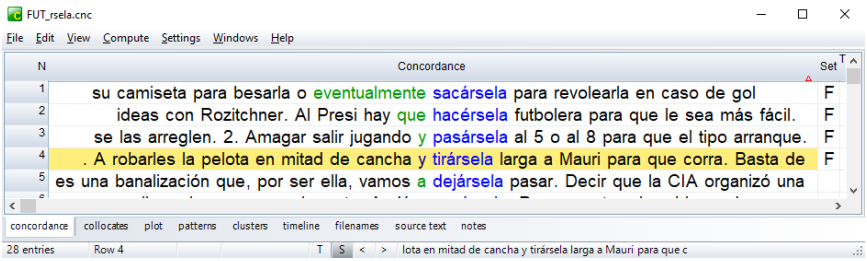
ao domínio da política, a partir da análise das linhas de concordância; e (7) análise e descrição das relações metafóricas, à luz das UFEs extraídas, entre os domínios do futebol e da política.

Considerando as limitações de espaço, priorizaremos, na próxima seção, a análise de algumas UFEs que selecionamos, marcadas especificamente pela forma de verbos sintagmáticos e de advérbios de lugar, tais como os apresentados na Introdução e no Quadro teórico deste trabalho. Serão prioridade, nesse sentido, UFEs formadas por estruturas verbais com pronomes em posição proclítica ou enclítica, seguidas de advérbios de lugar.

4. Análises

Com o propósito de capturar as ocorrências de candidatos a UFEs, para posterior análise de eventuais metáforas, processamos o *corpus* de estudo dos 406 textos por meio da ferramenta *Concord*. Seguindo a estrutura dos exemplos apresentados nas seções anteriores, realizamos os seguintes tipos de buscas: ***rla***, ***rsela*** (para as formas enclíticas) e ***se la*** (para obter as formas pronominais proclíticas). Optamos por esse caminho com o objetivo de delimitar os resultados, uma vez que a busca pelos advérbios de lugar (***abajo, arriba, adelante*** etc.), a partir da lista de palavras-chave, obrigaria a uma seleção acentuada dos resultados a ser apresentados, devido à extensão dos dados. A figura, a seguir, ilustra a identificação e extração das UFEs.

Figura 1: Linhas de concordância com ***rsela** no *corpus*



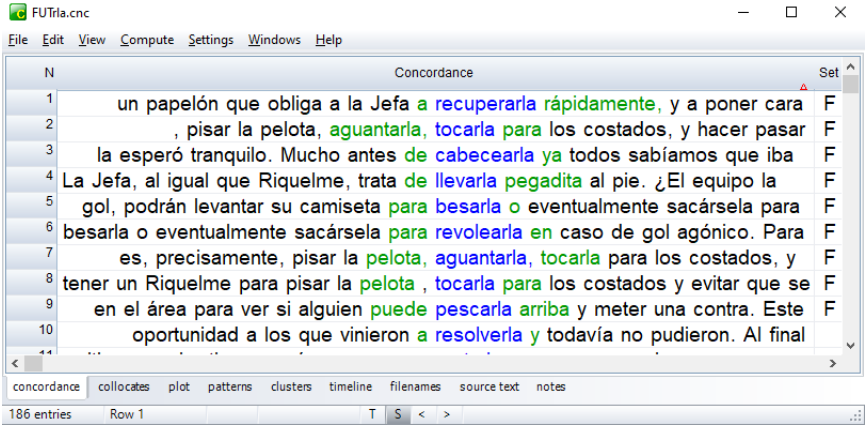
Fonte: Dados da pesquisa no *WordSmith Tools*, versão 7.0 (Scott, 2016)

A ferramenta *Concord* reportou 28 ocorrências, a partir da busca por ***rsela**, todas correspondentes à estrutura de verbos no infinitivo, seguidos de dois pronomes em posição enclítica, cumprindo a função sintática de objeto indireto e direto, respectivamente. As quatro primeiras linhas de concordância, como se aprecia na Figura 1, estão caracterizadas por sua pertinência ao domínio do futebol, a saber: (1) ***sacársela*** faz referência a um jogador tirar sua camisa, após ter feito um gol; (2) ***hacérsela futbolera*** se refere a tornar mais futebolística uma explicação ou ideia para o presidente, como um modo de facilitar sua compreensão; (3) ***pasársela*** corresponde a um jogador passar ou tocar a bola para seus companheiros de time (camisa 5 ou 8); e (4) ***tirársela larga*** remete a um passe da bola longo ou em profundidade para que outro corra, nesse caso, ***Mauri*** (ex-presidente argentino Macri).

Com exceção de (2), constatamos a referência a jogadas específicas de futebol (***sacársela*** [***la camiseta***] / tirar a camisa, ***pasársela*** [***la pelota***] / tocar a bola e ***tirársela larga*** [***una pelota***] / jogar uma bola longa, em profundidade), em que os jogadores são os políticos argentinos. Ainda que (2) não remeta a uma jogada, o domínio fonte é o futebol, uma vez que, para torná-la mais acessível, a explicação para o presidente deveria ser em termos de futebol. A consulta pelos diferentes verbos sintagmáticos, na versão dialetal do *Corpus del Español* (Davies, 2016), corroborou, pela frequência, sua utilização no âmbito do futebol.

Na busca por verbos no infinitivo, seguidos agora de apenas um pronome em posição enclítica, com função sintática de objeto direto, utilizamos a forma ***rla*** e obtivemos 186 resultados, dos quais 9 corresponderam ao domínio do futebol. A próxima figura ilustra os resultados.

Figura 2: Linhas de concordância com ***rla*** no *corpus*

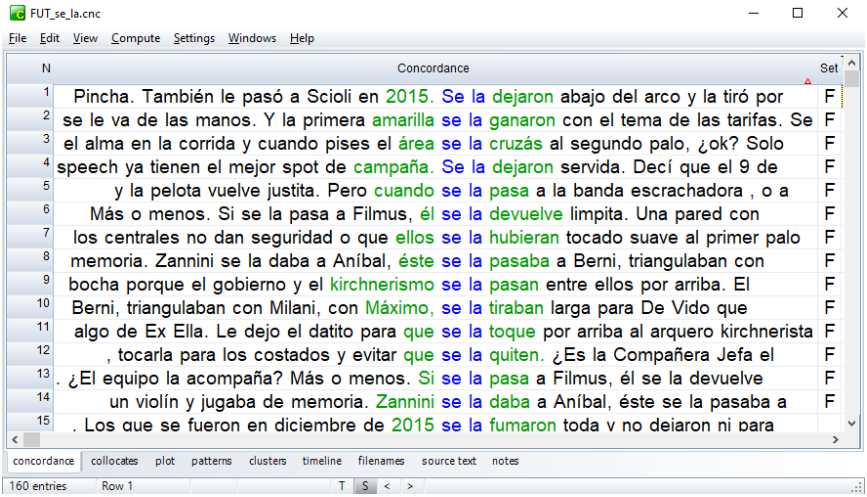


Fonte: Dados da pesquisa no *WordSmith Tools*, versão 7.0 (Scott, 2016)

O pronome ***la***, em posição enclítica, substitui o nome ***pelota*** (bola) em todas as linhas, menos em (5) e (6), em que a referência é à camisa de um jogador, no sentido de beijá-la (***besarla***) e agitá-la (***revolearla***), na comemoração de um gol. As demais linhas apresentam UFEs que registram jogadas de futebol, realizadas por políticos ou governantes, como se fossem jogadores. Assim, em (4) "La Jefa, al igual que Riquelme, trata de llevarla pegadita al pie", temos que ***Jefa*** (a chefe, por Cristina Kirchner, presidente da Argentina na época) é comparada explicitamente ao jogador de futebol Riquelme. As UFEs ***llevarla*** (conduzir a bola) e ***pegadita al pie*** (coladinha no pé) fazem alusão à maneira de conduzir o governo, de ter o controle da condução, assim como o jogador quanto ao domínio da bola. Também em (7) com ***aguantarla*** ou em (9) com ***pescarla***, observamos que se trata do controle da bola, primeiro, no sentido de segurá-la, de tocá-la para os lados, e depois de conseguir recuperar a bola no alto, de cabeça (***pescarla arriba***). Todas essas UFEs, que denotam

jogadas de futebol, funcionam como metáforas linguísticas, licenciadas pela metáfora conceptual POLÍTICA É FUTEBOL. A seguir, apresentamos a busca pelos pronomes em posição proclítica.

Figura 3: Linhas de concordância com ***se la*** no *corpus*



Fonte: Dados da pesquisa no *WordSmith Tools*, versão 7.0 (Scott, 2016)

Das 160 entradas registradas pela ferramenta *Concord* a partir da busca por ***se la***, 14 correspondem ao domínio do futebol como metaforização da política. O pronome ***la*** apenas não corresponde a ***pelota*** na linha (2), referindo-se, nesse caso, a cartão amarelo (***tarjeta amarilla***), por isso ***se la ganaron*** (receberam o cartão). Dentre as ações, temos as UFEs relacionadas a 'tocar a bola para outro jogador' em ***pasársela, dársela, devolvérsela*** (***se la pasa, se la pasan, se la pasaba, se la daba, se la devuelve***). A UFE presente em (9) "el gobierno y el kirchnerismo se la pasan entre ellos por arriba" corresponde à jogada em que os jogadores de um time (oficialismo na política) tocam a bola por cima, evitando que o time rival (oposição) consiga se apoderar da bola.

Outras frases como ***se la dejaron abajo del arco***, ***se la dejaron servida***, ***se la devuelve limpita*** denotam jogadas em que um jogador facilita um lance para que seu companheiro resolva em gol, com a bola na porta do gol, servida de bandeja ou limpinha para chutar. Já em ***se la cruzás al segundo palo*** (chuta cruzado na segunda trave), em ***se la hubieran tocado suave al primer palo*** (teriam dado um toque de leve na primeira trave) e em ***se la toque por arriba al arquero kirchnerista*** (chute por cima do goleiro) temos jogadas de definição com chutes a gol. Por sua vez, em ***se la tiraban larga para De Vido*** (chutavam uma bola longa para alguém) e em ***se la quiten*** (tirem a bola de alguém) percebemos primeiro um passe longo, depois uma roubada de bola. Em síntese, todo um conjunto de jogadas de ataque e defesa, como estratégias futebolísticas, que são metaforizadas para ilustrar o jogo político. A seguir, alguns apontamentos que decorrem desta análise.

5. Considerações

Neste trabalho, fizemos uma análise derivada dos resultados obtidos em nossa pesquisa de pós-doutoramento. Por meio de buscas específicas, com auxílio da ferramenta *Concord*, indagamos nosso *corpus* de estudo quanto a ocorrências de verbos no infinitivo, com pronomes em posição enclítica (****rsela***, ****rla***), além da sequência de pronomes átonos em posição proclítica (****se la***), que também exploramos a partir dos dados alcançados. Os procedimentos implicados viabilizaram a identificação, a análise e descrição dos itens que consideramos, inicialmente, candidatos a termos e a UFEs, com o propósito de alcançar o nível interpretativo da metáfora conceptual.

Pela perspectiva terminológica, portanto, foi necessário perceber os itens lexicais que se configurariam candidatos a termos, inclusive quando estivessem na forma de pronomes complemento átonos. Por meio das análises de linhas de concordância, de ocorrências em contexto, conseguimos determinar o caráter especializado das ocorrências e identificar as UFEs, algumas das quais mostraram a implicância de valores de espacialidade. Já na perspectiva fraseológica, por sua vez, as combinatórias analisadas atenderam às

condições classificatórias das fraseologias, seja pela presença de mais de um lexema (ainda que reunidos sob uma forma única, em formações com infinitivos e pronomes enclíticos, passíveis de separabilidade), pelo grau de fixação observado, assim como de uma maior ou menor idiomaticidade.

Na perspectiva metafórica, a identificação de metáforas linguísticas, entre os domínios fonte, mais concreto, do futebol e alvo, mais abstrato, da política, corroborou a ideia de serem licenciadas por metáforas conceptuais subjacentes. Isso revela uma estrutura mais profunda de pensamento, em que POLÍTICA É FUTEBOL, POLÍTICOS SÃO JOGADORES e ESTRATÉGIAS POLÍTICAS SÃO JOGADAS DE FUTEBOL. Quanto à perspectiva da Linguística de Corpus, destacamos fortemente que se trata de muito mais do que uma sequência de procedimentos metodológicos ou de um modo de olhar para os dados. Na identificação dos fatos linguísticos, a LC funciona como uma verdadeira aliada na observação da estrutura aparente do *corpus*, atiçando nossa introspecção para a percepção das ocorrências metafóricas.

Por outro lado, a partir da exploração e indagação de *corpora*, sempre é importante reforçar que o mapeamento da subjacência das metáforas conceptuais, como fenômeno cognitivo materializado em metáforas linguísticas, é um caminho significativo que oferece indícios suficientes acerca do poder de hipotetização e de teorização sobre a linguagem. Os trabalhos empíricos da LC somam recursos e propiciam o fortalecimento de pesquisas em metáfora conceptual e em Linguística Cognitiva.

Referências bibliográficas

Alves, M. *A representação do Brasil no ensino de espanhol: um estudo diacrônico baseado em corpus de textos acadêmicos*. 2013. Relatório (Iniciação Científica) – Instituto de Letras e Linguística da Universidade Federal de Uberlândia, Uberlândia, 2013.

Berber Sardinha, T. *Pesquisa em Lingüística de Corpus com WordSmith Tools.* Campinas, SP: Mercado das Letras, 2009.

Berber Sardinha, T. *Metáfora*. São Paulo: Parábola Editorial, 2007.

Bevilacqua, C. R. *Unidades Fraseológicas Especializadas Eventivas: descripción y reglas de formación en el ámbito de la energía solar.* Tesis doctoral. Orientadora: Maria Teresa Cabré. Barcelona: Instituto Universitario de Lingüística Aplicada, Universidad Pompeu Fabra, 2004.

Cabré, M. T. *La Terminología: Representación y Comunicación.* Barcelona: IULA / Universitat Pompeu Fabra, 2005.

Cabré, M. T. *La terminología: teoría, metodología, aplicaciones*. Traducción castellana: Carles Tebé. Barcelona: Editorial Empúries, 1993.

Davies, M. *Corpus del Español: Web/Dialectos*, 2016. Disponível em: https://www.corpusdelespanol.org/web-dial/. Acesso em: 03 mar. 2022.

Deignan, A. *Metaphor and Corpus Linguistics.* Amsterdam-Philadelphia: John Benjamins Publishing Company, 2005.

Novodvorski, A.; Bevilacqua, C. De 'marcar la cancha' a una 'canchereada' na metaforização da política pelo futebol: análise de unidades fraseológicas especializadas em corpus jornalístico argentino. *Revista de Estudos da Linguagem*, 29(2), 2021, 1191-1228. http://dx.doi.org/10.17851/2237-2083.29.2.1191-1228.

Novodvorski, A. *Unidades fraseológicas especializadas na metaforização da política pelo futebol: uma descrição guiada por corpus jornalístico de língua espanhola.* Relatório final (Pós-doutorado em Letras). Porto Alegre: Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul – UFRGS, 2020.

Pamies, A. B. El verbo sintagmático en las lenguas románicas. In: Briz, A. et al. (eds.). *Estudios lingüísticos en homenaje a Emilio Ridruejo.* Valencia: Universidad. Vol II: 2019, 1057-1070.

Scott, M. *WordSmith Tools* (7.0) [Programa computacional]. Liverpool: Lexical Analysis Software, 2016.

The C-ORAL-BRASIL corpora as a source of spoken language lexical information

Os corpora C-ORAL-BRASIL como uma fonte de informação lexical da fala

Guilherme Nunes[1]
Heliana Mello [2]

1 DCIDA, Localiza
2 UFMG

Abstract: In this paper we report a first round of lexicon-based analysis of the C-ORAL-BRASIL spontaneous speech corpora. The C-ORAL-BRASIL corpora currently comprise a corpus of informal spontaneous speech (C-ORAL-BRASIL I) along with media, telephonic and formal speech corpora (C-ORAL-BRASIL II) portraying Brazilian Portuguese (BP). BP lexical corpora-based studies have been carried out focusing mostly on written language. Spoken language, given the effort it takes to be properly documented, has received scarce attention. Through our study, we intend to start addressing this gap by providing an overview of results obtained through a series of methodologies that are geared towards profiling the lexical repertoire and variation found in spoken BP. This study is part of a larger project that intends to provide a detailed mapping of lexical phenomena and a basic vocabulary of spoken BP. The corpora .txt files were initially treated through Python coding so that all levels of linguistic and metadata annotation could be removed but were kept accessible for different types of analyses in which text genre, informational and speaker metadata are relevant. Individual word forms were lemmatized and PoS annotated. The quantitative measures presented in this paper were obtained through Python coding in addition to visualization tools available through SketchEngine (Kilgarriff et al., 2014) and Voyant Tools (Sinclair & Rockwell 2016).

Keywords: C-ORAL-BRASIL; Lexicon; Spoken language; Brazilian Portuguese.

Resumo: Neste artigo reportamos os primeiros resultados de uma análise lexical baseada nos corpora de fala espontânea C-ORAL-BRASIL. Os corpora C-ORAL-BRASIL abarcam um corpus de fala espontânea informal (C-ORAL-BRASIL I), bem como corpora de fala midiática, fala espontânea formal e fala telefônica (C-ORAL-BRASIL II) do português brasileiro (PB). A língua falada, dado o custo para ser adequadamente documentada, ainda tem recebido pouca atenção. Através do nosso estudo, objetivamos iniciar o caminho para suprir essa lacuna, oferecendo resultados da aplicação de uma série de metodologias que buscam perfilar o repertório e variação lexicais do PB falado. Este estudo é parte de um projeto mais amplo que visa oferecer como resultado um mapeamento detalhado de fenômenos lexicais além de um vocabulário básico do PB falado. O tratamento metodológico deste estudo incluiu pré-processamento, a criação de um dataframe e a visualização dos elementos lexicais de relevância via codificação em Python e ferramentas oferecidas via SketchEngine (Kilgarriff et al., 2014) e Voyant Tools (Sinclair & Rockwell 2016).

Palavras-chave: C-ORAL-BRASIL; Léxico; Língua falada; Português brasileiro.

1. Introduction

Studies that account for the so-called basic lexicon of a given language are usually carried out taking written corpora as their source of data (Nation 2001). The first estimate for the basic spoken lexicon of English was offered more than half a century ago (Schonell et al. 1956) and found 2,000 lexical families as a baseline. More recent studies suggest that this figure should be increased, based on a CANCODE sample of 497,658 tokens (Adolphs & Schmitt 2003). Brazilian Portuguese (BP), on the other hand, has not yet been the subject of a corpus-based study of its basic lexicon. As part of a project that aims to make available the basic spoken lexicon of BP, this paper introduces the C-ORAL-BRASIL corpora, their treatment for the extraction of lexical information, the tools employed to achieve lexical information visualization, as well as some initial results that point towards the usage patterns of the most frequent lexical items found in the source corpora.

2. Methodology

The C-ORAL-BRASIL corpora feature: (a) an informal speech corpus, divided into public and family/private contexts, covering monologues, dialogues and conversations (Raso & Mello 2012); (b) a formal speech corpus, collected in natural contexts; (c) a media corpus covering radio and TV shows; and (d) a telephonic corpus encompassing both private and public calls. The corpora are constituted as follows: the informal corpus features 139 recording sessions (208,130 words); the formal in natural context corpus features 74 recording sessions (119,807 words); the media corpus features 101 recordings (139,647 words); as for the telephonic corpus, its private section includes 50 recording sessions (25,533 words), while the public section features 29 recording sessions (5,755 words).

Our research workflow followed the following steps: assessment of transcription and prosodic segmentation conventions in the corpora .txt files, assessment of parsed .XML files, importation of annotated texts into the Python environment, data manipulation for pre-processing, creation of a dataframe, and analysis through Python coding and visualization tools

SketchEngine (Kilgarriff et al. 2014) and Voyant Tools (Sinclair & Rockwell 2016). We generated frequency lists taking into account each individual corpus (for individuation of each corpus lexical profile) as well as the joint set of corpora (for spoken BP overall results) for word forms, lemmas and PoS classes and looked into collocates (through the typicality measure LogDice), and different sized ngrams.

3. Results

The main measures obtained through Python coding, as far as word distribution is concerned, are: Overall number of tokens: 498,712; Average word distribution per file: 1,190.45; per speaker: 326.94; per turn: 14.4; per utterance: 7.42; per information unit: 3.26. The token number in the C-ORAL-BRASIL corpora perfectly matches that utilized in the English study based on CANCODE (Adolphs & Schmitt 2003). This allows for a comparative study between BP and English basic lexicons, making it feasible and useful for future computational applications.

The most frequent words in the joint corpora are functional words, as expected. However their distribution is particular to each corpus, which gives us hints about interactional typologies. Their frequency per corpus is as follows: Informal speech: ***nũ*** (2,708); ***né*** (2,199); ***tá*** (2,059); ***cê*** (1,920); hum (1,518); Formal speech: ***que*** (3,194); ***a*** (2,248); ***de*** (2,241); ***o*** (1,926); ***e*** (1,689); Media: ***que*** (4,807); ***a*** (4,122); ***de*** (3,982); ***o*** (3,941); ***e*** (3,512); Telephone: ***é*** (1077); ***que*** (1011); ***eu*** (888); ***tá*** (823); ***e*** (537). These numbers show that dialogic particles are more frequent in informal and telephonic interactions, while formal and media interactions present conjunctions and a preposition pointing to more complex syntactic patterns.

