



Pragmática de Corpus: o que é e onde estamos

Corpus Pragmatics: what it is and where we are now

Giovani Santos

Mary Immaculate College, University of Limerick, Limerick / Irlanda

giovani.santos@mic.ul.ie

<http://orcid.org/0000-0003-4116-5613>

Mateus Miranda

Mary Immaculate College, University of Limerick, Limerick / Irlanda

mateus.desouza@mic.ul.ie

<http://orcid.org/0000-0003-2575-8769>

Resumo: Este trabalho objetiva apresentar um novo campo que emergiu a partir da intersecção entre a Linguística de Corpus e a Pragmática: a Pragmática de Corpus. Para tanto, através de uma revisão da literatura como ponto de partida, traçamos um panorama que abarca a origem, os aspectos teórico-metodológicos, e os desafios da nova área. Ademais, introduzimos as abordagens forma-função e função-forma, dois modelos investigativos que integram a disciplina. Finalmente, por meio de um estudo de caso, a fim de ilustrar um dos possíveis percursos de análise, investigamos o marcador pragmático *kind of* por meio da filtragem, método que compõe a abordagem forma-função, no discurso oral de brasileiros universitários. Os subcorpora que subsidiam a pesquisa são o *Spoken Corpus of Brazilian Portuguese and L2-English* (SCoPE²) e o *Brazilian Spoken English Learner Corpus* (BraSEL). Os resultados apontam que quando usado pragmaticamente, mesmo em contextos linguísticos distintos, *kind of* ocorre em seus três domínios funcionais (atitudinal, interpessoal, textual) e como parte constituinte de marcadores de linguagem vaga.

Palavras-chave: pragmática de *corpus*; forma-função; função-forma; *kind of*.

Abstract: This work aims to present a new field which has emerged from the intersection between Corpus Linguistics and Pragmatics: Corpus Pragmatics. To do so, through a literature review as a starting point, we offer an overview that encompasses the origin,

the theoretical and methodological aspects, and the challenges of the new field. In addition, we introduce the form-to-function and function-to-form approaches, two investigative models which integrate the discipline. Finally, by means of a case study in order to illustrate one of the possible analytical routes, we investigate the pragmatic marker *kind of* by employing sifting, a method which comprises the form-to-function approach, in the oral discourse of Brazilian university students. The subcorpora which support the research are the *Spoken Corpus of Brazilian Portuguese and L2-English* (SCoPE²) and the *Brazilian Spoken English Learner Corpus* (BraSEL). The results show that when used pragmatically, even in different linguistic contexts, *kind of* occurs in its three functional domains (attitudinal, interpersonal, textual) and as a constituent part of vague language markers.

Keywords: corpus pragmatics; form-to-function; function to form; kind of.

Submetido em 08 de outubro de 2020

Aceito em 14 de dezembro de 2020

1 Introdução

A linguagem para Firth (1957), segundo Sinclair (2004, p. 103), está atrelada ao contexto que integra fatores como a ação verbal e sofre influência das pessoas, coisas e eventos. Com base em postulados como os de Firth, a Linguística de Corpus (doravante LC), a qual conhecemos hoje, foi desenvolvida por neo-firthianos como Sinclair, que contribuíram para sua expansão, desenvolvendo estudos por meio da observação da linguagem em seu contexto real (McENERY; HARDIE, 2012). Tomando a definição de Sinclair (2005, p. 16), um corpus “é uma coleção de textos em formato eletrônico, selecionados de acordo com critérios externos para representar, o melhor possível, uma língua ou sua variação como fonte de dados para a pesquisa linguística”.¹ Nos anos 1960, o *Brown Corpus*, em formato eletrônico e com um milhão de palavras, foi um marco na área, inspirando, posteriormente, a compilação de outros. Hoje, corpora com milhões de palavras não são incomuns, como o *Corpus of Contemporary American English* (COCA), o Corpus do Português NOW, o *British*

¹ Nossa tradução para: “[A corpus] is a collection of pieces of language text in electronic form, selected according to external criteria to represent, as far as possible, a language or language variety as a source of data for linguistic research.”

National Corpus (BNC), entre outros.² A LC veio a se consolidar nos anos 1980, mas, se em uma perspectiva mundial podemos considerá-la uma disciplina jovem em relação a outras áreas da linguística, as primeiras publicações no Brasil só aconteceriam no final dos anos 1990 e início dos anos 2000 (BERBER SARDINHA, 1999, 2000).