Some of the most frequent ngrams found were: ***a gente tem que, nũ sei o quê, como é que é, eu acho que é***. All of them are phrases which occur in interactive situations, three of them expressing a deictic point of view centered on the speaker, thus asserting the predominance of first-person utterances (direct speech).

The most frequent verbs are: ***ser*** (18,259), ***ir*** (8,124), ***ter*** (7,642), ***fazer*** (4,638). As for verbs of saying, ***falar*** (2,778) and ***dizer*** (961) were the most frequent. The collocates of these two verbs are distinct, and show the preferences for specific constructions in language use. ***Falar*** collocates to the left preferably with personal pronouns ***eu, ele, cê*** (LogDice 10.05, 9.42 and 9.24), while ***dizer*** collocates with other verb forms in verbal compounds, such as ***quer dizer, vamos dizer, posso dizer*** (LogDice: 11.94, 9.26 and 8.79). ***Falar*** collocates to the right with ***com***, ***sobre*** and ***comigo*** (LogDice 9.89, 9.26 and 9.17), while ***dizer*** collocates with ***assim***, ***que*** and ***respeito*** (LogDice: 9.08, 8.77 and 8.47).

Modal verbs have to be further investigated as their orthographic form is usually multifunctional. Such is the case for the verb ***dar***, which can be used in several types of constructions, ranging from dative to modal ones. As an epistemic modal verb, ***dar*** collocates to the left preferably with ***nũ*** (119) and to the right with ***pra*** + V-INF (77). On the other hand, the verb ***poder*** occurs mostly as either a deontic or an epistemic modal (1,891), with a preference for ***nũ***, ***gente*** and ***cê*** as left collocates (LogDice 9.99, 9.62 and 9.55) and ***ser***, ***fazer*** and ***ter*** as right collocates (LogDice 11.26, 10.43 and 10.31).

The most frequent generic noun is ***coisa***, which collocates left with ***alguma, uma, outra*** (LogDice 11.78, 10.74 and 10.39) and ***assim***, ***que*** and ***interessante*** to the right (LogDice 9.27, 8.73 and 7.98). Another frequent generic noun is ***gente***, which shows up in bigrams such as: ***de gente, com gente, em gente, até gente, muita gente, gente a, gente de***.

4. Conclusion

This first brief incursion into the lexical information provided by the C-ORAL-BRASIL corpora has resulted in the generation of word, lemma, PoS, collocate and ngram frequency lists profiling the most typical occurrences in the corpora.

The overall results point towards some degree of specialization in the most frequent lexical items found in each one of the corpora that make up C-ORAL-BRASIL. The informal and the telephonic corpora exhibit ***tá***, ***né***, ***é*** as frequent indexes of agreement in dialogic interactions; whereas the formal and media corpora show conjunctions ***que*** and ***e***, and preposition ***de*** as indexes of a higher degree of syntactic density. High frequency verbs include copula, possession, dative and light verbs. Verbs of saying (***falar*** and ***dizer***) are also highly frequent and portray specific usage patterns. Verb use also indicates frequent modalization, which follows preferred patterns with verbs ***dar*** and ***poder***. Generic nouns ***coisa*** and ***gente*** are the most frequent ones and account for different collocational preferences.

Future work in this project will focus on a qualitative analysis anchored in the quantitative results already achieved in order to carry out more in-depth analyses conducive to an overall profiling of the basic spoken lexicon for BP.

References

Adolphs, S. & Schmitt, N. Lexical coverage of spoken discourse. Applied Linguistics, n. 24:4, 2003, pp. 425-438.

C-ORAL-BRASIL Corpora. Raso, T. & Mello, H. Available at: www.c-oral-brasil.org

Kilgarriff, A., Baisa, V., Bušta, J., Miloš Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P. & Suchomel, V. The Sketch Engine: ten years on. Lexicography, n.1, 2014, pp. 7-36.

Nation, I.S.P. Learning vocabulary in another language. Cambridge: CUP, 2001.

Python Programming Language. Available at: https://www.python.org/

Raso, T. & Mello, H. C-ORAL-BRASIL I: corpus de referência do português brasileiro falado informal. Belo Horizonte: Editora UFMG, 2012.

Sinclair, S. & Rockwell, G. 2016. Voyant Tools. Available at: http://voyant-tools.org/

Schonell, F.J., Meddleton, I. G., & Shaw, B. A. A study of the oral vocabulary of adults. Brisbane: University of Queensland Press.

Sketchengine. [Software]. Available at: https://www.sketchengine.eu/

Os discursos feministas no Brasil e na Alemanha: um estudo contrastivo assistido por corpus sobre suas estruturas temáticas

Feminist discourses in Brazil and Germany: a contrastive corpus-assisted study of their thematic structures

Andressa Costa[1]

1 Pesquisadora do pós-doutorado no departamento de Linguística aplicada e estudos da linguagem (LAEL) na PUC-SP. acosta.andressa@gmail.com.

***Resumo**:* Esse estudo tem como finalidade analisar e contrastar os discursos de feministas na Alemanha e no Brasil. Os procedimentos metodológicos seguem a abordagem da linguística de corpus. A base de dados compõe-se de dois corpora compilados para esse projeto – um em alemão (feminaDE) e outro em português (feminaBR) - compostos por textos de diferentes registros/gêneros textuais escritos por feministas. A modelagem de tópicos (*topic modeling*) com o modelo LDA (*latent dirichlet allocation*) foi o método usado para identificar temas latentes nos dois corpora. LDA é um modelo probabilístico gerativo cuja ideia básica é que textos são uma mistura de tópicos latentes e cada tópico é caracterizado por uma distribuição de palavras (BLEI; NG; JORDAN, 2003, p. 996). Nesse modelo, as palavras são as variáveis observáveis e a estrutura de tópicos, *clustesr* de palavras que coocorrem nos textos, são as variáveis escondidas. As variáveis observadas compõem-se de lemas de substantivos, verbos, adjetivos e advérbios. Os resultados mostram que há mais diferenças do que semelhanças em relação à composição dos tópicos. Mesmo havendo várias palavras-chave comuns nos dois corpora, elas coocorrem com diferentes palavras-chave criando diferentes estruturas temáticas. Por exemplo, apenas um tópico no corpus brasileiro apresenta estrutura temática similar a outros dois tópicos no corpus alemão.

***Palavras-chave**:* Análise do discurso; Linguística de corpus; Estudos culturais; Linguística contrastiva; Feminismo; Língua Alemã; Língua Portuguesa

***Abstract**:* This study aims to analyze and contrast the discourses of feminists in Germany and Brazil. The methodological procedures follow a corpus linguistics approach. The database consists of two corpora compiled for this project – one in German (feminaDE) and the other in Portuguese (feminaBR) – composed of texts from different registers/genres written by feminists. Topic modelling with the LDA (latent Dirichlet allocation) model was used to identify latent themes in the two corpora. LDA is a generative probabilistic model whose basic idea is that texts are a mixture of latent topics and each topic is characterized by a distribution of words (BLEI; NG; JORDAN, 2003, p. 996). In this model, the terms are the observable variables, and the topics, clusters of words that co-occur in texts, are the hidden variables. The observed variables are lemmas of nouns, verbs, adjectives and adverbs. The results show more differences than similarities regarding the composition of topics. Even though there are several common keywords in the two corpora, they co-occur with different keywords creating different thematic structures. For example, only one topic in the Brazilian corpus has a composition similar to other two topics in the German corpus.

***Keywords**:* Discourse analysis; Corpus linguistics; Contrastive studies; Contrastive linguistics; Feminism; German, Portuguese.

1. Introdução

Os discursos feministas estão cada vez mais presentes nos espaços públicos de modo global. Na Alemanha, a presença feminista nos meios de comunicação parece ser mais frequente e ter maior eco do que no Brasil. Apesar disso, nota-se um crescimento de vozes feministas no Brasil nos últimos anos. Esses discursos não são homogêneos, embora haja temas que podem ser considerados típicos feministas como emancipação da mulher, aborto, violência contra a mulher, entre outros. E mesmo esses temas similares parecem ter diferentes pesos e ser abordados de diferentes maneiras nas culturas alemã e brasileira.

O presente estudo apoia-se em duas ideias centrais que relacionam os conceitos de cultura, língua e sociedade: a) a língua tem como funções primárias servir de base para o desempenho de atividades e identidades sociais e servir de suporte para afiliação cultural, institucional e de grupos sociais (Glee 2005: 1); b) o uso padronizado da linguagem é resultado de atos de fala cooperativos e por isso determinados padrões de uso linguístico indicam determinados aspectos de atos de fala e podem, desse modo, ser interpretados como elementos centrais característicos de discursos (Bubenhofer 2009: 43).

Partindo-se desses pressupostos, considera-se que os padrões linguísticos identificados neste estudo revelam aspectos que são culturalmente relevantes para feministas brasileiras e alemãs. Assim, o estudo orienta-se pelas seguintes perguntas de pesquisa: a) Quais são os principais temas nos discursos feministas nas culturas alemã e brasileira? b) Que diferenças e similaridades temáticas há nos discursos feministas nessas culturas? A abordagem metodológica combina linguística de corpus, análise do discurso e linguística computacional. Isso implica o uso de corpora como base de dados e técnicas estatísticas para identificar padrões nos textos. O corpus compõe-se de textos em alemão e português escritos por feministas. Esse estudo é o primeiro de uma série de estudos que tem como finalidade analisar e contrastar os discursos de feministas alemãs, brasileiras, britânicas e americanas sob diferentes aspectos a fim de identificar similaridades e diferenças discursivas.

2. Métodos

2.1. Os dados da análise

O presente estudo tem como base um corpus bilíngue, composto por textos escritos por feministas brasileiras e alemãs. Foi compilado em grande parte da internet com a ferramenta BootCat (Baroni/Bernardini 2004) e uma parte foi escaneada. A estrutura do corpus é apresentada na tabela 1:

Tabela 1: Estrutura do corpus

Nome do Corpus	**FeminaBR**	**FeminaDE**
Registros	Blog, Website, livro, revista, ensaio,entrevista/ relatos, cartilha	Blog, Website, livro, revista, ensaios
Tamanho do corpus	286 textos	323 textos
	682.720 tokens	488.786 tokens
Período	de 2003 a 2020	de 2006 a 2020
Etiquetados	TreeTagger (1994)	TreeTagger (Schmid 1995)

A seleção do material foi feita a partir de um levantamento de autoras, organizações e grupos de mulheres que se autodenominam feministas. Definiu-se uma quantidade de no máximo 100 textos de cada registro, o que só foi possível para Websites, blogs e revistas. Por se tratar de corpus especializado, a quantidade de material disponível ainda é escassa para a construção de um corpus maior.

2.2. Modelagem de tópicos

Para a identificação de padrões temáticos usou-se a modelagem de tópicos com LDA (*Latent Dirichlet Allocation*), um modelo probabilístico desenvolvido para modelar dados textuais (Blei/Ng/Jordan 2003). Segundo Blei (2012), esse modelo parte do princípio de que cada documento é composto por múltiplos tópicos e cada documento apresenta esses tópicos em diferentes proporções. Nesse contexto, um tópico consiste em um *cluster* de palavras que coocorrem frequentemente nos textos. Os algoritmos, *topic models,* são usados para descobrir os temas principais presentes nos textos (Blei 2012).

Neste estudo, a modelagem de tópicos foi realizada em R com o pacote *topicmodel* (Grün/Hornik 2011). A unidade de observação foi o corpus todo de cada língua, já que o objetivo desse primeiro estudo é fazer uma comparação intercultural. As variáveis analisadas compõem-se dos lemas de substantivos, adjetivos, verbos e advérbios. A quantidade de tópicos a ser identificada pode ser determinada através de dois métodos:

1) coerência dos tópicos – examinando se as palavras nos tópicos fazem sentido. Assim, pode-se aumentar ou diminuir a quantidade de tópicos para torná-los coerentes.

2) medidas quantitativas – *Log-Likelihood* e *Perplexity*. O objetivo é maximizar *log-likehood* negativo e minimizar *perplexity* (McCracken 2020).

Os métodos quantitativos não foram muito úteis na determinação da quantidade de tópicos, pois os tópicos gerados, considerando-se os melhores índices de *log-likelihood* e perplexidade, não apresentaram uma composição interessante e possível de interpretar. O melhor índice do log-likelihood sugere 30 tópicos no corpus alemão (-1536747) e 29 tópicos no corpus brasileiro (-2189362.6), já o melhor índice da perplexidade sugere 27 tópicos no corpus alemão (2279.48) e 29 tópicos no corpus brasileiro (1339.93) como a quantidade mais adequada do ponto de vista estatístico. No entanto, além de ser um número muito elevado de tópicos, vários deles apresentam uma composição muito similar enquanto outros apresentam um conteúdo difícil de interpretar. Desse modo, decidiu-se analisar 10 tópicos, já que a sua composição é mais coerente e possibilita uma interpretação.

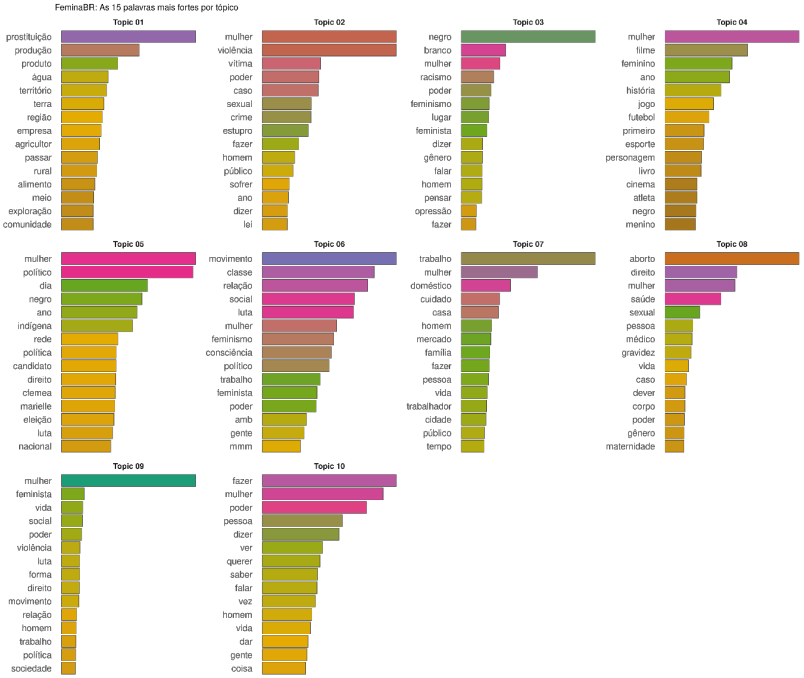
3. Resultados e discussão

A análise dos resultados baseou-se nas 15 palavras de maior peso de cada tópico. A identificação dos tópicos foi realizada por meio da modelagem de tópicos em R com o pacote *topicmodels* e teve os seguintes parâmetros:

```
feminaBR_mod = LDA(feminaBR_dtm, k=10, method="Gibbs",
                   control=list(alpha=0.001, seed=10005, burnin= 500,
                                delta= 0.1, iter=4000, thin= 100))
```

Nos gráficos 1 e 2 pode-se visualizar as 15 palavras mais marcantes de cada tópico. O gráfico 1 apresenta a estrutura temática do corpus do Português (feminaBR) e o gráfico 2 contém a estrutura temática do corpus alemão (feminaDE). Nos gráficos pode-se observar que há várias palavras-chave comuns aos dois corpora que ocorrem em diferentes constelações. Há também várias palavras específica de cada corpus.

Gráfico 1: as 15 palavras mais fortes por tópicos no corpus brasileiro

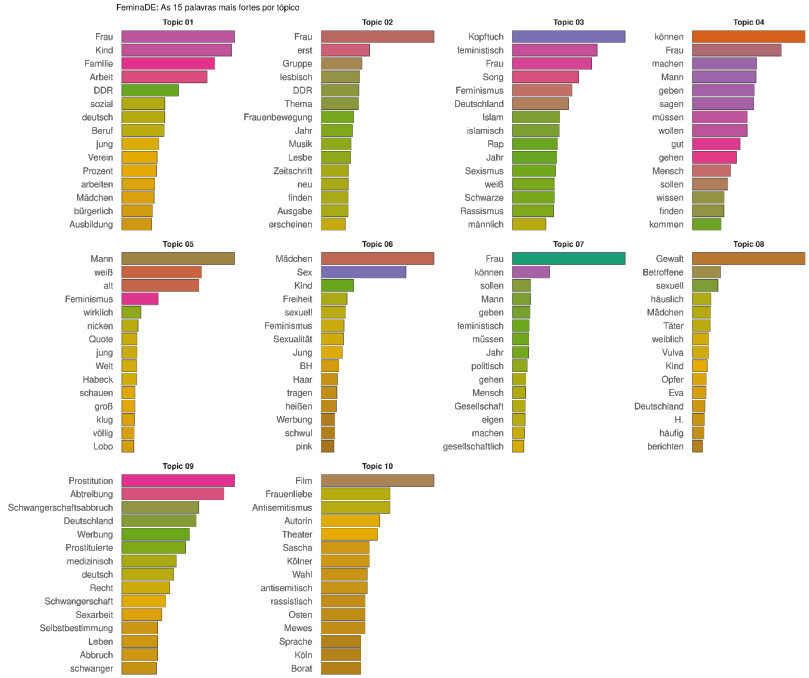


O tópico 1 do feminaBR compõe-se de palavras que fazem referência ao trabalho com a terra e à prostituição, trabalho sexual especialmente de mulheres. É o único tópico que não contém o termo 'mulher'. Todos os outros tópicos fazem referência à mulher e a algum outro aspecto. A interpretação dos tópicos baseou-se na análise dos 15 textos mais representativos de cada tópico. Assim temos os seguintes focos temáticos nos discursos de feministas brasileiras:

Tópico 01: Modelo de desenvolvimento capitalista
Tópico 02: Violência contra a mulher
Tópico 03: Racismo, gênero e o ser feminista
Tópico 04: Cinema, literatura e futebol
Tópico 05: Engajamento político de mulheres

Tópico 06: Feminismo como movimento social
Tópico 07: Trabalho doméstico e mercado de trabalho
Tópico 08: Direitos reprodutivos e saúde da mulher
Tópico 09: Militância feminista
Tópico 10: Sobre pessoas e suas experiências

Gráfico 2: as 15 palavras mais fortes por tópicos no corpus alemão



No gráfico 2 pode-se observar que a palavra ***Frau*** que significa mulher aparece em 5 dos 10 tópicos entre as palavras de mais peso. Isso é uma das diferenças entre os dois corpora já que no corpus brasileiro a palavra mulher aparece em 9 dos 10 tópicos também entre as palavras de maior peso. Com base nos 15 textos mais representativos de cada tópico do corpus alemão, foram identificados os seguintes temas:

Tópico 01: Família, educação e trabalho
Tópico 02: Rede de mulheres, arte e música
Tópico 03: Sexismo, racismo e feminismo pop
Tópico 04: Sobre pessoas e suas experiências
Tópico 05: Patriarcado, cota de mulheres e feminismo
Tópico 06: Estereótipos sobre feminilidade e masculinidade
Tópico 07: Engajamento e crítica social
Tópico 08: Violência sexual e doméstica
Tópico 09: Prostituição, aborto e autodeterminação
Tópico 10: Atividades políticas e ações artísticas

Nos gráficos pode-se observar que há várias palavras-chave comuns aos dois corpora que ocorrem em diferentes constelações. Há também várias palavras específica de cada corpus. Ao comparar a estrutura temática dos dois corpora pode-se observar que há apenas 3 tópicos similares: o tópico 10 do feminaBR (Sobre pessoas e suas experiências) tem uma composição bastante similar à dos tópicos 4 (Sobre pessoas e suas experiências) e 7 (Engajamento e crítica social) do feminaDE. Esses tópicos são constituídos por substantivos como mulher/***Frau***, homem/***Mann***, pessoa/***Mensch***, alguns verbos e apresentam um caráter mais geral e neutro em comparação com os outros tópicos, com exceção do tópico 7 do corpus alemão que, além dos termos já mencionados, também contém termos como ***feministisch***/feminista, ***politisch***/político, ***gesellschaftlich***/social. Os tópicos 10 do feminaBR e 4 do feminaDE contém 10 palavras comuns: fazer/***machen***, mulher/***Frau***, poder/***können***, pessoa/***Mensch***, dizer/***sagen***, querer/***wollen***, homem/***Mann***, dar/***geben***, gente/***Mensch***. Os 15 textos mais representativos dos dois tópicos abordam questões da vida em sociedade que mulheres e homens experienciam por conta do fato de serem mulheres e homens como por exemplo: ser mãe solteira ou pai solteiro, conciliar maternidade e profissão, a pobreza das mulheres que vivem sós. Por esse motivo, estes tópicos foram nomeados com o mesmo rótulo.

Outro aspecto interessante que se observa são as coocorrências do termo prostituição/***Prostitution*** nos dois corpora. No feminaBR aparece como a palavra-chave mais forte no tópico 1 (Modelo de desenvolvimento capitalista) e coocorre com outros termos relacionados ao trabalho no campo. Já no

feminaDE, o termo ***Prostitution*** aparece como a palavra-chave mais forte no tópico 9 (Prostituição, aborto e auto-determinação) e coocorre com termos relacionados a gravidez e aborto, como se pode observar na tabela 2:

Tabela 2: Tópicos com coocorrências do termo prostituição/***Prostitution***

FeminaBR Tópico 1	FeminaDE Tópico 9
Prostituição	Prostitution/prostituição
Produção	Abtreibung/aborto
Produto	Schwangerschaftsabbruch/aborto
Água	Deutschland/Alemanha
Território	Werbung/Propaganda
Terra	Prostituierte/prostituta
Região	medizinisch/medicinal
Empresa	deutsch/alemão
Agricultor	Recht/direito
Passar	Schwangerschaft/gravidez
Rural	Sexarbeit/trabalho sexual
Alimento	Selbstbestimmung/Auto-determinação
Meio	Leben/vida
Exploração	Abbruch/interrupção
comunidade	Schwanger/grávida

Com base na análise dos textos mais representativos dos tópicos apresentados na tabela 2, observa-se que, os textos de feministas brasileiras, a prostituição está relacionada a trabalho produtivo e de exploração assim como a trabalho rural. Os textos mais marcados nesse tópico tratam do modelo capitalista e abordam diferentes aspectos que mostram como esse modelo afeta de modo negativo a vida das mulheres através do seu trabalho. Em contrapartida, os textos de feministas alemãs tratam do debate público sobre o aborto, especialmente relacionado a questões legais como os parágrafos 218 e 219 da constituição alemã, ou sobre exploração das mulheres através da prostituição na Alemanha. Isso mostra como o mesmo tema pode ser abordado de maneira diferente em cada cultura.

Outro exemplo é o racismo. Esse aparece no tópico 3 (feminaBR: Racismo, gênero e o ser feminista; feminaDE: Sexismo, racismo e feminismo pop) nos dois corpora, também com certa diferença. No feminaBR, está claramente associado às pessoas negras, mulheres e homens. Em contrapartida, o racismo engloba, no feminaDE, não só as pessoas negras, mas também pessoas muçulmanas, o que se pode observar no gráfico 2 no qual palavras relacionadas ao mundo islâmico têm maior peso nesse tópico.

Por fim, há no corpus alemão referências à DDR (antiga República Democrática da Alemanha) a à própria Alemanha. Contudo, não há referências ao Brasil no corpus brasileiro. Em contrapartida, pode-se considerar o tópico 5 do feminaBR tipicamente brasileiro, uma vez que ele contém as palavras indígena, Marielle, cfemea (organização feminista brasileira) e o tópico 4 também, pois futebol tem uma forte associação com cultura brasileira. O tópico 5 do feminaDE parece ser típico da cultura alemã porque fala do velho homem branco e das cotas (de mulheres em posições de liderança) e o tópico 10 também por conter o termo antissemitismo, algo fortemente associado à cultura alemã. Todos os outros tópicos contêm várias palavras-chave comuns aos dois corpora, mas em diferentes composições, o que dever ser influenciado por questões específicas de cada cultura. Estas questões culturais e históricas necessitam de um estudo mais aprofundado.

Os resultados sugerem que os discursos de feministas nas culturas brasileira e alemão abordam questões relacionadas à condição de mulher na sociedade que se constituem, por um lado, em denúncias contra a violência, preconceitos raciais, sexuais e de gênero, exploração sexual e, por outro lado em crítica social que pode ter como objetivo conscientizar a sociedade sobre ideias preconcebidas (a mulher tem que se comportar segundo regras impostas pela sociedade) e comportamentos preconceituosos (que julgam e condenam mulheres por não seguirem regras impostas). A maior parte dos textos trata da mulher heterossexual e alguns poucos consideram lésbicas e mulheres trans. Apenas no tópico 2 (Rede de mulheres, arte e música) do corpus alemão aparecem as palavras ***lesbisch*** (adjetivo lésbica) e ***Lesbe*** (substantivo lésbica) e no tópico 6 (Estereótipos sobre feminilidade e masculinidade) a palavra ***schwul*** (gay).

4. Considerações finais

O objetivo deste primeiro estudo foi obter uma visão geral sobre os temas abordados nos discursos de feministas alemãs e brasileiras. Os resultados mostraram que os discursos feministas não são homogêneos, apenas três tópicos apresentam similaridade temática: tópicos 4 e 7 do corpus alemão, e tópico 10 do corpus brasileiro.

Com base na análise dos dados, pode-se concluir que, nas duas culturas, o feminismo tem um caráter de engajamento social, político e intelectual. Além disso, nota-se que as duas culturas compartilham temas que, todavia, são abordados em diferentes contextos como vimos com o termo 'prostituição' e 'aborto'. Os tópicos refletem questões centrais influenciadas pelo contexto social, histórico, político de cada cultura.

Referências bibliográficas

Baker, Paul. *Using Corpora in Discourse Analysis*. London, New York: continuum, 2006.

Baroni, Marco; Bernardini, Silvia. BootCaT: Bootstrapping corpora and terms from the web. *Proceedings of LREC 2004.* Disponível em: https://home.sslmit.unibo.it/~baroni/publications/lrec2004/bootcat_lrec_2004.pdf

Blei, David. Probabilistic Topic Models. Communications of the acm, vol. 55 no. 4, 2012, pp. 77-84. doi:10.1145/2133806.2133826.

Blei, David; Ng, Andrew; Jordan, Michael. Latent Dirichlet allocation. *Journal of Machine Learning Research,* 3: 993–1022, January 2003. Disponível em: http://www.cs.columbia.edu/~blei/publications.html.

Bubenhofer, Noah. *Sprachgebrauchsmsuter. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse*. Berlin, New York: de Gruyter, 2009.

Gee, Jame P. *An Introduction to Discourse Analysis: Theory and Method*, 2nd Ed. London, New York: Routledge, 2005.

Grün, Bettina; Hornik, Kurt. topicmodels: An R Package for Fitting Topic Models. *Journal of Statistical Software*, Volume 40, Issue 13, 2011, pp. 11-30. Disponível em: https://www.jstatsoft.org/article/download/v040i13/480.

McCracken, Cel. *TED Talks: AI and Topic Modelling*. RPubs, 2020 Disponível em: https://rpubs.com/CelMcC/645438

Schmid, Helmut. Improvements in Part-of-Speech Tagging with an Application to German. *Proceedings of the ACL SIGDAT-Workshop*. Dublin, Ireland, 1995.

Schmid, Helmut. Probabilistic Part-of-Speech Tagging Using Decision Trees. *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK, 1994.

Scripts da modelagem de tópicos em R: https://github.com/cristinnebr/feminims.git

We use "excuse me" when our body makes noises: A corpus-based contrastive study of pragmatic routines in undergraduate translation courses in China

Usamos "excuse me" quando nossos corpos fazem barulhos – um estudo contrastivo de rotinas pragmáticas baseado em corpus em cursos de tradução na China

Malila Prado[1]
Adriana Mendes Porcellato[2]
Xiao Wang[3]

1 Beijing Normal University/Hong Kong Baptist University United International College
2 University of São Paulo
3 Fujian University of Technology

Abstract**:* Pragmatic routines pose challenges to language learners and translators due to their context-dependent uses and meanings, which differ across languages and cultures. This paper describes how analyzing the meaning and use of pragmatic routines assisted pedagogical work in Corpus Translation Studies at a university in China. The pragmatic routine selected was ***excuse me, which was first introduced to the learners through teacher-designed worksheets with data collected from the Spoken BNC2014 corpus. Then students were encouraged to explore the use of ***excuse me*** in spoken language corpora available online using a bottom-up approach (Kádár & House, 2020c) which helped them identify and categorize six different meanings of the pragmatic routine under investigation. Based on the established categories, learners designed a discourse completion task (DCT) to collect data on Chinese behaviors in situations where ***excuse me*** would be expected in English. Finally, the Chinese pragmatic routines that emerged from the DCT data were scrutinized through Chinese corpora to arrive at a better understanding of their use at the sociopragmatic level. Based on the analyses of the work performed by students and the discussions in which they engaged, this pedagogical task proved highly useful to the use of Corpus Linguistics tools and the promotion of pragmatics awareness in a Corpus Translation undergraduate course in China.

***Keywords**:* Pragmatic Routines; Corpora; Corpus-based Pedagogy.

Resumo**:* As rotinas pragmáticas apresentam desafios para aprendizes de línguas e tradutores devido a seus usos e significados, que são estreitamente dependentes do contexto e diferem de uma língua e cultura para outra. Este trabalho descreve como a análise do significado e do uso de rotinas pragmáticas auxiliou o ensino de ***Corpus Translation Studies em uma universidade na China. A rotina pragmática selecionada foi ***excuse me***, apresentada pela primeira vez aos alunos através de atividades elaboradas pela professora com dados coletados do ***Spoken BNC2014***. Em seguida, os estudantes foram encorajados a explorar o uso de ***excuse me*** em corpora de linguagem oral disponíveis on-line, usando uma abordagem ***bottom-up*** (Kádár & House, 2020c) que os ajudou a identificar e categorizar seis significados diferentes da rotina pragmática em análise. Com base nas categorias estabelecidas, os alunos elaboraram um DCT para coletar dados sobre o comportamento dos chineses em situações em que se esperava o uso de ***excuse me*** em inglês. Por último, as rotinas pragmáticas chinesas que emergiram dos dados do DCT foram examinadas através de corpora chineses a fim de obter uma melhor compreensão de seu uso no nível sociopragmático. Com base nas análises do trabalho realizado pelos estudantes e das discussões em que se envolveram, esta unidade pedagógica se demonstrou útil no que diz respeito ao uso das ferramentas de Linguística de Corpus e à promoção da consciência pragmática em um curso de graduação de ***Corpus Translation*** na China.

***Palavras-chave**:* Rotinas pragmáticas; Corpora; Pedagogia baseada em corpus.

1. Introduction

Pragmatic routines as understood in this paper are "highly conventionalized pre-patterned expressions whose occurrence is tied to more or less standardized communication situations" (Coulmas 1981: 2). Examples include ***thank you***, ***please***, ***excuse me***, and ***good morning***, and their usage and interpretation is determined by social and cultural conventions (Bardovi-Harlig 2012).

One of the challenges faced by translation undergraduate students in China is finding equivalent translations to pragmatic routines that appear to be used in certain situations in English but not in Chinese (House et al. 2021). While several translations are possible, they should account for cultural differences in the context in which speech acts are performed or expected (Wierzbicka 2003). This calls for a framework that considers the nature of the situation rather than solely focusing on speech acts or even on pragmatic routines. In other words, such a framework, as suggested in Kádár and House (2020), should take into consideration the concepts of facework and rituals proposed by Goffman (1955; 1967).

This challenge prompted us to use pragmatic routines, more specifically ***excuse me***, to teach translation students how to explore corpora in a Corpus Linguistics course. This was in line with Bernardini (2000)'s suggestion of creating real problem-solving scenarios for students to facilitate their exploration and understanding of Corpus Linguistics tools.

This study adopts a bottom-up approach to the analysis of pragmatic routines, which starts from the expression under examination and then moves on to the analysis of the situations it is produced in, or, as Kádár and House (2020) put it, the ritual frame. To this end, it aims to investigate:

A) The effectiveness of Corpus Pragmatics (Clancy & O'Keefe 2015) activities designed for Chinese undergraduate students of translation to raise their awareness of the use of the pragmatic routine ***excuse me*** and its possible translations into Chinese;

B) The extent to which such activities help students develop an understanding of the use of Corpus Linguistics tools.

Another aim of this research was to understand the extent to which such activities help students translate pragmatic routines into Chinese; however, this last phase is ongoing and therefore not addressed in this paper.

This paper reports part of our investigation conducted on the basis of a pedagogical unit we used in the first two classes of the Corpus Linguistics course held at a provincial vocational college in the southeast of China.

2. Literature review

From a pragmatic perspective, when we speak, we perform actions, or speech acts, that influence the world around us (Austin 1962). These speech acts convey our intentions, which are then interpreted by our interlocutors thanks to the linguistic (or pragmalinguistic) and social (or sociopragmatic) conventions we share with them (Searle 1969).

In this framework, pragmatic routines can be defined as fixed or formulaic expressions, which, based on socio-cultural conventions, help indicate what speech act is being performed (Bardovi-Harlig et al. 2015; Bardovi-Harlig & Mossman 2016). Examples of typical pragmatic routines are ***Hello*** and ***Good morning*** to greet, ***I'm sorry*** to apologize, ***Thank you*** to express gratitude, and so on. However, this does not mean that there is a one-to-one correspondence between pragmatic routines and speech acts, as the same formulaic expressions can be used to express different speech acts. For instance, depending on the intonation, ***I'm sorry*** can also be also used as a request or as a complaint (cf. House et al. 2021).

Being "essential in verbal handling of everyday life" (House 1996: 228), pragmatic routines are an important aspect of learning a second language; yet, because of their intrinsic features, they can be difficult to fully grasp. According to Kecskes, "people using a particular language and belonging

to a particular speech community have ***preferred ways of saying things***” [author’s emphasis] (2014: 106), which directly reflects on the use of formulaic expressions. Since there are differences in the ways pragmatic routines are used across languages and cultures, it is possible that lingua franca speakers and L2 learners may “pick up these expressions without comprehending the socio-cultural load that they carry” (Kecskes 2014: 117). As a result, they may not fully understand their meaning and especially their usage. Consequently, it is not uncommon for learners to misuse such formulas either because they translate them directly from their L1 equivalents or because they use them when they are not expected to and, conversely, do not use them when they are (Bardovi-Harlig 2006). In addition, a recent study by House, Kádár, Liu, and Bi (2021) showed that Chinese learners of English have difficulties understanding and translating conventionalized expressions such as ***thank you very much*** or ***hello*** used with non-conventional functions in conversation.

Due to the difficulties they pose to language learners, pragmatic routines have been investigated from both a contrastive pragmatics and a teaching perspective. Interestingly, both types of approaches have commonly been based on Corpus Linguistics, which demonstrates that corpora are resourceful tools when it comes to investigating pragmalinguistic features.

In the field of contrastive pragmatics, Kádár and House (2020a; 2020c) recently proposed investigating pragmatic routines from a new framework based on Goffman’s concepts of facework and rituals. According to Goffman, facework designates “the actions taken by a person to […] counteract ‘incidents’ – that is, events whose effective symbolic implications threaten face” (Goffman 1955: 220), while face refers to “the positive social value a person effectively claims for himself” (Goffman 1955: 213). Rituals, on the other hand, are standard situations built up socially and culturally and often recognized by means of conventionalized utterances. Based on these premises, Kádar and House (2020a; 2020c) consider that pragmatic routines are in fact “ritual frame indicating expressions” (RFIE), that is, “pragmatically-heavy expression[s] indicating ritual standard situations” (Kádár; House 2020b:7).

Generally, pragmatics studies have adopted a top-down approach starting from the investigation of a speech act or another pragmatic phenomenon and then focusing on the language used to perform it. In contrast, Kádár and House (2020b) propose addressing contrastive pragmatics investigations from a bottom-up perspective using Corpus Linguistics. The authors suggest starting from the pragmalinguistic level, that is from pragmatic routines (or RFIEs) such as ***please*** and ***sorry***, which can be compared with their counterparts in other languages, for instance ***qing*** 请 and ***duibuqi*** 对不起 in Chinese. By filtering searches in a corpus for type of interaction (i.e. dyadic, multiparty, public, etc.) and type of situation (i.e. institutional with power-salience, informal with no power salience, ceremonial, etc.), the authors show that it is possible to identify sociopragmatic differences and similarities in the use of routines in different languages.

In a related vein, studies focused on pedagogical interventions have been carried out to verify the effects of teaching on the acquisition of pragmatic routines. Bardovi-Harlig et al. (2015), Bardovi-harlig and Mossman (2016) and Bardovi-Harlig et al. (2017), who concentrated on the use of teacher-designed materials based on corpus research to teach important speech acts in the academic domain, point out that such materials are helpful in drawing learners' attention to those speech acts. Another study (Bardovi-Harlig et al. 2019), in which a group of learners was assisted in corpus searches by a teacher, showed that this type of instruction engaged the learners in a discovery process and was particularly effective in drawing learners' attention to the pragmatic routines and their use in context. This study also emphasized the importance of corpus searches in teacher-designed materials.

Based on the above-mentioned studies, this paper will describe a pedagogical unit in which Corpus Linguistics was used as a tool to raise awareness of pragmalinguistic and sociopragmatic uses of the pragmatic routine ***excuse me*** and their Chinese counterparts.

3. Methodology

75 third-year students taking a Translation Course at a provincial vocational college in the southeast of China participated in this study. They were all enrolled in a Corpus Translation Studies course delivered by Author 1 from August to December 2021. All students gave written consent to taking part in this study.

In an attempt to adopt a bottom-up approach that “encompasses both interactional and pragmalinguistic / corpus-based methodologies” (Kádár & House 2020b: 7), we departed from collecting different situations in which ***excuse me*** was employed in favor of a corpus of authentic colloquial interactions produced by speakers in the United Kingdom and presenting social information about the participants and the context: the Spoken BNC2014 corpus (Love et al. 2017). Students analyzed and grouped these situations under usage-based categories such as interrupting someone, entering a room, asking someone to repeat what they said, and apologizing “when our body makes noises.” These situations assisted students in designing a Discourse Completion Task (DCT), a widely used instrument for collecting data in pragmatics studies that consists of a situational description or context “designed to constrain the response so that it elicits the desired communicative act” (Kasper 2008: 292). The DCT designed by the students and corrected by Author 3 was completed by native speakers of Chinese to verify whether or not they use pragmatic routines in the same situations as those collected from the British spoken corpus. The answers to the DCT were analyzed by the students, who were instructed to search for the expressions used in the responses in a spoken Chinese corpus.

The situations in which Chinese behaviors differ substantially from British behaviors were investigated in a multimodal English corpus; this corpus enabled us to collect short video clips which were presented to the same students. The students were then tasked with translating the video scenes by coming up with possible subtitles.

The next phase of the project – not yet described here - will be to collect and analyze the translating strategies deployed by the students to perform the task as well as the reasons why they made specific linguistic choices in their translations so as to gain insights into their awareness of Chinese and British cultural differences regarding the use of ***excuse me***.

4. Analysis

Turning to the application of the pedagogical unit, in this section, we first present the input given as well as the sources from which it was collected, followed by a description of how the pedagogical materials were explored and the discussions they gave rise to.

Following Bardovi-Harlig et al. (2019), the first phase was a teacher-designed pedagogical unit that extracted and displayed spoken data collected from the Spoken BNC2014 corpus. The students were asked to read conversations in which ***excuse me*** was used and identify what functions the expression fulfilled in each context. Some of the extracts used can be found in Table 1:

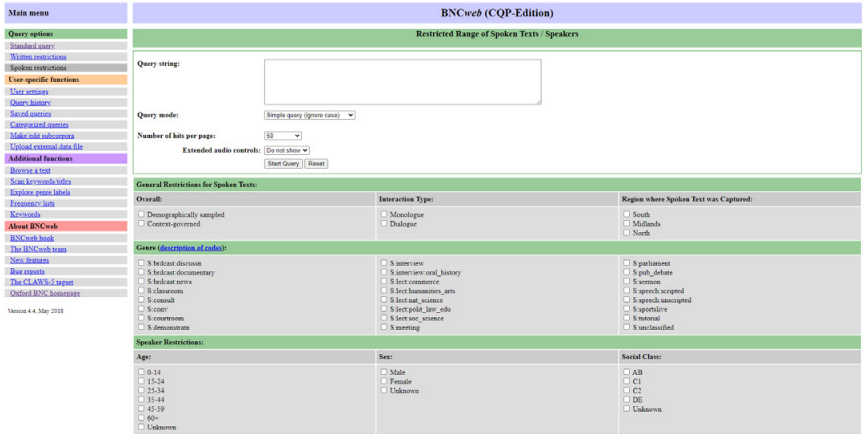
Table 1: Extracts of spoken dialogues taken from the Spoken BNC2014 corpus

1.	
Sue	I mean I think the special needs thing can be addressed…
Keith	Yeah.
Sue	…in a slightly different way.
	I think the management thing can be addressed in a different way <cough> erm **excuse me** erm, oh I've lost my thread now but, oh yeah that's right.
Keith	Yeah.

2.	
Harry	You can go and have a look at the radar if you want.
Unknown person	**Excuse me.** Do you know if Maria’s down there?
James	Yeah, bottom of the stairs, straight through that open sliding door, straight ahead of you, well, ju—just, to your right

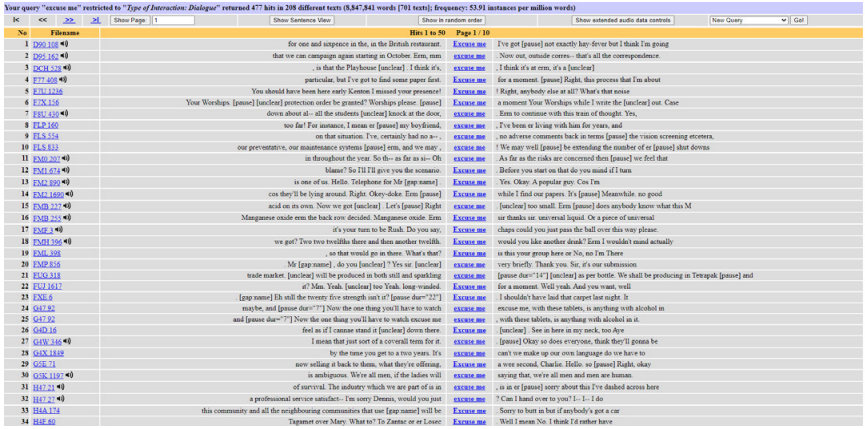
Working in groups, students received two extracts to work with, the task consisting of contextualizing the extracts (Who are the speakers?; Where are they? What are they doing? etc.) and identify the function of ***excuse me*** in each context. Following this initial exploration of the teacher-collected data, the students were asked to look for the expression ***excuse me*** in the Spoken BNC2014 corpus themselves. They received brief instructions on how to reach the page shown in Figure 1, where they could insert the search expression (***excuse me***) and select ***dialogue*** in the interaction type (as suggested in Kádár and House (2020a, 2020c)) since we were interested in the diverse functions played by ***excuse me*** in interactions.