Da mesma forma, Gomes de Matos observava em 1982 que os estudos da Pragmática ainda não estavam inclusos nas ementas dos cursos de pós-graduação no Brasil, e que as pesquisas nesse campo ainda eram incipientes (RAJAGOPALAN, 1999), apesar de o termo pragmática, em seu uso moderno, ser atribuído ao filósofo Charles Morris na década de 1930 (LEVINSON, 1983). Para Rajagopalan (1999, p. 323), a própria indefinição do termo pragmática era um fator dificultador para a impopularidade da área no país e também afetava pesquisadores pelo mundo. Em uma publicação mais recente, Rajagopalan (2016, p. 203) argumenta que os pragmaticistas ainda divergem no que tange às temáticas e prioridades de investigações da área e afirma que há uma enorme carência de obras sobre a Pragmática em língua portuguesa.

Destacamos ainda a interdisciplinaridade da LC que possibilita, por meio de corpora, investigações que contemplam o léxico, a sintaxe, o discurso, o ensino e a produção de materiais para a sala de aula de línguas, os estudos de registros e gêneros, os estudos literários e tradução, a sociolinguística, dentre outros (cf. O'KEEFFE; McCARTHY, 2010). Em oposição à crítica de que o marasmo se instalou no campo dos estudos linguísticos, Rajagopalan (2006, p. 160) faz menção à LC e sua interdisciplinaridade, com foco na lexicografia e na sintaxe, como uma área revolucionária que “questiona um dos postulados da linguística tradicional (teórica), com base em uma perspectiva de ordem eminentemente prática”. Nesse sentido, o autor destaca que a lexicografia, vista como um exercício da semântica intencional, e norteadas pelo racionalismo, não abre espaço para variações dialetais e geográficas de sentidos, ou para novas acepções da língua, o que contrasta com o interesse da LC em registrar os diferentes usos de cada item lexical em contextos variados.

Como vimos, tanto a LC quanto a Pragmática podem ser consideradas disciplinas relativamente novas e, de fato, foram por muito

² Cf. Tagnin (2010) para uma lista de corpora nas línguas alemã, espanhola, francesa, inglesa, italiana e portuguesa.

tempo exclusivas devido às suas diferenças metodológicas (ROMERO-TRILLO, 2008b). No entanto, recentemente, a partir de seu caráter interdisciplinar, a LC se fundiu com a Pragmática em uma nova subárea que vem a ser denominada Pragmática de Corpus e que pode caracterizar-se como um avanço para os estudos da linguagem.

Considerando o panorama da LC e da Pragmática no Brasil, e no espírito desta chamada para discorrer sobre as conquistas e desafios da LC, este trabalho se propõe a apresentar a Pragmática de Corpus (uma conquista mútua da LC e da Pragmática) e seu estado da arte. Para tanto, a estrutura deste artigo está disposta da seguinte forma: após esta introdução, perpassamos as Seções 2 e 3 que abordam, respectivamente, a LC e a Pragmática. O resultado da fusão entre as duas disciplinas, a Pragmática de Corpus, é descrito na Seção 4 a partir de seu percurso teórico e metodológico. Tal seção é ainda subdividida em três partes em que discorreremos sobre as limitações e desenvolvimentos da nova área, além de focarmos em suas duas abordagens de análise: forma-função e função-forma. Na seção 5, complementamos, a fim de exemplificação, o trajeto das seções anteriores com um estudo de caso através de corpora orais. Por fim, tecemos na última seção as considerações finais.

2 Linguística de Corpus

Estudos no campo da LC têm aumentado de maneira consistente e substancial nos últimos 30 anos, e os resultados confirmam que a LC é um meio eficiente para se fazer análises da linguagem em uma vasta gama de contextos linguísticos (cf. BERBER SARDINHA, 2000; McCARTHY; O'KEEFFE, 2010 para uma perspectiva histórica).

Segundo McCarthy e O'Keeffe (2014, p. 271), “[a] evidência estatística e contextual que pode ser obtida através do uso de um software [de LC] nos permite fazer interpretações confiáveis das intenções comunicativas dos falantes e escritores.”³ Esta combinação estatística e contextual permite que pesquisadores analisem seus dados em termos qualitativos e quantitativos, através, por exemplo, do estudo empírico de listas de frequência e de concordância, reduzindo significativamente o risco de se introduzir quaisquer inclinações pré-concebidas.

³ Nossa tradução para: “[the] statistical and contextual evidence the [CL] software can provide us with enables us to make more reliable interpretations of speakers’ and writers’ communicative purposes.”

Validade e confiabilidade são, de fato, características representativas dos resultados obtidos pela LC, uma vez que as ferramentas principais aplicadas numa pesquisa baseada em corpus são computacionais, o que se opõe a possíveis limitações humanas. Em outras palavras, enquanto a análise manual é suscetível ao erro com relação às mais simples ocorrências de palavras e padrões linguísticos, abordagens baseadas em corpus oferecem ao analista a oportunidade de processar, de maneira meticulosa, grandes quantidades de dados. Tais processos não apenas têm o potencial de recuperar cada ocorrência de um item em questão, como também destacam estruturas complexas de padronização linguística e associações de palavras.