Figure 1: Main screen: Spoken BNC2014 Corpus



The students were then exposed to concordance lines (Figure 2) and instructed to select lines presenting words (either to the left or the right of the node expression) they considered interesting. Some of the lines show a sound button, indicating that an audio file is available. Students seemed to choose concordance lines based on the availability of the audio file rather than on an attentive analysis of the collocates within the lines.

Figure 2: Concordance lines with ***excuse me*** in the center



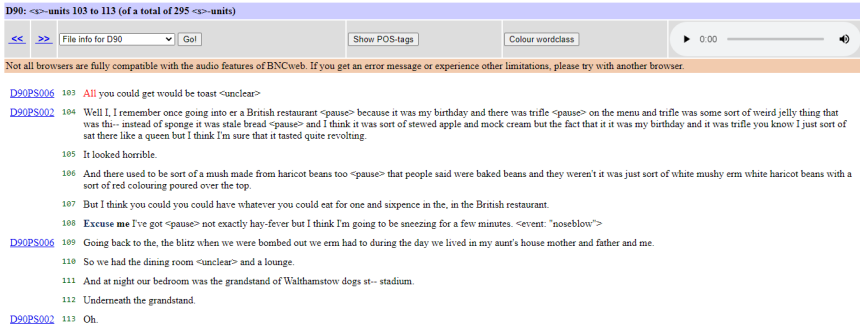
Your query "excuse me" restricted to "*Type of Interaction: Dialogue*" returned 477 hits in 208 different texts (8,847,841 words [701 texts]; frequency: 53.91 instances per million words)

|< << >> >| Show Page: 1 Show Sentence View Show in random order Show extended audio data controls New Query Go!

No	Filename	Hits 1 to 50	Page 1 / 10	
1	D90 108	for one and suxpence in the, in the British restaurant.	Excuse me	I've got [pause] not exactly hay-fever but I think I'm going
2	D95 162	that we can campaign again starting in October. Erm, mm	excuse me	. Now out, outside corres– that's all the correspondence.
3	DCH 528	, is that the Playhouse [unclear] . I think it's,	excuse me	, I think it's at erm, it's a [unclear]
4	F77 408	particular, but I've got to find some paper first.	Excuse me	for a moment. [pause] Right, this process that I'm about
5	F7U 1236	You should have been here early Kentne I missed your presence!	Excuse me	! Right, anybody else at all? What's that noise
6	F7X 156	Your Worships. [pause] [unclear] protection order be granted? Worships please. [pause]	Excuse me	a moment Your Worships while I write the [unclear] out. Case
7	F8U 430	down about at– all the students [unclear] knock at the door,	excuse me	. Erm to continue with this train of thought. Yes,
8	FLP 160	too far! For instance, I mean ur [pause] my boyfriend,	excuse me	, I've been er living with him for years, and
9	FLS 554	on that situation. I've, certainly had no a–,	excuse me	, no adverse comments back in terms [pause] the vision screening etcetera,
10	FLS 833	our pre\ventative, our maintenance systems [pause] erm, and we may ,	excuse me	! We may well [pause] be extending the number of er [pause] shot downs
11	FMD 203	in throughout the year. So th– as far as si– Oh	excuse me	. As far as the risks are concerned then [pause] we feel that
12	FMJ 674	blame? So I'll I'll give you the scenario.	Excuse me	. Before you start on that do you mind if I turn
13	FMJ 890	is one of us. Hello. Telephone for Mr [gap name]	Excuse me	. Yes. Okay. A popular guy. Cos I'm
14	FMJ 1690	cos they'll be lying around. Right. Okay-doke. Erm [pause]	Excuse me	while I find our papers. It's [pause] Meanwhile. no good
15	FMB 227	acid on its own. Now we got [unclear] . Let's [pause] Right	excuse me	. [unclear] too small. Erm [pause] does anybody know what this M
16	FMB 255	Manganese oxide erm the back rov decided. Manganese oxide. Erm	Excuse me	sir thanks sir. uni\versal liquid. Or a piece of universal
17	FMF 3	it's your turn to be Rush. Do you say,	Excuse me	chaps could you just pass the ball over this way please.
18	FMH 396	we got? Two two twelfths there and then another twelfth.	Excuse me	would you like another drink? Erm I wouldn't mind actually
19	FML 398	, so that would go in there. What's that?	Excuse me	is this your group here or No, no I'm There
20	FMP 856	Mr [gap name] , do you [unclear] ? Yes sir. [unclear]	excuse me	very briefly. Thank you. Sir, it's our submission
21	FUG 318	trade market. [unclear] will be produced in both still and sparkling	excuse me	[pause dur="1.4"] [unclear] as per bottle. We shall be producing in Tetrapak [pause] and
22	FUJ 1617	it? Mm. Yeah. [unclear] too Yeah. long-winded.	Excuse me	for a moment. Well yeah. And you want, well
23	FXE 6	[gap name] Eh still the twenty five strength isn't it? [pause dur="22"]	Excuse me	. I shouldn't have laid that carpet last night. It
24	G47 92	maybe, and [pause dur="7"] Now the one thing you'll have to watch	excuse me	excuse me, with these tablets, is anything with alcohol in
25	G47 92	and [pause dur="7"] Now the one thing you'll have to watch excuse me	excuse me	, with these tablets, is anything with alcohol in it.
26	G4D 16	feel as if I cannae stand it [unclear] down there.	Excuse me	[unclear] . See in here in my neck, too Aye
27	G4W 346	I mean that just sort of a coverall term for it.	excuse me	[pause] Okay so does everyone, think they'll gonna be
28	G4X 1849	by the time you get to a two years. It's	excuse me	can't we make up our own language do we have to
29	G5E 71	now selling it back to them, what they've offering,	Excuse me	a wee second, Charlie. Hello. so [pause] Right, okay
30	G5K 1197	is ambiguous. We're all men, if the ladies will	excuse me	saying that, we're all men and men are human
31	H47 21	of survival. The industry which we are part of is in	excuse me	, is in er [pause] sorry about this I've dashed across here
32	H47 27	a professional service satisfact– To Avery Dennis, would you just	excuse me	? Can I hand over to you? I– I– I do
33	H4A 174	this community and all the neighbouring communities that use [gap name] will be	Excuse me	. Sorry to butt in but if anybody's got a car
34	H4F 60	Tagarnet over Mary. What to? To Zaenic or er Losec	Excuse me	. Well I mean No. I think I'd rather have

Selecting a line takes the analyst to the co-text page, where a longer extract with turns before and after the utterance can be found, as seen in Figure 3.

Figure 3: Part of a dialogue in which ***excuse me*** is used



D90: <s>-units 103 to 113 (of a total of 295 <s>-units)

<< >> File info for D90 Go! Show POS-tags Colour wordclass 0:00

Not all browsers are fully compatible with the audio features of BNCweb. If you get an error message or experience other limitations, please try with another browser.

D90PS006 103 All you could get would be toast <unclear>

D90PS002 104 Well I, I remember once going into er a British restaurant <pause> because it was my birthday and there was trifle <pause> on the menu and trifle was some sort of weird jelly thing that was thi-- instead of sponge it was stale bread <pause> and I think it was sort of stewed apple and mock cream but the fact that it it was my birthday and it was trifle you know I just sort of sat there like a queen but I think I'm sure that it tasted quite revolting.

105 It looked horrible.

106 And there used to be sort of a mush made from haricot beans too <pause> that people said were baked beans and they weren't it was just sort of white mushy erm white haricot beans with a sort of red colouring poured over the top.

107 But I think you could you could have whatever you could eat for one and sixpence in the, in the British restaurant.

108 **Excuse me** I've got <pause> not exactly hay-fever but I think I'm going to be sneezing for a few minutes. <event: "noseblow">

D90PS006 109 Going back to the, the blitz when we were bombed out we erm had to during the day we lived in my aunt's house mother and father and me.

110 So we had the dining room <unclear> and a lounge.

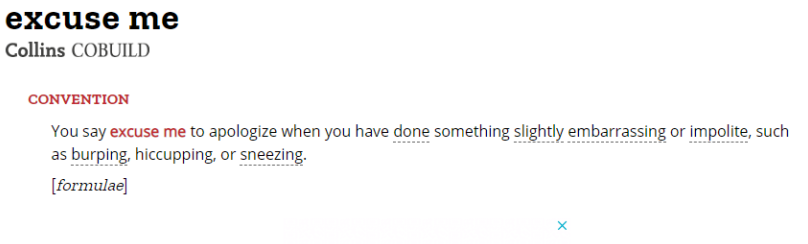
111 And at night our bedroom was the grandstand of Walthamstow dogs st-- stadium.

112 Underneath the grandstand.

D90PS002 113 Oh.

The students selected examples of ***excuse me*** in small groups and then discussed what functions the expression fulfilled. They categorized their investigations under six different headings: (1) entering a room; (2) interrupting someone; (3) apologizing; (4) asking someone to repeat what they said; (5) leaving a room or group of people; and (6) apologizing “when our body makes noises.” The last category is worth addressing: when Author 1 heard of this category, she realized the importance of enabling the students to carry out their own investigations on their own terms since her cultural values would have imposed a different category – that of doing something impolite – on the students. Consulting a dictionary with the students yielded the following information (Figure 4):

Figure 4: a dictionary entry of ***excuse me***



excuse me

Collins COBUILD

CONVENTION

You say **excuse me** to apologize when you have done something slightly embarrassing or impolite, such as burping, hiccupping, or sneezing.

[*formulae*]

Source: https://www.collinsdictionary.com/dictionary/english/excuse-me

The students were surprised with this dictionary definition of ***excuse me***, which corresponded to the way Author 1 conceived of the expression in use. A discussion with the students revealed that burping, sneezing, or coughing cannot always be controlled or suppressed; in some cultures, burping, for example, is desirable as an indicator of satisfaction after a meal (see Freitag-Hild, 2016 for similar cases). Kádár and House (2020b: 3) remind us that "[a]ll ritual forms of ritual behavior are very familiar to those who practise them, even though they may be 'exotic' or 'unusual' to those outside the lingua-culture in which they occur." Since describing such spontaneous actions as impolite reveals cultural bias, such a conclusion would most likely not have been reached had the students not identified this function in the data first.

In an attempt to identify what expressions – if any – would be used in Chinese in comparable situations, the students were invited to design a DCT based on the categories of ***excuse me*** they had raised. Student-generated DCTs serve as a means of promoting metapragmatics as students are invited to ponder a diversity of situations in which a particular language item can be used (McLean 2005: 151). This type of instrument can also be employed as a way of eliciting data from a large number and wide variety of speakers in terms of gender, age, etc. In addition, DCTs allow for the manipulation of relevant pragmatic variables such as power relations, social distance, or degree of imposition of the speech act (Kasper 2008; Ishihara & Cohen 2010), all of which are necessary in corpora-based investigations. Ishihara & Cohen (2010)'s suggestions of DCTs were used to assist students in designing their own. The students once again worked in groups, with each group generating a situation based on the categories previously established, writing it up in English, and then translating it into Chinese, with Author 3 verifying the final translation. Table 2 shows an example of the spreadsheets used by the students to design their DCTs:

Table 2: Student-generated DCT

Group No.	Situation	Chinese translation
1	You are in a hurry and need to talk to your teacher, but she is talking to someone else. What will you do or what do you say?	你有急事要跟老师说 , 但她在跟别人讲话。你会怎么做？
2	Your two friends are discussing their homework, and you are going to the supermarket and you want to know if they need you to buy something for them.	你的两个朋友在讨论作业 , 你要去超市 , 想知道需不需要你帮他们买点东西。你会怎么做/说什么?
3	You are talking with your friends about idols. One of your friends says that an idol is not that charming and not good at singing at all. This idol happens to be your favorite. What do you do or say?	“你正在和你的朋友谈论偶像。你的一个朋友说偶像没那么迷人 , 一点也不擅长唱歌。这个偶像恰好是你最喜欢的。你做什么/说什么？”
4	Today is Monday. You woke up late as you did not hear the alarm clock. You get up and run very fast. Unfortunately, you bump into a person. What do you do or say?	今天是星期一。你没听到闹钟所以起床晚了。你迅速起床跑去教室。不幸的是 , 你撞到了一个人。你会做什么/说什么？
5	You are having a date with your boyfriend or girlfriend in a restaurant and you eat a lot while chatting, so you burp. What do you do or say?	你和你的男朋友或女朋友在餐厅约会 , 你吃了很多 . 当你们在聊天的时候你打了个嗝 , 这个时候你会怎么做/说什么?

As homework, the students were requested to send out a link of the DCT

to classmates or family members. A total of 76 people participated in the DCT. Part of the spreadsheet presenting this data can be seen in Figure 5:

Figure 5: Data collected through the DCT

G	H	I	J	K
等她说完	不问 直接去	救救孩子	你懂个屁	对不起
等老师聊完再说；或者我先走，之后再说	我要去超市买东西，你们有要带的吗	你们先讨论，实在抱歉	没事，每个人都有喜欢的人，我觉得他好就行	不好意思，不好意思，对不起
等老师说完我再说	打扰一下，我要去超市买东西，你们有什么需要我带的东西吗？	抱歉，我有事需要立马去医院，你们先自己讨论。	不好意思，人无完人，我就是喜欢。	对不起
等着他说完	直接问	直接说	是的，你说的都对	抱歉
等待老师完成谈话再上前说话	直接询问	告知组员情况	并不反驳人各有志	迅速道歉留下名字
老师不好意思打断一下。	我要去超市，你们需要我带什么吗？	不好意思，我胃疼。	你认真的吗？	对不起！
不好意思，老师，打扰一下。我有急事。	朋友们，我要去超市，要不要顺带点啥。	不好意思，我现在很不舒服，我可能不能继续完成作业了，我得先去医院。	我觉得还是不要随意评价一个人吧。我很喜欢这个idol，虽然他不擅长唱歌，但是他能成为偶像就说明他有令人喜欢的点。	不好意思，不好意思，我赶时注意看路。
老师，打扰一下	你们需要我帮你们带点东西吗	抱歉，我得先走一步	我觉得你说的不对，你不了解他	对不起
等老师说完我在跟老师说。	发消息给同学	我会让他们商量完后，通知我	尊重别人的想法，遵从自己内心	会跟被撞的人说对不起
等她说完	发消息在他微信上	我会让他们先讨论过后通过微信或者钉钉告诉我	每个人都有自己的想法和看待别人的眼光，我会对他说的话表示认同，同时坚持自己的想法	我会说对不起然后请求对方原
打扰了，老师	你们需要我带点什么嘛	对不起，我有点不舒服，离开一下	-	对不起

Following this data collection, the students investigated the expressions used by the DCT participants in the BCC Modern Chinese Corpus, a 9.5-billion-word corpus consisting of newspapers, books, and dialogues from a social media platform and television subtitles (http://bcc.blcu.edu.cn/help#intro). Part of this data is shown in Figure 6, with one of the search expressions – ***buhaoyisi*** (不好意思) – in the center:

Figure 6: Concordance lines for buhaoyisi (不好意思)

Source: BCC Modern Chinese Corpus

The screen shown in Figure 6 generated fruitful discussion as the expression ***buhaoyisi*** (不好意思) appears to be used differently from its most direct English counterpart ***I’m sorry***, even though it is commonly used for apology-making. One of the students mentioned that in one of the lines, “不好意思 ***is not used as an apology. It is a way to refuse an offer, but he is not exactly refusing***.” The same conclusion was drawn in the study by Cheng et al. (1995), who confirmed that refusals are a strategy used by Chinese speakers to indicate consideration toward the interlocutor and to avoid threatening the positive social value of those involved in the interaction (facework). If the offer is sincere, it will be repeated, and only then can it finally be accepted.

5. Conclusion and future directions

The bottom-up approach investigation of ***excuse me*** first helped the teacher collect data to design pedagogical materials oriented to the needs of Chinese undergraduate students of translation, and then assisted learners' investigations of pragmatic routines, promoting their pragmatic awareness.. Searching corpora and assigning different occurrences of ***excuse me*** to categories based on the situational context helped students to be aware of the pragmatic routine ***excuse me*** not only at the pragmalinguistic level but also at the sociopragmatic level by understanding its contextual, social, and cultural use.

Moreover, while looking for the expression ***excuse me*** along with the Chinese expressions elicited through the DCT from different corpora, the students showed increasing comprehension of how to use the tools available to them and what kinds of searches corpora are particularly useful for. As a result, the students were motivated to pursue other investigations proposed throughout the Corpus Translation course and seemed confident about exploring and extracting data for other purposes and activities. Pragmatic routines offer real-life scenarios or problem-solving situations for students to tackle through Corpus Linguistics. Furthermore, the analysis of authentic data helped students recognize cultural differences they would not otherwise have noticed.

The findings of this study will inform translation courses regarding the raising of pragmatic awareness through the use of Corpus Linguistics. This activity kept students engaged from beginning to end, allowing them to understand how to contextualize spoken data and thus make appropriate translation choices.

References

Austin, J. L. *How to do things with words*. London: Oxford University Press, 1962.

Bardovi-harlig, K. On the role of formulas in the acquisition of L2 pragmatics. *Pragmatics and language learning*, 11, Honolulu, 2006, pp. 1–28.

Bardovi-Harlig, K. Formulas, routines, and conventional expressions in pragmatics research. *Annual Review of Applied Linguistics*, 32, Cambridge, 2012, pp. 206–227.

Bardovi-Harlig, K.; Mossman, S.; Vellenga, H. E. The effect of instruction on pragmatic routines in academic discussion. *Language Teaching Research*, 19/3, London, Jul 2015, pp. 324–350.

Bardovi-harlig, K.; Mossman, S. Corpus-based materials development for teaching pragmatic routines. In: Tomlinson, B. (Ed.); *SLA research and material development for language learning.* New York, Abingdon: Routledge, 2016: 250-267.

Bardovi-Harlig, K.; Mossman, S.; Su, Y. The effect of corpus-based instruction on pragmatic routines*. Language Learning & Technology*, 21/3, 2017, pp. 76–103.

Bardovi-Harlig, K.; Mossman, S.; Rothgerber, J.; Su, Y.; Swanson, K. Revisiting clarifications: Self- and other-clarifications in corpus-based pragmatics instruction. In: Sato, M.; Loewen, S. (Eds.); *Evidence-based second language pedagogy: A collection of instructed second language acquisition studies.* New York: Routledge, 1st ed, 2019: 52-80.

Bernardini, S. *Competence, capacity, corpora: A study in corpus-aided language learning***.** Bologna: CLUEB, 2000.

Cheng, X., Ye, L., & Zhang, Y. Refusing in Chinese. In: Kasper, G. (Ed.) *Pragmatics of Chinese as a native and target language.* Honolulu, HI: University of Hawai›i Press, 1995**:** 119-163.

Clancy, B., & O´Keefe, A. Pragmatics. In: Biber, D., & Reppen, R. *The Cambridge Handbook of English Corpus Linguistics.* Cambridge: Cambridge University Press, 2015: 235-251.

Coulmas, F. *Conversational routine: Explorations in standardized communication situations and prepatterned speech.* The Hague: Mouton, 1981.

Freitag-Hild, B. Identity, intercultural relationships and growing up in the 1970s: Teaching Anita and me to promote inter- and transcultural learning. In: Viebrock, B. (Ed.) *Feature films in English language teaching.* Tübingen: Narr Francke Attempto, 2016: 207-216.

Goffman, E. On Face-Work: An analysis of ritual elements in social interaction. *Psychiatry*, 18/3, Philadelphia, Aug 1955, pp. 213–231.

Goffman, E. Interaction ritual: Essays on face to face behavior. New York: Doubleday, 1967.

House, J. Developing pragmatic fluency in English as a foreign language: Routines and metapragmatic awareness. *Studies in Second Language Acquisition*, 18/2, Jun 1996, pp. 225–252.

House, J., Kádár, D., Liu, F., & Bi, Z. Altered speech act indication: A problem for foreign language learners? *System,* 101, Oxford, Oct 2021, pp. 1-11.

Ishihara, N.; Cohen, A. D. *Teaching and learning pragmatics: Where language and culture meet.* Edinburgh: Pearson Education Limited, 2010.

Kádár, D. Z.; House, J. Revisiting the duality of convention and ritual: A contrastive pragmatic inquiry. *Poznan Studies in Contemporary Linguistics*, 56/1, Berlin, Mar 2020a, pp. 83–111.

Kádár, D. Z.; House, J. The pragmatics of ritual: An introduction. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)*, 30/1, Antwerpen, Mar 2020b, pp. 1–14,

Kádár, D. Z., & House, J. Ritual frames: A contrastive pragmatic approach. *Pragmatics,* 30/1, Amsterdam, Dec 2020c, pp. 142–168.

Kasper, G. Data collection in pragmatics. In: Spencer-Oatey, H. (Ed.); *Culturally speaking*. 2nd ed., New York: Continuum, 2008: 316–341.

Kecskes, I. *Intercultural Pragmatics*. New York: Oxford University Press, 2014.

Love, R., Dembry, C., Hardie, A., Brezina, V. & McEnery, T. The Spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics,* 22/3, Amsterdam, Nov 2017, pp. 319-344.

McLean, T. Why no tip?: Student-generated DCTs in the ESL classroom. In: Tatsuki, D. (Ed.) *Pragmatics in language learning, theory, and practice.* Tokyo: Pragmatics Special Interest Group of the Japan Association for Language Teaching, 2005: 150-156.

Searle, J. *Speech Acts*. Oxford: Cambridge University Press, 1969.

Wierzbicka, A. *Cross-cultural pragmatics: The semantics of human interaction* (2nd. ed.). Berlin/New York: Mouton de Gruyter, 2003.

Um estudo da função conativa em um corpus paralelo através da pesquisa de verbos no modo imperativo por meio do Sketch Engine

A study of the conative function in a parallel corpus through the research of verbs in the Imperative Mood through the use of Sketch Engine

Anna Catharina de Mendonça Paes[1]

1 Aluna de mestrado na Universidade de São Paulo – USP.

Resumo: Este trabalho é um estudo de um corpus paralelo composto de um conjunto de treze textos fonte em inglês e seus respectivos treze textos traduzidos para português do antigo website da Escola Moderna de Mistérios (sede brasileira). Os textos foram traduzidos pela autora deste trabalho entre 2012 e 2013 e foram analisados em sua monografia de graduação (PAES, 2016). Com os métodos da linguística de corpus, utilizando-se o software Sketch Engine (KILGARRIFF et al, 2004), os mesmos textos foram analisados revelando novas informações sobre o corpus, a saber: o uso da função conativa que se evidencia pela alta frequência dos pronomes "*you*" e "você". Para investigar mais a fundo a função conativa no corpus paralelo, foram levantados os verbos com maior ocorrência e comparou-se a ocorrência de verbos no modo imperativo tanto nos textos traduzidos como nos textos fonte.

Palavras-chave: corpus paralelo; função conativa; Sketch Engine; Estudos da Tradução.

Abstract: This paper is a study of a parallel corpus composed of a set of thirteen source texts in English and their respective thirteen translated texts in Portuguese from the earlier version of the Brazilian Modern Mystery School's website. The texts were translated by the author of this paper between the years of 2012 and 2013, these texts were analyzed in her undergraduate monography (PAES, 2016). Following corpus linguistics methods, using the software Sketch Engine (KILGARRIFF et al, 2004), the same texts were analyzed revealing new information about the corpus: the use of the conative function that presents itself through the high frequency of occurrences of the pronouns "you" and "*você*". To investigate further the conative function in the parallel corpus, the verbs with highest frequency of occurrences were listed, and a comparison was made between the occurrences of verbs in the Imperative Mood in the translated texts and in the source texts as well.

Keywords: parallel corpus; conative function; Sketch Engine; Translation Studies.

1. Introdução

Este trabalho se propõe a voltar a um conjunto de treze textos fonte e dos respectivos treze textos traduzidos do antigo website https://brasilmms.com/[2], sendo a autora deste artigo a tradutora desses textos, os quais já foram analisados em sua monografia de graduação pela Universidade Federal de Santa Catarina (USFC). Essa análise anterior foi sob a perspectiva da competência tradutória e de suas subcompetências conforme são definidas pelo Grupo PACTE (Beeby et al, 2008). Porém, desta vez, vale-se da Linguística de Corpus e de suas ferramentas, para reunir os textos fonte e textos traduzidos em um corpus paralelo, personalizado ou "DIY" (*Do It Yourself* – Faça Você Mesmo).

Os textos fonte do corpus de estudo, são em parte textos em inglês canadense escritos pelo setor administrativo e instrutores da sede internacional da *Modern Mystery School* (MMS - Escola Moderna de Mistérios) em Toronto e em parte escritos pelo Fundador da MMS, Gudni Gudnason, natural da Islândia, falante não-nativo da língua inglesa. Os textos traduzidos para o português brasileiro foram encomendados pelos presidentes da filial da Escola Moderna de Mistérios (MMS) no Brasil, Liza Rossi e Eric Thompson para facilitar a divulgação dos cursos e da MMS no Brasil, que se define como uma instituição de ensino que fornece "formação espiritual avançada" e autoconhecimento, conforme o adágio "conhece-te a ti mesmo", recorrente no corpus.

O objetivo geral desta pesquisa é realizar um experimento de aplicação de conhecimentos básicos de Linguística de Corpus, reaproveitando um conjunto de textos disponíveis e ainda não estudado por outros pesquisadores da área. Os objetivos específicos derivaram de um levantamento inicial através da análise automatizada que encontrou uma alta ocorrência dos pronomes "*you*" e "você", sétima e décima palavras mais frequentes em seus respectivos subcorpora. A partir desse levantamento inicial, surgiram dois objetivos:

---

2 Website da MMS Brasil – Modern Mystery School – Escola Moderna de Mistérios no Brasil, versão do site do ano de 2012, o site já foi alterado/atualizado, de modo que os textos não estão mais disponíveis online, pois foram substituídos/reescritos. Link do website atual: https://escolamodernademisterios.com.br/

1) buscar dados que comprovassem a predominância da função conativa, uma hipótese levantada a partir da alta ocorrência desses pronomes pessoais; 2) verificar através da observação da tradução dos verbos se houve uma tendência a manter os elementos da função conativa nos textos traduzidos, em comparação com os textos fonte.

Nesta pesquisa, o conjunto de textos foi alinhado em um corpus paralelo e os subcorpora em inglês (textos fonte) e em português (textos traduzidos) foram analisados por meio do software de análise de corpus Sketch Engine (KILGARRIF et al, 2004) separadamente, em comparação com corpora de referência nas línguas respectivas para obtenção não apenas da *Word List* (Lista de Palavras) mas também para obter a *Keyword List* (Lista de Palavras-chave). Com o software, foi possível verificar palavras colocadas (*collocates*) e palavras muito frequentes que na primeira análise não foram levadas em conta, porém *collocates* não foram alvo deste estudo.

Este artigo está organizado da seguinte forma: 1) Introdução, esta seção; 2) Quadro Teórico, que traz o aporte teórico deste estudo; 3) Métodos, seção que descreve os métodos usados passo-a-passo; 4) Resultados, em que são apresentados os resultados obtidos; 5) Discussão do Resultados, em que os resultados são comentados levando em conta o processo da pesquisa do corpus; 6) Considerações Finais, a qual traz sugestões de novas pesquisas.

2. Quadro Teórico

2.1. Linguística de Corpus nos Estudos da Tradução: fornecimento de novas informações

Mona Baker, em seu texto seminal de 1993, "Linguística de corpus e estudos da tradução: implicações e aplicações", tinha a convicção de que um ponto de virada nos Estudos da Tradução era iminente. Ela defendia que:

> [...] esse ponto da virada chegará como consequência [...] do acesso a grandes corpora de textos originais e traduzidos e do desenvolvimento de métodos e ferramentas específicos para interrogar esses corpora. E [grandes corpora] também nos permitirão explorar, em uma escala maior do que jamais foi possível, os princípios que regem o comportamento do tradutor e as restrições sob as quais ele opera. (BAKER 1993: 235, tradução nossa) [3]

Um exemplo disso é o estudo deste corpus paralelo de website, em que houve informações que surgiram a partir das ferramentas de busca no âmbito da linguística de corpus e que não constavam da análise anterior (PAES, 2016), pois a leitura e releitura de textos não revelam tudo o que a pesquisa automatizada de corpus revela, como palavras e expressões recorrentes. Essas palavras e expressões recorrentes podem mostrar aspectos antes não tão claros sobre a forma que os significados se constroem no texto.

---

3 No texto fonte, lê-se: "[...] this turning point will come as a direct consequence of access to large corpora of both original and translated texts, and of the development of specific methods and tools for interrogating such corpora in ways which are appropriate to the needs of translation scholars. Large corpora will provide theorists of translation with a unique opportunity to observe the object of their study and to explore what it is that makes it different from other objects of study, such as language in general or indeed any other kind of cultural interaction. It will also allow us to explore, on a larger scale than was ever possible before, the principles that govern translational behaviour and the constraints under which it operates."

2.2. A Classificação do corpus deste estudo

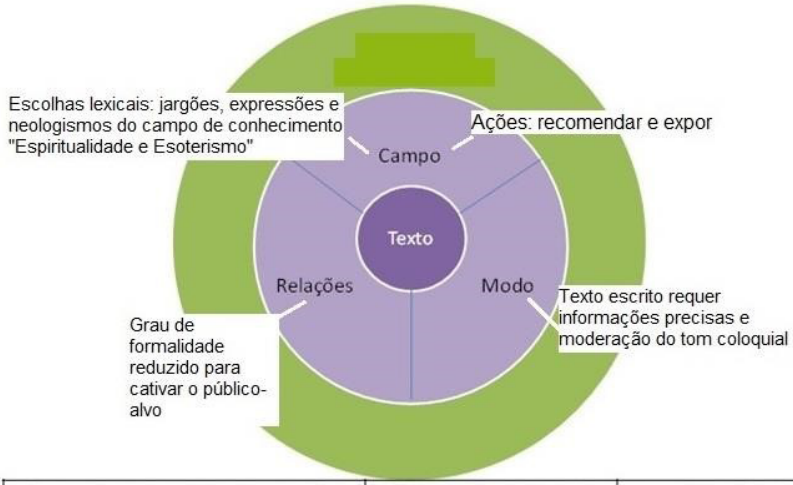
Nos Estudos da Tradução baseados em Corpora, podemos classificar um corpus pelos seguintes critérios: Campo de estudos – tradutório ou linguístico; Domínio – geral ou restrito; Modo – escrito ou falado; Tipo de relação entre os textos – comparável ou paralelo; Restrição temporal – diacrônico ou sincrônico; Número de línguas – monolíngue, bilíngue ou multilíngue; Direcionalidade – unidirecional, bidirecional ou multidirecional (FERNANDES 2006: 91, tradução nossa).

Levando-se em conta os critérios de classificação acima, este corpus é: tradutório; restrito (domínio Autoconhecimento e Espiritualidade); escrito; paralelo (composto de textos fonte e de suas traduções); sincrônico; bilíngue; unidirecional (tradução na única direção do inglês para o português).

Quanto ao tamanho, este pode ser classificado como um corpus minúsculo (ASTON, 1997 apud SARDINHA, 2002: 108), pois tem menos de 20 mil palavras (o subcorpus em inglês tem 7.918 palavras e o subcorpus em português, 8.435; um total de 16.353 palavras).

Levando em consideração as questões de **uso/registro** de linguagem aplicados à tradução (HATIM & MASON, 1990), é possível identificar o seguinte sobre o corpus paralelo estudado:

Figura 1- Registro do corpus paralelo.



ATIVIDADE ACONTECENDO (CAMPO)	PARTICIPANTES (RELAÇÕES)	LINGUAGEM (MODO)
Website de divulgação da MMS Brasileira e dos cursos oferecidos – atividades: recomendar e expor no campo de conhecimento "Espiritualidade e Esoterismo"	MMS Brasileira e seus estudantes em potencial – reduzir formalidade para uma aproximação	Texto escrito – evitar excesso de coloquialismos

Fonte: Monografia de graduação (PAES, 2016: 23).

2.3. Funções de Linguagem

Os seis fatores constitutivos do processo linguístico/ato de comunicação verbal, segundo Jakobson (2008: 122-123) são: remetente; mensagem; destinatário; contexto; código; contato ("um canal físico e uma conexão psicológica entre o remetente e o destinatário [...]").

> Cada um desses seis fatores determina uma diferente função da linguagem. Embora distingamos seis aspectos básicos da linguagem, dificilmente lograríamos, contudo, encontrar mensagens verbais que preenchessem uma única função. A diversidade reside não no monopólio de algumas dessas diversas funções, mas numa diferente ordem hierárquica de funções. A estrutura verbal de uma mensagem depende basicamente da função predominante. (Jakobson, 2008: 123)

As correspondências entre as funções de linguagem e os seis fatores do ato comunicativo podem ser observadas no quadro da figura 2 abaixo, conforme as posições dos itens no quadro. Por exemplo, a função de linguagem orientada para o contexto é a referencial.

Figura 2 - quadro de correspondências entre seis fatores e funções da linguagem (Jakobson, 2008: 123; 129)

	CONTEXTO			REFERENCIAL	
REMETENTE	MENSAGEM	DESTINATÁRIO	EMOTIVA	POÉTICA	CONATIVA
	CONTATO			FÁTICA	
	CÓDIGO	p. 123		METALINGUÍSTICA	p. 129

Quando se fala em "recomendar" (ver figura 1), temos a função de linguagem conativa ou apelativa, segundo a classificação das funções de linguagem de Roman Jakobson (2008) e "expor" (ver figura 1) corresponde à função referencial ou denotativa. Isso porque recomendar é uma atividade típica dos

anúncios, das críticas literárias ou cinematográficas, livros de autoajuda e todo tipo de propaganda que são textos voltados aos leitores/ouvintes, ou, "destinatário", termo usado por Jakobson (2008: 125).

Mesmo que sutil, a função essencial dos textos do website é anunciar a existência da sede sul-americana da MMS (uma espécie de filial da sede internacional ocidental – *headquaters* - de Toronto) no Brasil e recomendar ao(à) leitor(a) a experiência de frequentar os cursos desta instituição de ensino espiritual como algo com potencial de lhe proporcionar muitos benefícios. A propaganda do website convida o(a) leitor(a) a tornar-se estudante e, por meio da função referencial, são fornecidas informações que lhe auxiliam a escolher quais cursos deseja frequentar. Os textos informativos realizam a função referencial, porém aqui a função referencial está a serviço da conativa: as informações visam a recomendar ao(à) leitor(a) a experiência dos cursos oferecidos, pois ao ler sobre os cursos, o seu interesse é despertado por afinidade com o conteúdo, de modo que a oportunidade de estudar com a MMS se torna desejável.

Os textos fonte em inglês do website canadense da MMS, conforme a análise automatizada, em termos numéricos (contagem de ocorrências – *tokens*), são escritos em sua maior parte com a função referencial ("expor"), com orientação para o "contexto" ocorrendo na maior parte do texto (Jakobson, 2008: 122-123) e, portanto, os textos fonte visam a informar o leitor. Os textos fonte oferecem informações sobre: os cursos da instituição de formação espiritual Escola Moderna de Mistérios Internacional (*headquaters* de Toronto), a origem da MMS e a biografia de seu fundador, Gudni Gudnason. A função referencial é muito comum em textos como artigos científicos, textos enciclopédicos, livros didáticos e apostilas. Vale reforçar que, neste caso, a função referencial é um meio para um fim, um modo de colaborar para a realização do propósito dos textos do corpus: divulgar os cursos da MMS e anunciar a presença da MMS no Brasil, com sede em Florianópolis-SC.

2.4. Skopostheorie: a teoria do escopo/propósito ou da finalidade da tradução

De modo semelhante às funções de linguagem de Jakobson (2008), existe na escola dos teóricos funcionalistas nos Estudos da Tradução uma visão de tipologias textuais do ponto de vista do modo de uso da linguagem e do propósito do texto. Os três tipos de texto são: informativo, expressivo e operativo (Reiss & Vermeer, 2014: 182). Se compararmos com o modelo de Jakobson, o texto do tipo expressivo pode ser entendido como uma combinação das funções emotiva e poética, o texto do tipo informativo corresponde à predominância da função referencial e o operativo, à função conativa ou apelativa. Tanto para Jakobson e Reiss e Vermeer os tipos puros de textos classificados conforme função/tipo costumam ser menos comuns que textos híbridos ou com trechos em que predominam funções/tipos diferentes.

É importante dar-se conta de que o propósito, ou *Skopos*, dos textos traduzidos e dos textos fonte são parecidos, com a diferença de que os textos traduzidos buscam anunciar não somente os cursos, mas também a presença da MMS no Brasil, muito mais recente que a presença da MMS no Canadá. A comissão (também conhecida como "encomenda") de tradução tem a finalidade de divulgar e anunciar a MMS e seus cursos para estudantes em potencial. Conforme a teoria tradutória funcionalista *Skopostheorie* (a teoria do escopo ou da finalidade da tradução), o conceito de comissão/encomenda de tradução pode ser entendido da seguinte maneira:

> Uma pessoa traduz como resultado ou de sua própria iniciativa ou da iniciativa de outra pessoa: em ambos os casos, isto é, alguém traduz de acordo com uma "comissão" *(Auftrag)*.
>
> Deixe-nos definir a comissão como a instrução, dada por si mesmo ou por outro alguém, para realizar uma dada ação – aqui: traduzir. (Ao longo do presente artigo, tradução também inclui interpretação.)
>
> Hoje em dia, na prática, comissões normalmente são dadas explicitamente *(Por favor traduza o texto anexo)*, apesar de raramente ser dito algo referente ao propósito final do

> texto. Na vida real. a especificação de propósito, público-alvo, etc. é na maioria das vezes suficientemente aparente a partir da própria situação em que a comissão é dada [...]. (Vermeer 2012: 198-199, grifos do autor, tradução nossa)[4]

Além disso, dada a combinação das funções referencial e conativa, é possível dizer que o escopo não é homogêneo nos subcorpora – textos fonte e textos traduzidos – havendo segmentos com funções diferentes e, portanto, a finalidade dos segmentos varia de acordo com a função de linguagem ou tipo textual. Segundo Vermeer (2012: 192), "O conceito de escopo também pode ser usado no que diz respeito a segmentos de um *translatum* [texto traduzido], onde isso parece razoável ou necessário. Isso nos permite afirmar que uma ação e, portanto, um texto, não precisa ser considerada um todo indivisível" (tradução nossa)[5].

Tendo em vista a comissão/encomenda de tradução, percebe-se que para preservar a finalidade ou o propósito geral dos textos fonte, necessita-se da manutenção da função conativa nos textos traduzidos. Isso pode ser observado neste estudo através da ocorrência semelhante dos pronomes "*you*" e "você" no corpus, bem como através da tradução da maioria absoluta dos verbos no modo imperativo em inglês por verbos no mesmo modo verbal em português.

---

4 No texto fonte, lê-se: "One translates as a result of either one's own initiative or someone else's: in both cases, that is, one acts in accordance with a "commission" (Auftrag).
Let us define a commission as the instruction, given by oneself or by someone else, to carry out a given action—here: to translate. (Throughout the present article translation is taken to include interpretation.)
Nowadays, in practice, commissions are normally given explicitly (Please translate the accompanying text), although seldom with respect to the ultimate purpose of the text. In real life, the specification of purpose, addressees etc. is usually sufficiently apparent from the commission situation itself […]".

5 No texto fonte, lê-se: "The skopos concept can also be used with respect to segments of a translatum, where this appears reasonable or necessary. This allows us to state that an action, and hence a text, need not be considered an indivisible whole."

3. Métodos

3.1 Compilação do corpus

Segundo Tagnin (2015: 27), antes de mais nada, ao começar a compilar um novo corpus, é preciso definir o objetivo da compilação desse corpus. Neste trabalho, o objetivo geral foi fazer um experimento aplicando conhecimentos adquiridos ao cursar a disciplina FLM 5245-6 Linguística de Corpus[6] a um conjunto de textos fonte e textos traduzidos já disponíveis para uso, sobre o qual já se tinha conhecimento.

Os objetivos específicos não foram definidos previamente, pois derivaram de um levantamento inicial através da pesquisa automatizada que encontrou uma alta ocorrência dos pronomes "*you*" e "você". A partir desse levantamento inicial, surgiram dois objetivos: 1) verificar se a alta ocorrência desses pronomes, possível indicativo da predominância da função de linguagem conativa, coincidiria com uma alta ocorrência de verbos no imperativo, outro indicativo da presença da função conativa, ou seja, buscar dados que comprovassem a predominância da função conativa; 2) verificar através da observação da tradução dos verbos no modo imperativo se houve uma tendência a manter os elementos da função conativa nos textos traduzidos, em comparação com os textos fonte - uma tentativa de averiguar se o tipo textual operativo foi preservado no processo tradutório.

Conforme os parâmetros de composição de corpus em Tagnin (2015: 27-28), o corpus deste estudo é caracterizado da seguinte forma: a) corpus estático, sem atualizações; b) apenas textos escritos; c) bilíngue; d) paralelo; e) textos dos websites MMS e MMS Brasil; f) 13 textos de cada website;

6 A pesquisa que originou este artigo foi feita para redigir o trabalho final da disciplina de pós-graduação na USP "FLM 5245-6 Linguística de Corpus", cursada no primeiro semestre de 2021 como disciplina isolada (aluna especial) antes do ingresso no mestrado. As professoras Stella Esther Ortweiler Tagnin e Luciana Carvalho Fonseca ministraram a disciplina.

g) domínio Autoconhecimento e Espiritualidade; h) fonte: internet; i) textos completos; j) o tamanho foi estabelecido pelo número de textos, ao todo, 26 (13 textos fonte e seus 13 respectivos textos traduzidos).

3.2 Tentativas de alinhamento

No caso de um corpus paralelo, faz-se necessário o alinhamento dos subcorpora. Foram feitas diversas tentativas de alinhamento com softwares de análise de corpus. O alinhamento de corpus paralelo é um procedimento ainda pouco aperfeiçoado pelos softwares gratuitos voltados para a linguística de corpus aqui utilizados. Houve problemas por conta de acentuação (WordSmith Tools), erros de upload (Sketch Engine), falhas ao pesquisar o corpus (AntPConc), conforme será enumerado abaixo na ordem sequencial das tentativas e resultados.

A. **Arquivos TXT** – fez-se a segmentação dos subcorpora por frase/período (na tradução, não foram omitidas sentenças e não foram unidas 2 frases em 1 período, portanto o alinhamento deveria ser simples).

B. **AntPConc** – não houve alinhamento de fato, era possível pesquisar palavras, mas a numeração das linhas de concordância não correspondia entre os dois idiomas e os resultados das buscas eram falhos.

C. **WordSmith Tools** (SCOTT, 2020) – neste programa ocorreu a desconfiguração[7] das palavras com acento no subcorpus em português, mesmo colocando nas configurações que aquele arquivo de texto estava em português e o outro arquivo em inglês, mas houve alinhamento, apesar da pouca legibilidade em português.

---

7 Pode ser que a desconfiguração fosse solucionada salvando-se os arquivos TXT (no programa Bloco de Notas do Windows) como UTF8, mas só se soube disso após a pesquisa estar concluída e o período de trial test gratuito do WordSmith Tools versão 8.0 ter expirado, o que dificultou acessar outra vez essa versão do software para refazer a compilação.

D. **Sketch Engine** – a tentativa de compilar "*multilingual corpus*" falhou em várias tentativas com vários formatos de arquivo, na última tentativa, passou 12 horas compilando sem sucesso; com isso, não houve acesso às ferramentas "*Parallel Concordance: translation search*" (alinhador) e "*Bilingual Terms: bilingual terminology extraction*"[8].

E. **Planilha Eletrônica Excel** – o alinhamento correu bem, mas houve dificuldade na busca por palavras, já que o programa mostra a frase sem colocar em destaque a palavra ali encontrada, de modo a não formar linhas de concordância propriamente ditas.

F. Transposição da planilha do Excel para uma tabela no **editor de textos Word** – o alinhamento precisou de ajustes, mas, em compensação, a busca por palavras foi mais satisfatória que no programa Excel, pois as palavras buscadas ficaram em destaque.

Apesar de a pesquisa automatizada inicial dos subcorpora, que levantou dados que instigaram esta pesquisa do artigo, ter sido realizada no AntConc (ANTHONY, 2020) com o uso dos arquivos TXT, em que se notou pela primeira vez a alta ocorrência dos pronomes "*you*" e "você", recorreu-se a outro programa. De modo que pesquisa foi aprofundada com a transformação dos arquivos TXT do corpus em planilhas Excel, as quais puderam ser analisadas mais detalhadamente pelo programa Sketch Engine (Kilgarriff et al, 2004).

3.3 Uso das ferramentas

Portanto, conforme o relato da seção anterior, depois de tentar usar o AntConc (ANTHONY, 2020) e o WordSmith Tools (Scott, 2020), optou-se pelo Sketch Engine (Kilgarriff et al, 2004). Na utilização da ferramenta Keyword list, recorreu-se aos corpora de referência oferecidos pelo Sketch Engine: para o subcorpus em inglês, usou-se o corpus de referência *English Web* 2020 (enTenTen20); em português, *Portuguese Web* 2011 (ptTenTen11).

8 Esta é uma função de extração de termos.

A ferramenta *Wordlist* foi a mais usada (figura 3 abaixo). Ao consultar a lista de palavras do subcorpus em inglês, percebeu-se a alta frequência do pronome "*you*" (sétima palavra mais frequente, pronome mais frequente, com 142 ocorrências). A alta frequência desse pronome chamou atenção para a presença da função conativa no texto já que a função conativa explora o uso do discurso em 2ª e 3ª pessoa, utilizando pronomes como tu e você. Como a **função conativa** possui foco sobre o interlocutor, destinatário da mensagem comunicada, de modo que busca engajá-lo como se autor e leitor estivessem interagindo, e normalmente o remetente/autor possui o intuito de recomendar algo, essa função de linguagem **utiliza-se de verbos no modo imperativo.** Decidiu-se então o foco desta pesquisa: o estudo dos verbos no imperativo. Utilizou-se a função avançada da Wordlist: busca de verbos ao fornecer lista retirada do corpus em inglês.

Figura 3 - Busca avançada de verbos.

Fonte:<https://app.sketchengine.eu/>.

Seria possível buscar apenas no documento Word ou na planilha Excel, mas o Sketch Engine já etiqueta o corpus de modo que é possível detectar quando uma mesma palavra funciona como verbo ou outra classe gramatical. Portanto a busca avançada (ver figura 3 acima) agiliza a busca, além de facilitar a visualização ao criar linhas de concordância, de modo que, no Sketch

Engine, a visualização é mais clara que em tabelas Word ou planilhas Excel. Por uma questão de disponibilidade de tempo, restringiu-se a busca aos verbos com ao menos 2 ocorrências (137 verbos).

Na planilha Excel em que o corpus paralelo foi alinhado, buscavam-se os verbos no imperativo em inglês identificados inicialmente no Sketch Engine, os quais foram formatados em negrito e vermelho para facilitar a identificação e localização. Os verbos traduzidos como imperativo, também foram formatados em vermelho e negrito; já a tradução de verbos no imperativo por formas diferentes do verbo no imperativo ocorreu raramente, e essas ocorrências foram formatados em negrito azul nas tabelas. A partir disso, foram elaboradas as tabelas 1 e 2, que estão na seção seguinte que trata dos resultados.

4. Resultados

A busca de verbos no imperativo nos subcorpora dos textos fonte e dos textos traduzidos visou a esclarecer se esses verbos foram traduzidos do inglês para o português, mantendo o mesmo modo verbal, de modo a preservar a função de linguagem conativa/apelativa ou não no processo tradutório. Como se verificou que o modo verbal se manteve na tradução, isso indica que a função conativa deve ter sido preservada e, portanto, acredita-se que a finalidade da encomenda de tradução do tipo textual operativo foi atendida, produzindo no público-alvo dos textos traduzidos um efeito bastante semelhante àquele produzido nos leitores dos textos fonte.

Vale lembrar que o corpus paralelo é o conteúdo de websites de divulgação de uma instituição de ensino internacional e que esse conteúdo visa a informar sobre a MMS e seus cursos, a fim de recomendar os cursos como experiências positivas para alunos em potencial, sendo, em última análise, material de publicidade, o exemplo mais comum do tipo textual operativo.

Os verbos escolhidos para a análise foram obtidos através do uso da ferramenta *Wordlist* do software on-line Sketch Engine e buscou-se em linhas de concordância por verbos da lista de verbos do subcorpus em inglês, neste caso, verbos com ao menos duas ocorrências (137 verbos - *types*).

Notou-se a predominância da função referencial, pois dentre esses 137, apenas 19 verbos (*types*) foram conjugados no imperativo com um total de 36 ocorrências de verbos (*tokens*) conjugados no imperativo em inglês, conforme consta na primeira coluna da última linha da tabela 2 abaixo. Apenas 2 verbos no imperativo foram traduzidos sem uso do imperativo em português (verbo "*know*", linha 340 da tabela 1; "*master*", linha 324 da tabela 2) e na linha 434 nas tabelas 1 e 2, há dois verbos no imperativo, "*speak*" e "*contact*", mas na tradução há apenas um verbo no imperativo, "confira". Ao encontrar um verbo imperativo, na mesma linha de concordância, podia-se encontrar mais outro verbo no imperativo, isso ocorre em 6 linhas (por exemplo, a linha 434, mencionada acima), as quais estão registradas uma vez para cada verbo, portanto, estão duplicadas e foram destacadas em amarelo na última coluna das tabelas 1 e 2. Vale notar que a coluna "Total de oc." registra quantas vezes cada verbo ocorre (*tokens*) no subcorpus em inglês.

Tabela 1 – Ocorrências de verbos no imperativo no corpus.

	Verbo	Total de oc.	Oc. no imp.	Frase	Frase traduzida	Nº da linha no corpus
1	Be	364	1	If you would like more information about [...] please **be** sure to contact us.	Se você tiver interesse em obter mais informações sobre [...] por favor, não **hesite** em nos contatar.	66
2	Know	24	6	**Know** Thyself [*subtitulo*]	**Conhece**-te a ti mesmo	69
				[...] the ancient decree of these esoteric schools has always been '**Know** Thyself'.	[...] o antigo decreto dessas escolas esotéricas sempre foi "**Conhece**-te a ti mesmo".	110
				Know Thyself [*subtitulo*]	**Conhece**-te a ti mesmo	174
				Once you've completed the Empower Thyself I & II programs, you move to the **Know** Thyself program.	Depois de ter concluído os cursos [...], poderá prosseguir e fazer o curso "**Conhece**-te a ti Mesmo".	255
				Initiating Teachers (**Know** Thyself Program) into the Great Brotherhood and Sisterhood of Light.	Iniciar Professores (**Segunda Iniciação**) na Grande Fraternidade e Irmandade de Luz.	340
				Through this study we [...] fulfilling the ancient decree: **Know** Thyself!	Através deste estudo, [...], cumprindo o antigo decreto: **Conhece**-te a ti mesmo!	445
3	Take	17	1	**Take** a moment and look around you right now.	**Pare** por um instante e agora olhe ao seu redor.	180
4	Empower	14	10	**Empower** Thyself [*nome de curso*]	**Empodera**-te a ti Mesmo Linhas: 234, 234, 249, 253, 253, 255, 292, 293, 299, 339	
5	Learn	13	1	**Learn** the Truth [*título de página do site*]	**Aprenda** a Verdade	202
6	See	13	1	(**see** Golden Pyramid of Peace)	(**veja** a Pirâmide Dourada da Paz)	58
7	Build	5	1	**Build** and refine your intuition and guidance to stay more consistently and fully aligned with your truth.	**Construa** e refine sua intuição e sua receptividade às orientações dos guias espirituais [...].	378
8	Contact	4	3	For a list of certified instructors please click here or **contact** us.	Para acessar uma lista de instrutores certificados, por favor clique aqui ou **entre** em contato.	321
				For more information on this program please speak to a local Certified Guide or **contact** us today!	**Confira** nossa AGENDA aqui - gostaria de me cadastrar.	434
				If you would like to receive more information [...], please **contact** us or check our calendar.	Se você tiver interesse em receber mais informações sobre [...], por favor, **entre** em contato conosco ou confira nossa agenda.	467

Fonte: própria autora.

Tabela 2 - Ocorrências de verbos no imperativo no corpus (continuação).

9	Enhance	4	1	Crystal Dreaming – **enhance** your ability to recall your dreams.	Sonhando com Cristais – **potencialize** sua habilidade de recordar sonhos.	366
10	Join	4	2	**Join** us for a day that will give you practical tools you can use immediately to enhance your health, home, and life.	**Junte**-se a nós por um dia de curso em que vai receber ferramentas práticas que você pode usar imediatamente para melhorar a sua saúde, o seu lar e a sua vida.	372
				Join this full day of journeys to explore yourself and build your "spiritual muscles"!	**Inscreva**-se para participar destas viagens para explorar a si mesmo(a) e fortalecer seus "músculos espirituais" durante um dia inteiro!	383
11	Look	3	1	Take a moment and **look** around you right now.	Pare por um instante e agora **olhe** ao seu redor.	180
13	Master	2	1	Third Step Initiation - Guide (**Master** Thyself) [*nome de curso*]	Terceira Iniciação – Guia (**Maestria** Pessoal)	324
14	Observe	2	1	**Observe** what you see.	**Observe** o que você vê.	181
15	Refine	2	1	Build and **refine** your intuition and guidance to stay more consistently and fully aligned with your truth.	Construa e **refine** sua intuição e sua receptividade às orientações dos guias espirituais para manter-se em coerência e em alinhamento total com a sua verdade.	378
16	Click	2	2	For more information on healings offered by certified MMS practitioners, **click** here.	Para mais informações sobre terapias holísticas oferecidas por profissionais certificados da Escola Moderna de Mistérios, **clique** aqui.	239
				For a list of certified instructors please **click** here or contact us.	Para acessar uma lista de instrutores certificados, por favor **clique** aqui ou entre em contato.	321
17	Let	1	1	**Let's** make sure she is taken care of by living a prosperous and green life!	**Vamos** nos assegurar de que ela seja bem cuidada se vivermos uma vida próspera e verde!	45
18	Speak	1	1	For more information on this program please **speak** to a local Certified Guide or contact us today!	**Confira** nossa AGENDA aqui - gostaria de me cadastrar.	434
19	Check	1	1	If you would like to receive more information [...], please contact us or **check** our calendar.	Se você tiver interesse em receber mais informações [...], por favor, entre em contato conosco ou **confira** nossa agenda.	467
Total de ocorrências no imp.=36				Ocorrência não trad. por imp.=2	2 verbos no imp. juntos numa linha= 6 linhas	

Fonte: própria autora.

Em síntese: 1. total de ocorrências examinadas foi de 137 verbos (*types*), com 1177 ocorrências (*tokens*) - 88,96% do total do subcorpus em inglês (em ocorrências/*tokens*); 2. total de ocorrências verbais no subcorpus dos textos fonte (em inglês): 283 *types*, 1323 *tokens*; 3. dentre as ocorrências examinadas no subcorpus em inglês (13 textos fonte); 3.a. verbos no imperativo - 19 verbos (*types*), 36 ocorrências de verbos (*tokens*) - 3,06% das 1177 ocorrências (*tokens* examinados); 3.b. ocorrência de verbos em outros modos verbais diferentes do imperativo - 96,94%; 4. Comparando os subcorpora de textos traduzidos e textos fonte – 4.a. tradução por verbo em modo diferente do imperativo - 2 verbos dentre 36 no imperativo - 5,55%; 4.b. ocorrências em que os textos traduzidos preservaram o imperativo - 34 de 36 ocorrências (*tokens*) - 94,45% das ocorrências de verbos no imperativo examinadas no corpus.

5. Discussão dos Resultados

Quanto ao objetivo geral, foi um experimento bem-sucedido, pois esta pesquisa foi uma oportunidade enriquecedora de aplicação de conhecimentos básicos de Linguística de Corpus.

No que tange os objetivos específicos:

1) Apesar de os dados numéricos indicarem a predominância da função referencial, pelo menos referente à conjugação verbal, não se deve desprezar o contexto comunicativo, visto que o corpus é o conteúdo de websites de divulgação em que a função geral é a conativa, ou seja, os dados contrariam a visão mais ampla do corpus.

2) Foi demonstrado que a tradutora manteve sempre que possível os verbos no imperativo ao traduzir do inglês para o português, de modo a colaborar para a manutenção da finalidade geral do texto que se identifica com a função conativa do texto fonte.

Outra forma de preservar a função conativa foi a manutenção aproximada da frequência do pronome de tratamento “você” (décima palavra mais

frequente no subcorpus em português), visto que o equivalente em inglês "you" é a sétima palavra mais frequente. Foi mantida essa alta frequência de "você" mesmo havendo tendência a omitir pronomes em português como convenção da língua escrita, convenção essa que busca evitar o excesso de repetições, utilizando-se, por exemplo, da elipse de sujeito.

6. Considerações Finais

Este estudo demonstrou como podemos abordar um mesmo conjunto de textos que já foi estudado antes e fazer novas descobertas por meio das ferramentas de análise da Linguística de Corpus. A abordagem deste estudo foi do tipo *bottom-up*, pois não havia uma questão de pesquisa antes da análise do corpus, havia a princípio apenas um objetivo geral de observar estratégias tradutórias sob um novo ângulo, diferente do trabalho em Paes (2016). Com a observação inicial da Lista de Palavras, a alta taxa ocorrência do pronome "you" deu início a todo o trabalho registrado neste artigo.

Outras questões podem vir a ser investigadas, tal como os colocados (*collocates*) de "*you*" e "você" que podem vir a ser estudados com a ferramenta *N-grams*. As palavras de conteúdo da Lista de Palavras que indicam que a instituição Brasil MMS se dedica à atividade educativa podem ser o início de outra(s) pesquisa(s): *information, knowledge, wisdom, teach, taught, teachings, teachers, learn*. Poderá ser investigado ainda o tema da preferência semântica e da prosódia semântica, tão importante quanto nos dizem pesquisadores como Sinclair (2004), de modo que esse tema pode ser trabalhado em estudos futuros.

Neste estudo, percebeu-se que uma função de linguagem – neste caso, a referencial – pode dar suporte à função geral mesmo que sejam funções distintas, até porque a finalidade da tradução pode variar dentro de um mesmo texto em seus diferentes segmentos, de modo que sempre se deve levar em conta o propósito geral de uma encomenda de tradução – neste caso, divulgar e anunciar uma instituição educativa e seus cursos - antes de iniciar o processo tradutório.

Referências bibliográficas

Anthony, Laurence. (2020). AntConc (Version 3.5.9) [Computer Software]. Tokyo, Japan: Waseda University. Disponível em: <https://www.laurenceanthony.net/software/antconc/>. Acesso em: 18 jul. 2023.

Baker, M. Corpus linguistics and translation studies: Implications and applications. In: Baker, M.; Francis, G.; Tognini-Bonelli, E. (Eds.). *Text and Technology: in honour of John Sinclair*. 1ª ed. Amsterdã/Philadelphia: John Benjamins, 1993: 233-250.

Beeby, A. et al. First Results of a Translation Competence Experiment: 'Knowledge of Translation' and 'Efficacy of the Translation Process'. In: Kearns, J. (Ed.). *Translator and Interpreter Training: Issues, Methods and Debates*. Londres: Continuum, 2008: 104-126.

Fernandes, L. Corpora in Translation Studies: revisiting Baker's typology. 2006. *Fragmentos*, n. 30, Florianópolis, jan-jun/2006, pp. 87-95. Disponível em: <https://periodicos.ufsc.br/index.php/fragmentos/article/view/8217>. Acesso em: 18 jul. 2023.

Hatim, B.; Mason, I. Context in Translating: Register Analysis. In:_____. *Discourse and the Translator*. 1ª ed. Londres e Nova York: Longman, 1990: 36-54.

Jakobson, R. *Linguística e comunicação*. 19ª ed. São Paulo: Cultrix, 2003 [1969].

KilgarriFf, A. et al. Itri-04-08 the sketch engine. *Information Technology*, 2004. Software SKETCH ENGINE. Disponível em: <http://www.sketchengine.eu>. Acesso em: 25 jun. 2021.

Paes, A. C. M. *Competência Tradutória*: o estudo de um caso real de tradução em uma organização internacional. Florianópolis: Universidade Federal de Santa Catharina, 2016. Disponível em: <https://secretariado.ufsc.br/2016/07/11/competencia-tradutoria-o-estudo-de-um-caso-de-traducao-em-uma-organizacao-internacional/>. Acesso em: 18 jul. 2023.

Reiss, K.; Vermeer, H. J. *Towards a general theory of translational action*: skopos theory explained. Nova York: Routledge, 2014.

Sardinha, T. B. Tamanho de Corpus. *The Especialist*, n. 2/ v. 23, São Paulo, abr 2003, pp. 103-122. Disponível em: <https://revistas.pucsp.br/esp/article/download/9381/6952/23377>. Acesso em: 18 jul. 2023.

Scott, M., 2020, WordSmith Tools version 8, Stroud: Lexical Analysis Software.

Sinclair, J. *Trust the Text:* language, corpus and discourse. Londres/Nova York: Routledge, 2004.

Tagnin, S. E. O. A Linguística de Corpus na e para a Tradução. In: Viana, V.; tagnin, S. (orgs.) *Corpora na Tradução*. São Paulo: HUB, 2015: 19-56.

Vermeer, H. J. Skopos and Comission in Translational Action. Tradução Andrew Chesterman. In: *The Translation Studies Reader*. Venuti, L. (Ed.). Nova York: Routledge, 2012 [2000]: 191-202.

Análise da tradução de termos da área de Meteorologia da Fraseologia Padrão Aeronáutica

Analysis of the translation of aeronautical standard phraseology terms in the field of Meteorology

Patrícia Tosqui-Lucks[1]
Rafaela Araújo Jordão Rigaud Peixoto[2]

1 Instituto de Controle do Espaço Aéreo (ICEA), São José dos Campos, SP, Brasil.
2 Departamento de Controle do Espaço Aéreo (DECEA), Rio de Janeiro, RJ, Brasil.

***Resumo**:* A Fraseologia Padrão Aeronáutica é composta por um conjunto de frases e expressões pré-estabelecidas cujo objetivo é permitir a troca de informações entre pilotos e controladores de tráfego aéreo de forma clara, concisa e precisa, em situações normais de voo, e costuma ser suficiente para lidar com a maioria das situações encontradas na prática diária do controle de tráfego aéreo. Porém, em situações inesperadas ou não usuais, como, por exemplo, em condições meteorológicas adversas, essa linguagem pode se mostrar limitada para a comunicação (PEIXOTO; PIMENTEL, 2021). Nesses casos, é necessário recorrer à linguagem comum, que também deve ser restrita ao contexto aeronáutico e submetida aos mesmos padrões que caracterizam a fraseologia aeronáutica (SCARAMUCCI; TOSQUI-LUCKS; DAMIÃO, 2018). Neste artigo, analisaremos alguns exemplos de traduções de fraseologia, vertidas da língua portuguesa para a inglesa, referentes a termos da área da Meteorologia, além de exemplos de fraseologismos identificados no uso de *plain language*. A análise, realizada utilizando os procedimentos metodológicos da Linguística de Corpus, por meio do software AntConc, consiste em busca por *clusters* de termos equivalentes em um corpus composto por Fraseologias elaboradas originalmente em língua inglesa e amplamente utilizadas pela instituição norte-americana *Federal Aviation Administration* (FAA), pela instituição europeia *Eurocontrol* e pela organização internacional OACI. As traduções identificadas como imprecisas ou problemáticas receberão propostas de traduções mais adequadas, com base nos documentos consultados.

***Palavras-chave**:* Terminologia; Meteorologia Aeronáutica; Fraseologia.

***Abstract**:* The Aeronautical Standard Phraseology is composed of a set of pre-established phrases and expressions whose objective is to allow the exchange of information between pilots and air traffic controllers in a clear, concise and precise way, in regular flight situations, and is usually sufficient to handle most situations encountered in the daily practice of air traffic control. However, in unexpected or unusual situations, such as, for example, in adverse weather conditions, this language can be limited for communication (PEIXOTO; PIMENTEL, 2021). In these cases, it is necessary to resort to common language, which must also be restricted to the aeronautical context and submitted to the same standards that are unique to aeronautical phraseology (SCARAMUCCI; TOSQUI-LUCKS; DAMIÃO, 2018). In this paper, we will analyze some examples of phraseology translations, from Portuguese to English, referring to terms in the Meteorology area, in addition to examples of phraseologisms identified in the use of plain language. The analysis, carried out using the methodological procedures of Corpus Linguistics and the AntConc software, consists of a search for clusters of equivalent terms in a corpus composed of Phraseology originally written in English and broadly used by the US institution Federal Aviation Administration (FAA), by the European institution Eurocontrol and by the international organization ICAO. Translations identified as inaccurate or problematic will receive proposals for more appropriate translations, based on the documents consulted.

***Keywords**:* Terminology; Aeronautical Meteorology; Phraseology.

1. A fraseologia aeronáutica

A comunicação que ocorre entre pilotos e controladores de tráfego (ATCO) apresenta características e peculiaridades que a diferenciam de outras comunicações no contexto da aviação. Ela tem por base a fraseologia padrão aeronáutica, composta por um conjunto de frases e expressões pré-estabelecidas que utilizam um vocabulário de aproximadamente 400 palavras, no qual artigos, pronomes, verbos de ligação, verbos auxiliares e algumas preposições costumam ser excluídas ou evitadas. Cerca de 50% das frases estão no imperativo ou na voz passiva e as nominalizações são privilegiadas. Os ATCOs brasileiros devem seguir o Manual de Fraseologia publicado pelo Comando da Aeronáutica (MCA 100-16, 2020), que apresenta fraseologia bilíngue, com as frases em português e seus equivalentes em inglês ao lado, como veremos nas próximas seções. Esse documento é periodicamente revisado e a versão atualmente vigente foi publicada em 2020.

O objetivo dessa fraseologia é permitir a troca de informações de forma clara, concisa e precisa, em situações normais de voo e suficiente para lidar com a maioria das situações rotineiras de controle de tráfego aéreo. Porém, em situações inesperadas ou não usuais, como em condições meteorológicas adversas, essa linguagem pode mostrar-se limitada para a comunicação. Nesses casos, é necessário recorrer à linguagem comum (*plain language*), que também deve ser restrita ao contexto aeronáutico e submetida aos mesmos padrões de concisão, precisão e clareza que caracterizam a fraseologia aeronáutica (OACI, 2010). No caso de voos internacionais, a língua inglesa é considerada a *lingua franca*, sendo que o "inglês aeronáutico" (SCARAMUCCI, TOSQUI-LUCKS e DAMIÃO, 2018; TOSQUI-LUCKS; SILVA, 2020a e 2020b; TOSQUI-LUCKS; PRADO, 2020) é o conjunto de interações trocadas por esses profissionais, caracterizado pelo uso da Fraseologia Padrão combinado com o uso de uma linguagem comum (*plain language*) que extrapola a fraseologia, nos momentos em que ela não é suficiente. Passaremos a uma melhor diferenciação desses dois conceitos.

1.1. Fraseologia e *Plain English*

Segundo o Doc 9835, itens 3.2.5 a 3.2.7, fraseologia é uma "sublíngua/sublinguagem"[3] caracterizada por fórmulas específicas e vocabulário especializado, ou "o código formulaico feito de palavras específicas que, em contextos de operações aeronáuticas, têm um significado operacional singular e preciso" (OACI, 2010, inciso 6.2.8.4, p. 6-6, tradução nossa)[4]. Tosqui-Lucks e Prado (2020) complementam tal definição, acrescentando que a fraseologia permite a troca de informações exclusivamente entre dois profissionais da aviação, a saber, pilotos e ATCOs, de forma clara, concisa e segura, em situações normais de voo. Encontramos, ainda, outra definição de fraseologia, mais recente e bastante completa, que além de elencar as características linguísticas dessa "ferramenta especializada" (BOROWSKA, 2017 p. 77), também aponta seus usuários e contexto de uso. Segundo a autora, a fraseologia aeronáutica padrão é uma ferramenta de comunicação aeronáutica que, juntamente com o Inglês Comum Aeronáutico, forma o conjunto de frases e palavras necessários para comunicações aeronáuticas bem-sucedidas. A autora segue afirmando que esse código é utilizado para construir enunciados diretos e significativos usados apenas para propósitos de comunicação aeronáutica de rotina durante operações de aeronaves no solo e no ar, por pilotos, ATCOs ou pessoal de solo, de modo a facilitar a compreensão em ambientes de alto risco, para o melhor desempenho linguístico (BOROWSKA, 2017, p. 77-78, tradução nossa).

Também segundo a OACI, a fraseologia, sempre que possível, deve ser a ferramenta específica a ser utilizada nas comunicações aeronáuticas por radiotelefonia (OACI, 2010, p.4-2, item 4.3.3), mas ela não é suficiente para abarcar a ampla variedade de situações emergenciais que poderão ocorrer durante um voo. Assim sendo, em casos em que a fraseologia padrão não é suficiente, os participantes da interação podem fazer uso de outra sublinguagem do inglês aeronáutico (BOROWSKA, 2017, p. 90), chamada *plain*

3 No original: "sublanguage".

4 No original: "the formulaic code made up of specific words that in the context of aviation operations have a precise and singular operational significance".

English, que é uma subdivisão do termo *plain language*, definido como "o uso espontâneo, criativo e não codificado de uma dada língua natural" (OACI, 2010, p. 3-5, item 3.3.14, tradução nossa)[5].

Como destacam Tosqui-Lucks e Silva (2020a e 2020b), ao contrário da fraseologia aérea padrão, o *plain English*, mesmo na língua inglesa, ainda é um conceito vago e relativamente indefinido. Diferentes autores oferecem definições complementares para esse termo e, em alguns casos, até os termos mudam um pouco, como no caso de '*plain aviation English*' ou '*plain Aeronautical English*'.

Justamente por constituir um dos elementos das comunicações por radiotelefonia, o *plain English* deve ser utilizado segundo as mesmas regras de concisão, precisão, objetividade, inteligibilidade e não ambiguidade que regem o uso da fraseologia (OACI, 2010, p. 3-5, item 3.3.14). Por essa razão, somente pode ser associado a "inglês comum" em oposição ao termo fraseologia, não tendo, de forma alguma, a conotação de inglês para uso em situações comuns do cotidiano (BOROWSKA, 2017, p. 91), tampouco para uso nos demais contextos da aviação, que fujam à comunicação por radiotelefonia.

A fim de compreender como o uso da linguagem especializada ocorre em situações específicas, é preciso compreender não apenas a dinâmica das necessidades e do uso da fraseologia aeronáutica, mas também a recorrência de padrões terminológicos em instituições, tema abordado no próximo tópico.

5 No original: "the spontaneous, creative and non-code use of a given natural language".

2. Padrões terminológicos em instituições

Inicialmente, cumpre destacar que, conforme os princípios da Linguística de Corpus, os padrões terminológicos e fraseológicos são identificados com base na co-ocorrência de termos (colocados, coligações, etc) e na formação de categorias sintagmáticas (HUNSTON, 2010), de forma a indicar usos costumeiros de linguagem especializada e de fraseologia.

Nesse sentido, têm-se que a compreensão do uso de terminologia especializada não depende somente do conhecimento sobre a área temática, mas sobretudo das relações linguísticas próprias do contexto de uso, isto é, do contexto léxico-semântico dos termos (FINATTO, 2001), sobretudo em instituições especializadas.

Isto posto, os padrões normativos da área de Meteorologia Aeronáutica propriamente dita e de Meteorologia Geral aplicada à Navegação Aérea são preconizados por duas principais organizações, responsáveis pela elaboração de regras que devem ser parâmetro para todos os países: a Organização Meteorológica Mundial (OMM), fundada em 1919; e a Organização da Aviação Civil Internacional (OACI), fundada em 1947. A atribuição de cada uma dessas instituições, no âmbito da aviação civil internacional, é definida pelo Doc 7473, publicado em 1963 (PEIXOTO, 2020b).

No caso do Brasil, o Departamento de Controle do Espaço Aéreo (DECEA) é a instituição regulatória oficial para o segmento de Meteorologia Aeronáutica no território brasileiro. Mais recentemente, a gestão das atividades relativas a essa especialidade foi unificada no Centro Integrado de Meteorologia Aeronáutica (CIMAER), criado em 2017 (PEIXOTO, 2020b), subordinado ao DECEA.

Nesse sentido, as instituições OMM e OACI, além da FAA, foram utilizadas como parâmetro para os textos compilados em inglês por Peixoto, em sua pesquisa de pós-doutoramento (2020) em andamento, para a área de Meteorologia Aeronáutica, como explicitado a seguir.

3. Metodologia

As etapas metodológicas deste trabalho compreenderam seleção de textos institucionais em língua inglesa, análise em programa concordanciador (*AntConc)* e busca por *clusters* de equivalentes em inglês. O corpus foi composto por documentos da FAA (08), da OACI (21) e da OMM (30), cujas evidências nortearam as propostas de tradução mais adequadas ao contexto aeronáutico.

4. Análise de termos da fraseologia aeronáutica

Passamos a discutir quatro termos (em português e seu equivalente em inglês), extraídos do Manual de Fraseologia (MCA 100-16) mais recente (DECEA, 2020). Primeiramente, demonstramos o contexto de uso do termo; em seguida, procedemos a uma análise em relação à coerência e à adequabilidade dos termos vertidos para o inglês, conforme consulta ao corpus mencionado acima; e, a partir disso, apresentamos uma sugestão de versão para o inglês.

As ocorrências dos termos (1) 'área de mau tempo', (2) 'área intensa de mau tempo', (3) 'formações pesadas' e (4) 'vento instantâneo', e de suas versões para o inglês no MCA foram analisadas, conforme ilustrado na figura abaixo:

ITENS DE FRASEOLOGIA DO MCA 100-16 (2020)

(1)	FAB 4515, área de mau tempo entre os azimutes 300 e 030, a 50 milhas, deslocamento leste, com 10 nós, topo FL 250.	FAB 4515, adverse weather area between azimuth 300 and 030, 50 miles, moving east at 10 knots, top FL 250.

(DECEA, 2020, p.24)

(2)	FAB 4515, área intensa de mau tempo entre os azimutes 300 e 030, a 50 milhas, deslocamento Leste, com 10 nós, topo FL 250.	FAB 4515, intensive weather area between azimuth 300 and 030, 50 miles, moving east at 10 knots, top FL 250.
	UAL 861, área intensa de mau tempo, 30 milhas à frente.	UAL 861, intensive weather area, 30 miles ahead.
		(DECEA, 2020, p.24-5)
(3)	AAL 7201, formações pesadas reportadas sobre Confins, topo acima do nível 300, reporte se for necessário desvio.	AAL 7201, heavy weather area reported over Confins, top above flight level 300, advise if deviation will be necessary.
		(DECEA, 2020, p.25)
(4)	*TIB 5561 MELO aos 45', FL 330, estima PAG aos 03, FNP próxima, vento instantâneo 270 graus com 59 nós, temperatura menos 45, condições de voo instrumentos no topo, turbulência leve.	*TIB 5561 MELO at 45', FL 330, estimate PAG at 03, FNP next, spot wind 270 degrees at 59 knots, temperature minus 45, flight conditions instruments on top, light turbulence.
		(DECEA, 2020, p.39)

FONTE: Elaborado pelas autoras

Em relação aos termos (1) e (2), o equivalente '*weather area*' apenas aparece em um documento da OMM. No geral, as condições climáticas são referenciadas, no corpus, como '*weather conditions*'; e os termos '*bad weather*' e '*average bad weather*' são usados pelas três instituições, ou seja, são bastante difundidos. Nesse sentido, a comparação de intensidade em relação a diferentes áreas com mau tempo é marcada pela gradação *bad → severe*. Por outro lado, não foram encontradas ocorrências de '*intensive weather*' nem '*intense weather*'.

Quanto ao termo (3), o equivalente *'formation'* geralmente não co-ocorre com adjetivo de intensidade, mas apenas com o indicativo do tipo de formação ('*fog formation*' e '*cloud formation*'). E há ocorrências de '*heavy precipitation*', '*heavy rain*' e '*heavy clouds*' nos documentos da OMM e da FAA.

Em relação ao termo (4) 'vento instantâneo', o equivalente '*spot wind*', apesar de ser um termo estabelecido, apenas apresenta ocorrências em documentos da OMM e da OACI. Na FAA, é utilizado o termo '*sudden wind*'.

Com base nessa discussão e nos dados do corpus, é proposta a seguinte revisão dos termos em inglês: (1) 'área de mau tempo' seria vertido como '*bad weather conditions*'; (2) 'área intensa de mau tempo', como '*severe weather conditions*'; (3) 'formações pesadas', como '*heavy clouds*'; e (4) 'vento instantâneo', como '*spot wind*' ou '*sudden wind*'.

5. Considerações finais

Neste artigo, foram analisados alguns exemplos de traduções problemáticas de Fraseologia Aeronáutica, vertidas da língua portuguesa para a inglesa, referentes a termos da área da Meteorologia, bem como foram propostas versões mais adequadas, com base em parâmetros de documentos publicados por instituições de referência na área. A análise dessas ocorrências foi baseada em procedimentos e princípios da metodologia da Linguística de Corpus.

A versão dos termos estudados para o inglês recorreu a estruturas pouco usuais na língua inglesa, uma vez que, no inglês, a tendência é nomear situações mais concretamente (como no caso de '*heavy clouds*', '*heavy precipitation*' ou '*heavy rain*'). Além disso, verificou-se que termos difundidos em organizações internacionais, tais como OMM e a OACI, não são necessariamente utilizados em organizações nacionais, mesmo que tenham amplo alcance, como a FAA)

Buscamos, com este estudo, apresentar uma contribuição para os estudos terminológicos, na área de especialidade de Meteorologia Aeronáutica. Além disso, acreditamos que a reflexão fomentada seja outra forma de agregar contribuição para a área de Terminologia e de Tradução.

Do ponto de vista de contribuições sociais, as aplicações dessa pesquisa são voltadas para a comunidade usuária da Fraseologia, sobretudo pilotos e controladores de tráfego aéreo, como uma resposta às indagações desses

profissionais quanto à melhor forma de expressar os termos na língua inglesa. Nesse sentido, o artigo traz contribuições para o ensino da Fraseologia e do *plain English* nos treinamentos oferecidos a esses profissionais.

Por fim, as propostas de tradução aqui oferecidas podem ser apresentadas para o Processo de Revisão Normativa (PRENOR), do Departamento de Controle do Espaço Aéreo (DECEA), que recebe e analisa sugestões para melhorias das comunicações aeronáuticas. Assim, acreditamos que o estudo ora realizado extrapola as contribuições para a Linguística, sendo também socialmente relevante para a segurança da aviação no espaço aéreo brasileiro.

Referências bibliográficas

ANTHONY, L. *AntConc* (Version 3.5.8) [Computer Software]. Tokyo, Japan: Waseda University. 2019. Disponível em: <https://www.laurenceanthony.net/software>. Acesso em: 20 fev. 2022.

BOROWSKA, A. P. *Avialinguistics:* The Study of Language for Aviation Purpose. Frankfurt: Peter Lang, 2017. 332 p.

BRASIL. Comando da Aeronáutica. Departamento de Controle do Espaço Aéreo. *ICA 105-12*: Fraseologia Volmet. Rio de Janeiro, 2014. Disponível em: <https://publicacoes.decea.gov.br/?i=publicacao&id=4072>. Acesso em: 20 set. 2021.

BRASIL. Comando da Aeronáutica. Departamento de Controle do Espaço Aéreo. *MCA 100-16*: Fraseologia de Tráfego Aéreo, 2020. Rio de Janeiro. Disponível em: <https://publicacoes.decea.gov.br/?i=publicacao&id=4072>. Acesso em: 10 nov. 2021.

FINATTO, M. DA G. K. *Definição terminológica*: fundamentos teórico-metodológicos para sua descrição e explicação. Doutorado em Estudos da Linguagem. Porto Alegre: Universidade Federal do Rio Grande do Sul, 2001.

HUNSTON, S. How can a corpus be used to explore patterns? In: O'KEEFFE, A.; McCARTHY, M. The Routledge Handbook of Corpus Linguistics. London & New York: Routledge; Taylor & Francis Group, 2010: 152-166.

INTERNATIONAL CIVIL AVIATION ORGANIZATION. *Air Traffic Management*. Doc 4444. Montreal, Canada, 2016.

INTERNATIONAL CIVIL AVIATION ORGANIZATION. *Annex 3 of the Convention on International Civil Aviation*. Meteorological Service for International Air Navigation: parts I and II. 20. ed. Montreal, 2018.

PEIXOTO, R. A. J. R. Aeronautical Meteorology glossary: a discussion on term definition in the ANACpedia termbase. *The Especialist*, v. 42 (1), 2020a.

PEIXOTO, R. A. J. R. Terminology of Aeronautical Meteorology codes: a systematization by using corpus. *TradTerm*, v. 37 (1), 2020b.

PEIXOTO, R. A. J. R.; TOSQUI-LUCKS, P. Weather events in air traffic control standards and communication: discourse patterns and implications for language teaching and assessment. *ReLin*: Revista de Estudos da Linguagem. Belo Horizonte, v. 29 (2), 2021, pp. 1443-1484. Disponível em:<http://periodicos.letras.ufmg.br/index.php/relin/article/view/17491> Acesso em: 20 fev. 2022.

SCARAMUCCI, M. V. R.; P. TOSQUI-LUCKS; S. M. DAMIÃO (Eds.). 2018 *Pesquisas sobre inglês aeronáutico no Brasi*l. Campinas: Pontes Editores, 2018. 313 p.

TOSQUI-LUCKS, P.; PRADO, M. C. de A. 2020. Corpora de inglês aeronáutico: desafios para o estudo da área e proposta de trabalho conjunto. *Tradterm*, v. 37 (1), 2020.

TOSQUI-LUCKS, P. ; SILVA, A. L. B. C. Aeronautical English: Investigating the nature of this specific language in search of new heights. *The Especialist*, v. 42 (1), 2020a.

TOSQUI-LUCKS, P.; SILVA, A.L.B.C. Da elaboração de um glossário colaborativo à discussão sobre os termos "inglês para aviação" e "inglês aeronáutico". Estudos Linguísticos, São Paulo, 49 (1), pp. 97-116, 2020b. Disponível em: <https://revistas.gel.org.br/estudos-linguisticos/article/view/2561>. Acesso em: 20 fev. 2022.

UNITED STATES OF AMERICA. Federal Aviation Administration. U.S. Department of Transportation. *Air Traffic Organization Policy.* ORDER JO 7110.65W: Air Traffic Control. Washington, D.C., 2015.

UNITED STATES OF AMERICA. U.S. Department of Transportation. *Blue Lightning Initiative*. Disponível em: <https://www.transportation.gov/administrations/office-policy/blue-lightning-initiative>. Acesso em: 25 fev. 2022.

WORLD METEOROLOGICAL ORGANIZATION. *Guide to Practices for Meteorological Offices serving Aviation*. Geneva. (WMO, n.732), 2003.

Transdisciplinaridade na Tecnologia da Linguagem

Transdisciplinarity in Language Technologies

Ana Claudia Zandavalle[1]
Livy Real[2]

1 Americanas S.A.
2 Americanas S.A.

Abstract*:* The need for greater transdisciplinarity in the development of technologies involving natural language is currently increasing. In this chapter, we share cases in which transdisciplinarity added value to language technologies in the electronic retail industry, highlighting how one can learn from experiences in different areas of knowledge, including group discussions and the benefit of having a holistic view of a given project. We also highlight the main challenges in relation to daily practices related to this approach, such as the importance of knowing the profile of the work team and how to obtain better results from it. We also discuss new ways to bring groups from different areas of knowledge closer together.

Keywords*:* transdisciplinarity; computational linguistics; language technologies.

***Resumo**:* A necessidade de maior transdisciplinaridade no desenvolvimento das tecnologias que envolvam a linguagem natural é cada vez mais evidente. Neste trabalho, compartilhamos casos nos quais a transdisciplinariedade agregou valor às tecnologias da linguagem na indústria do varejo eletrônico, destacando como podemos aprender com experiências de diferentes áreas do conhecimento, discussão em grupo e visão holística de um projeto. Destacamos os principais desafios em relação à prática diária dessa abordagem, como a importância de conhecer o perfil da equipe de trabalho e como obter melhores resultados a partir dela, bem como pensar em novas formas de aproximar grupos de diferentes áreas do conhecimento.

***Palavras-chave**:* transdisciplinaridade; linguística computacional; tecnologias da linguagem.

1. Introdução

Vivemos tempos cada vez mais complexos, nos quais grandes projetos necessitam de diferentes saberes e olhares para serem executados e satisfatoriamente concluídos. Seja na educação, no meio científico ou no dia-a-dia do mercado de trabalho, a colaboração constante entre indivíduos de áreas do conhecimento que habitualmente não se comunicam tem sido uma prática crescente (ROSS & MITCHELL, 2018).

É neste contexto que emerge o termo "transdisciplinariedade". A transdisciplinaridade veio à tona a partir de uma crítica em relação à organização padrão do conhecimento em disciplinas do currículo de ensino superior, promovendo reflexões sobre questões éticas e morais. Atualmente, é caracterizada pelo engajamento entre *stakeholders* de ciências diversas com o objetivo de encontrar a melhor solução para um grande problema, sendo socialmente responsável (BERNSTEIN, 2015).

Na literatura, são encontrados três conceitos que são utilizados com uma certa ambiguidade: multidisciplinaridade, interdisciplinaridade e transdisciplinaridade. A multidisciplinaridade reúne diferentes áreas do conhecimento, mas ainda limitada às disciplinas. A interdisciplinaridade estuda, resume e combina as conexões entre as disciplinas em um todo estruturado e racional. E a transdisciplinaridade integra áreas do conhecimento como ciências sociais, exatas e linguística, em um contexto de humanidades e transcende suas fronteiras tradicionais (CHOI; PAK, 2006).

A partir do momento em que o ser humano tem cada vez mais contato com 'máquinas', torna-se mais necessária ainda a convergência de disciplinas para o desenvolvimento de tecnologias que envolvem a linguagem. Entende-se por 'linguagem natural' a linguagem utilizada para a comunicação entre seres humanos (PUSTEJOVSKY; STUBBS, 2013) e o campo de 'processamento da linguagem natural' é responsável por 'projetar e construir aplicativos que facilitem a interação humana com máquinas e outros dispositivos através do uso de linguagem natural' (BIRD; KLEIN; LOPER, 2009) como, por exemplo, *chatbots*, assistentes virtuais, tradução automática, entre outros.

Para além da linguística e da ciência da computação, áreas como psicologia, sociologia, biblioteconomia e outras trazem qualidade, diversidade e impacto social às tecnologias geradas. Neste artigo, mostramos como a prática da transdisciplinaridade contribui diretamente para a geração das tecnologias que fazem uso da linguagem, quais os desafios enfrentados e pontos de melhorias que podem ser aplicados para obter experiências mais eficazes com essa abordagem.

2. Transdisciplinaridade em cenários reais

A Americanas S.A. é um grande *marketplace*, que é uma plataforma de varejo online cuja finalidade é conectar lojas parceiras que querem vender seus produtos com consumidores que desejam encontrar determinado produto para efetuar a compra; assim, o seu negócio se baseia na venda de produtos. Para facilitar o cliente a encontrar o que deseja comprar, precisamos conectar a linguagem utilizada pelos clientes com a forma como essa informação é fornecida pela loja parceira.

Nesse contexto, temos um grande projeto de classificação da informação do produto, que consiste em atribuir um significado, um conceito ao produto, para que ele possa ser encontrado e vendido. Por exemplo: quando pensamos no universo de cadeiras, temos uma variedade de cadeiras à venda e, a priori, poderíamos pensar que todas elas consistem no produto ‘cadeira’. No entanto, ao visualizar diferentes imagens de cadeira, tais como ’cadeira gamer’, ‘cadeira design’, ‘cadeira de praia’, ‘cadeira de jardim’ etc., levantam-se alguns questionamentos: será que toda essa variedade de cadeiras está bem representada informacionalmente somente como ‘cadeira’? Ou seria mais inteligente que cada tipo de cadeira tivesse uma nomenclatura mais personalizada à necessidade do usuário?

Considerando que os vendedores atribuem nomes diferentes a um produto, e que os usuários têm um conhecimento de mundo particular para pesquisar o que eles desejam comprar, nosso objetivo é dar um único significado, um único conceito, e assim, os usuários conseguem encontrar o que querem de forma rápida e, facilmente, realizar a compra.

A conceitualização de produto permite a conexão entre muitas áreas da companhia como o motor de busca, que se refere à recuperação da informação; a classificação mercadológica, a qual representa a organização da informação na indústria oriunda da estruturação de itens em um ambiente físico (ponto de venda); e as especificações técnicas do produto, também chamadas de 'ficha técnica'.

Nesse cenário, a participação dos *stakeholders* faz toda a diferença porque traz visões complementares ao conhecimento técnico do time de analistas, partindo da necessidade de informação do usuário. Assim, mesmo com um propósito diferente entre os times, o grupo foi capaz de alcançar o objetivo de negócio: facilitar a busca pelo produto e agilizar o processo de compra.

Outro projeto em que trabalhamos está relacionado ao engajamento com a empresa, cujo objetivo foi transformar os *feedbacks* dos usuários em informação acionável à área de negócios.

Em grandes empresas, tem sido cada vez mais importante entender automaticamente o *feedback* de seus usuários, pois a partir dessa compreensão pode-se priorizar ou não determinada ação.

Dessa forma, precisávamos transformar milhares e milhares de *feedbacks* em dados confiáveis. Como uma tarefa de classificação, focamos tanto nas temáticas mais frequentes comentadas pelos usuários, como também, nas temáticas que são relevantes para o negócio. Assim, nossa tarefa de classificação multirrótulo não foi tendenciosa considerando apenas os dados analisados, nem tendenciosa para o que a equipe de negócios estava interessada. O resultado quantitativo e qualitativo dessa tarefa foi apresentado no formato de *dashboard*, no qual os analistas de negócios conseguiam facilmente entender as demandas de um grande grupo.

Em relação à transdisciplinaridade nesse projeto, dois pontos merecem destaque: 1) seleção da amostra de dados para a anotação e 2) combinação do conhecimento de analistas de dados, cientistas de dados e time de negócios.

Entre as etapas do fluxo de anotação, uma delas é selecionar uma amostra de dados para construir o corpus anotado. Nessa fase do projeto, inicialmente, nossa equipe tomou decisões de forma isolada e descobrimos que estavam erradas, pois olhamos apenas para os dados, e não para as necessidades da área de negócios. A partir disso, notamos a importância de unir os esforços de analistas, desenvolvedores e área de negócio a fim de ter um completo entendimento da qualidade dos dados, do impacto na anotação, do impacto no balanceamento do conjunto de dados, e consequentemente, na entrega final. Assim, olhamos para a frequência, para os fenômenos linguísticos e para as necessidades de informação do negócio, o que resultou em dados de treinamento mais representativos do desafio de negócio. Esse aprendizado nos permitiu concluir que a seleção da amostra de dados inclui decisões que precisam ser tomadas em conjunto, pois uma etapa influencia o resultado da outra.

O segundo ponto em que a transdisciplinaridade fez a diferença e que também está relacionada à união do conhecimento de diferentes áreas, foi na construção e aplicação de uma taxonomia para a classificação de subtópicos, utilizando-se expressões regulares. Conversando com as unidades de negócio, percebemos que muitos padrões importantes eram melhor recuperados por meio de regras do que por sistemas de aprendizagem de máquina.

Ao codificar as regras no pré-processamento dos dados, notamos que a ordenação das etapas de processamento como normalização, lematização e reconhecimento de *multiwords* poderia distorcer o resultado final e levar a uma decisão errônea. Assim, a combinação de pessoas com um *background* variado trouxe outras dimensões do problema e possibilitou um olhar mais atento a todas as etapas. Na prática, vimos que não podemos confiar no pré-processamento padrão, mas que todas as etapas de um projeto dependem da tarefa que temos nas mãos.

De um modo geral, a equipe deve estar ciente de que cada etapa de um projeto mudará um resultado-chave da empresa, independentemente de quem a executa, pois uma pequena coisa, como uma anotação de dados errada ou um pré-processamento ingênuo, pode trazer uma visão míope da realidade. A transdisciplinaridade chama a atenção para o pensamento crítico em todas as fases da resolução de um problema.

3. Os desafios da aplicação da transdisciplinaridade

Trabalhar com pessoas de diferentes *backgrounds* no mesmo ambiente é enriquecedor e, ao mesmo tempo, desafiador. Um dos principais desafios que encontramos é entender o perfil profissional de cada um, o que envolve tanto as habilidades profissionais quanto o comportamento pessoal.

Considerando desafios de gerenciamento, isto engloba tanto compreender como tirar o melhor do profissional quanto como formar o melhor grupo para resolver determinado tipo de problema. Já por parte do time como um todo, requer alto poder de escuta para que, através da existência de diferentes opiniões baseadas em diferentes perspectivas, possa-se chegar a um consenso para convergir visões em prol do resultado almejado. No geral, trata-se de um longo caminho a percorrer, de aprendizado contínuo, a partir de testes e adaptação para ambas as partes sempre que necessário.

Ainda sobre o perfil do profissional, o aprendizado colaborativo é também um desafio, pois é importante incentivar uma equipe transdisciplinar a desenvolver novas habilidades, mas respeitando o interesse, afinidade com o assunto, a curva de aprendizado e as limitações de cada indivíduo. Em um ambiente mercadológico, o aprendizado colaborativo ainda deve estar em equilíbrio com o objetivo do time na estratégia de negócio da Companhia.

Outro desafio encontrado é a comunicação. Identificamos a necessidade de ter uma estratégia de comunicação com o objetivo de garantir o entendimento do problema entre os times, ter clareza do impacto de cada decisão no resultado final e ter certeza de que estamos aplicando o que foi discutido para desenvolver a melhor solução e não apenas trocando informações.

Além disso, a utilização de termos mais técnicos é uma dificuldade presente quando trabalhamos com equipes de áreas diversificadas, pois quando não estamos acostumados com o vocabulário técnico de uma área diferente e vice-versa, isso dificulta o entendimento e a discussão entre os *stakeholders*. Assim, ter um léxico comum é essencial para alcançar bons resultados com uma equipe transdisciplinar, pois facilita a comunicação e possibilita a compreensão de todos.

4. Conclusão

Neste trabalho, apresentamos a transdisciplinaridade aplicada a projetos de tecnologia da linguagem, com o objetivo de compartilhar casos reais e estimular essa cultura em diferentes grupos de trabalho, tenham eles objetivos científicos ou mercadológicos. Os resultados das experiências relatadas mostraram o valor de unir esforços de profissionais de diferentes áreas do conhecimento para a resolução de determinado problema, mas também os desafios existentes na prática dessa abordagem. Por fim, ressaltamos práticas simples e efetivas que se mostraram eficientes para enriquecer ainda mais as experiências de times transdisciplinares.

Referências

Bernstein, J. H. Transdisciplinarity: A review of its origins, development, and current issues. *Journal of Research Practice*, 11(1), Article R1, 2015. Disponível em: < http://jrp.icaap.org/index.php/jrp/article/view/510/412 >

Choi B.C.K., Pak, A.W.P. Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services, education and policy: 1. Definitions, objectives, and evidence of effectiveness. *Clin Invest Med*. 2006 Dec;29(6):351-364. PMID: 17330451.

Pustejovsky, J., Stubbs, A. *Natural Language Annotation for Machine Learning*. O'Reilly Media, 2013.

Ross, K., Mitchell, C. Transforming transdisciplinarity: An expansion of strong transdisciplinarity and its centrality in enabling effective collaboration. in Fam, D et. al. *Transdisciplinary Theory, Practice and Education: The Art of Collaborative Research and Collective Learning*, Springer 2018, 39-56.

Steven, B., Klein, E., Loper E. *Natural Language Processing with Python*. O'Reilly Media, 2009.

Biodatas

Adriana Mendes Porcellato holds a Ph.D. in linguistics from the University of São Paulo in Brazil and the Sapienza University of Rome in Italy and is currently working as a temporary teacher at the University of São Paulo. She has worked as an Italian and English teacher in different contexts in Brazil for 20 years. Her research interests include second language acquisition, interlanguage and cross-cultural pragmatics, identity and culture in language pedagogy, and materials design.

E-mail: adriana.porcellato@outlook.com

Ana Zandavalle holds a degree in Librarianship Studies from the Federal University of Santa Catarina. She currently works as an Artificial Intelligence Analyst at Americanas S.A. specializing in Natural Language Processing (NLP). She was previously involved in the development of digital products, applying Computational Linguistics and NLP among self-sufficient squads using agile methodologies. She has gained experience in data intelligence in various environments such as startups, consulting companies and advertising agencies.

E-mail: ana.zandavalle@gmail.com

Andressa Costa é mestre em Língua e Literatura Alemã pela UPS e doutora em Linguística Alemã pela Universidade de Mannheim, Alemanha. Atualmente desenvolve um projeto sobre Variação de Registro no Alemão Contemporâneo na PUC-SP. Paralelamente, desenvolve um projeto sobre os discursos de feministas em uma perspectiva multicultural. Seus interesses são: Linguística de Corpus, Variação de Registro e Análise do discurso com base em corpus.

E-mail: acosta.andressa@gmail.com

Anna Catharina de Mendonça Paes é aluna de mestrado em Tradução no programa de pós-graduação em Letras Estrangeiras e Tradução da Universidade de São Paulo (USP), Especialista em Docência do Ensino Superior pela Universidade Paulista (UNIP) e Bacharela em Secretariado Executivo pela Universidade Federal de Santa Catarina (UFSC).

E-mail: annacmp2009@gmail.com

Ariel Novodvorski é Professor associado do Instituto de Letras e Linguística da Universidade Federal de Uberlândia (ILEEL/UFU). Doutor em Estudos Linguísticos pela UFMG e Pós-doutorado pela UFRGS. Atua no curso de Graduação em Letras Espanhol e no Programa de Pós-Graduação em Estudos Linguísticos (PPGEL). Seus interesses de pesquisa são Fraseologia, Terminologia, Linguística de Corpus, Estudos da Tradução, Estudos Descritivos/Contrastivos e Linguística Sistêmico-Funcional. Conta com experiência de mais de vinte anos na docência, pesquisa e tradução, além de publicações em diversos periódicos indexados e em livros. Atualmente é Diretor do ILEEL/UFU (2017-2025).

E-mail: arivorski@ufu.br
Link para acesso ao CV Lattes:
http://lattes.cnpq.br/2882362453894798

Carlos Henrique Kauffmann é doutor e mestre em Linguística Aplicada e Estudos da Linguagem pela PUC-SP. Atualmente é bolsista de pós-doutorado na PUC-SP pela Capes. Foi pesquisador visitante na Northern Arizona University (EUA). É graduado em Linguística pela FFLCH-USP e possui especialização em marketing (ESPM). É pesquisador do GELC (PUC-SP) e parecerista das revistas Texto Livre e Intercâmbio. Suas áreas de interesse abrangem multimodalidade, dimensões discursivas e variação de registros literários e jornalísticos.

E-mail: chkauffmann@corpuslg.org

Elaine Alves Trindade é doutora em Letras-Estudos da Tradução pela FFLCH-USP, mestre em Letras-Semiótica/Linguística pela FFLCH-USP, pós-graduada em Didática para o Ensino Superior pela UNINOVE, Graduada em Letras-Tradutor/Intérprete pelo Centro Universitário Ibero-Americano. Atualmente é professora no curso de Letras-Tradução da PUC-SP (ministrando disciplinas relacionadas à Tradução) e Vice coordenadora desse curso. É também Vice-presidente da Associação dos Professores da PUC-SP. Além do currículo acadêmico, possui mais de 30 anos de experiência em tradução, tendo traduzido 18 livros, inúmeros textos técnicos e mais de 2500 filmes, séries e documentários para legendagem e dublagem.

E-mail: elainetrindade@gmail.com

Giovana de Castro Marchese Rampini é Doutoranda em Letras pelo PPGELLI/USP e mestra em Letras pelo TRADUSP/USP. Foi professora contratada no DLM/USP, no programa de Língua Inglesa, onde ministrou aulas na graduação em Letras. Atua como professora no curso de inglês para seniores da Universidade de Sorocaba (UNISO) e como tradutora parceira do projeto Tradução em Contexto, da Editora Lexikos. É pesquisadora do Projeto COMET (Corpus Multilíngue para Ensino) da FFLCH/USP. Ministra cursos e participa de projetos na área de Letras com ênfase em ensino de inglês, escrita acadêmica, linguística de corpus, lexicografia, tradução e análise do discurso.

E-mail: giovana.marchese01@gmail.com

Guilherme Nunes atua há 5 anos como Cientista e Analista de Dados no Localiza Labs. Possui especialidade na área de Machine Learning, e trabalha com foco em soluções de NLP (processamento de linguagem natural) e previsão de demanda para auxiliar na tomada de decisões estratégicas relacionadas a Revenue Management e People Analytics.

E-mail: gdmello.nunes@gmail.com
Link para acesso ao CV Lattes:
http://lattes.cnpq.br/4758773937085904

Heliana Mello é Professora Titular de Linguística na FALE/UFMG, onde atua desde 1998. Sua principal área de pesquisa é a Linguística de Corpus, com foco específico em compilação e metodologias para análises quantitativas de corpora orais e multimodais. Dentre seus projetos, destaca-se o C-ORAL-BRASIL (www.c-oral-brasil.org) e seus corpora.

E-mail: hmello@ufmg.br
Link para acesso ao CV Lattes:
http://lattes.cnpq.br/5724573734505786

Jamilly Brandão Alvino é Mestra em Estudos da Tradução, subárea Linguística de Corpus, pela Universidade de São Paulo. É bacharela e licenciada em português e chinês pela mesma instituição. Membro do Projeto CoMET – Corpus Multilíngue para Ensino e Tradução (FFLCH-USP) e, também, do Grupo de Estudos de Adaptação e Tradução (GREAT|FFLCH-USP). É tradutora e pesquisadora de literatura infantojuvenil, tradução e adaptação. Seus idiomas de trabalho são chinês, inglês e português.

E-mail: brandao.jamilly@gmail.com

Livy Real earned her Ph.D. in Linguistics from the Federal University of Paraná in 2014 and has been actively involved in Computational Linguistics since 2012. At present she holds the position of Natural Language Processing Manager at Americanas S.A. Her interests revolve around utilizing open tools and corpora, with a particular emphasis on Portuguese Processing, Natural Language Understanding tasks, and underrepresented communities.

E-mail: livyreal@gmail.com

Malila Prado is an assistant professor at BNU-HKBU United International College, a Sino-foreign university based in China. She has worked as an English language teacher for over 20 years. She holds a Master´s Degree and PhD from the Department of Modern Languages of University of Sao Paulo (Brazil), examining the language used by pilots and air traffic controllers in abnormal situations through corpus linguistics. Her current research interests lie in corpus linguistics, pragmatics, ESP, and teacher education.

E-mail: malilaprado@uic.edu.cn

Marina Leivas Waquil, doutora e mestre em Teorias Linguísticas do Léxico pelo Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul (UFRGS), tem pós-doutorado pelo Programa de Letras Estrangeiras e Tradução (LETRA) da Universidade de São Paulo (USP). É uma das coordenadoras do Grupo de Estudos e Ação em Feminismos, Gênero e Tradução da USP e atua principalmente nas áreas de Tradução, Tradução Feminista, Terminologia, Linguística de Corpus, Fraseologia e Língua Espanhola. Desde 2010 é Tradutora Pública e Intérprete Comercial, concursada pela Junta Comercial do Rio Grande do Sul.

E-mail: marinawaquil@gmail.com

Patrícia Tosqui Lucks é Doutora em Linguística pela Unesp e tem Pós-doutorado pela FFLCH-USP, na área de Linguística de Corpus. Desde 2009, coordena o Setor de Capacitação em Inglês Aeronáutico do Instituto de Controle do Espaço Aéreo (ICEA). Colabora nos programas de pós-graduação do ITA e da UNITAU. É líder do Grupo de Estudos em Inglês Aeronáutico (GEIA), e membro do ICAEA Research Group, um grupo internacional que realiza pesquisas sobre Aviation English. Tem publicado artigos e livros sobre o assunto, e organizou edições temáticas dos periódicos The ESPecialist e TradTerm em 2020. Atua com ensino de inglês para fins específicos, avaliação de proficiência, linguística de corpus aplicada ao desenvolvimento de materiais didáticos, formação de professores e tecnologia aplicada à educação.

E-mail: patricialucks@gmail.com

Rafaela Rigaud Peixoto é Tradutora e Pesquisadora do Departamento de Controle do Espaço Aéreo, Professora da Universidade da Força Aérea e Pós-doutoranda em Linguística de Corpus pela Universidade de São Paulo. Pesquisadora do Núcleo de Estudos Interdisciplinares em Ciências Aeroespaciais (NEICA/UNIFA) e do Grupo de Estudo em Inglês Aeronáutico (GEIA/ICEA). Seus principais interesses de pesquisa abrangem Terminologia, Tradução, Linguística de Corpus, Identidade, Narrativas e Relações Interculturais. Atualmente desenvolve pesquisa sobre terminologia na área de Aviação e de Defesa, com interface discursiva.

E-mail: rafaela.peixoto@gmail.com

Dra. Regiani Aparecida Santos Zacarias é professora Livre-Docente em Língua Inglesa junto ao Departamento de Letras Modernas da FCL-UNESP/Assis. É professora credenciada junto ao PPG em Mestrado Profissional Docência para a Educação Básica do Departamento de Educação da FC-UNESP/Bauru e ao PPGLLP da FCLAr-UNESP/Araraquara. Desenvolve pesquisas no campo da Lexicografia Pedagógica Bilíngue com base na Linguística de Corpus como abordagem investigativa e assertiva com apoio da FAPESP, CAPES, CNPq e The National Endowment for the Humanities.

E-mail: regiani.zacarias@unesp.br

Stella E. O. Tagnin é professora sênior do Departamento de Línguas Modernas da Universidade de São Paulo, Brasil. Embora aposentada, continua ativa na pós-graduação. Lecionou disciplinas de Tradução no Curso de Especialização em Tradução por mais de 25 anos. Introduziu a Linguística de Corpus na USP em 1998. Coordena o Projeto COMET - Corpus Multilíngue de Ensino e Tradução. É autora de O Jeito que a Gente Diz (2013); organizou, em co-autoria com Vander Viana: Corpora no Ensino de Línguas Estrangeiras e Corpora na Tradução e com Cleci Bevilacqua Corpora na Terminologia. Atualmente trabalha num dicionário bilingue inglês-português de colocações verbais.

E-mail: seotagni@usp.br

Xiao Wang graduated from Fujian University of Technology with a degree in translation and interpretation. She has been providing language assistance to the community since her freshman year. She is an experienced English-Chinese translator and a native Chinese speaker. She is currently working as an English teacher in the K-12 age group for an E-learning company listed in China.

E-mail: wangxiao0728@hotmail.com

Sobre o livro
Tipologia: Source Sans Pro
Formato fechado: 210 x 148 mm (A5 retrato)
Visualização em página dupla (paisagem)
Impressão em formato aberto: 297 x 210 mm (A4 paisagem)
Finalizado em outubro de 2023


Corpus
CAPES
USP
unesp