Desta forma, imparcialidade, precisão, processabilidade, e o acesso à linguagem natural (autêntica) em grande escala são benefícios essenciais que a LC oferece aos pesquisadores. Por conseguinte, muito mais informação e conhecimento sobre a forma como a linguagem opera pode ser revelado e descoberto atualmente, principalmente quando comparado com uma época anterior à LC. Biber *et al.* (1998, p. 5) observam que, uma vez propriamente utilizado, um corpus bem arquitetado e compilado pode fornecer maior variedade de informações sobre a língua em seu uso real. De fato, questões linguísticas que preliminarmente pareciam quase impossíveis de serem abordadas, ou que ainda não haviam sido sequer consideradas, foram evidenciadas e continuam a ser investigadas com a ajuda das ferramentas da LC e da disponibilidade de muitos corpora gratuitamente acessíveis.

Em um relato sobre o processo analítico e técnico de um grande e influente projeto baseado em corpus, realizado no início dos anos 1980, Sinclair (1991, p. 2) refere-se ao fato de que como língua mais descrita no mundo, eram mínimas as chances de que a língua inglesa apresentasse quaisquer tipos de novas evidências reveladoras que, por séculos, ainda não haviam sido observadas. O projeto foi, inicialmente, desenvolvido para ajudar na construção de um dicionário inovador, nomeado *COBUILD Dictionary* (SINCLAIR *et al.*, 1987). Tal projeto necessitou do desenho e desenvolvimento de um grandioso corpus da língua inglesa, como também de novas técnicas para auxiliar na observação e análise da linguagem em uso. Conhecido pelo mesmo nome do dicionário, o projeto não só lançou uma nova abordagem para a construção de dicionários, como também foi fundamental para a introdução de uma nova perspectiva para a descrição da linguagem.

Uma das mais importantes contribuições que resultaram do projeto COBUILD segue na forma do trabalho de Sinclair, mais notavelmente em Sinclair (1991), no qual o princípio idiomático é proposto e delineado. Tal princípio estabelece que os usuários de uma língua nem sempre selecionam suas palavras de maneira em que espaços sintáticos são preenchidos com escolhas abertas. Ao invés disso, falantes e escritores são mais propensos a formar o seu discurso e textos usando porções de linguagem semiconstruídas que estão disponíveis no sistema linguístico e que variam em diferentes níveis de flexibilidade. Além da questão de a linguagem autêntica mostrar-se através de construções pré-fabricadas, ou padrões linguísticos, outra valiosa inferência resultante do princípio idiomático, e importante para a observação e descrição da língua, é o conceito de unidades de significado. Nas palavras de Sinclair (2003, p. 3), “[s]e estudarmos casos de uso real, descobrimos que palavras e frases adjacentes ajudam muito na determinação do significado.”⁴ Ou seja, o significado de muitas palavras, se não da maioria, é determinado pela sua interação e correlação com outras palavras (cotexto), sugerindo que as palavras se selecionam e que tal performance possui uma relação clara com o significado.

O projeto COBUILD documenta um exemplo prático e significativo do quão favorável a LC pode ser ao estudo da linguagem natural empregada em contextos reais. Desde então, os métodos e as técnicas da LC têm sido aprimorados, desenvolvidos, e de grande influência, além de aplicados a muitas outras subáreas da linguística, incluindo os campos da linguagem falada e da Pragmática.

É evidente que os estudos do discurso oral têm se beneficiado grandemente com os avanços da LC nos últimos anos (CAINES *et al.*, 2016; LOVE, 2020). Nesse sentido, podemos afirmar que outra contribuição considerável da LC foi o seu auxílio na investigação e exposição de características importantes da língua oral que, do contrário, não teriam sido facilmente acessadas e descobertas – especialmente com o uso de dados fabricados. Embora ainda seja uma tarefa desafiadora, devido a questões de disponibilidade de participantes e do trabalho transcritório, a construção de corpora orais tem sido facilitada consideravelmente pelos avanços tecnológicos. Consequentemente, pesquisadores hoje têm

⁴ Nossa tradução para: “[i]f we study instances of usage, we find that the surrounding words and phrases help a lot in determining the meaning.”

maior facilidade de acesso a dados linguísticos autênticos do discurso oral e, portanto, conseguem sistematicamente extrair dados confiáveis para a investigação da singularidade das interações desta modalidade. Como Sinclair (1991, p. 4) adequadamente aponta, “a capacidade de se examinar grandes corpora de maneira sistemática nos permite acesso a evidência de uma qualidade antes não disponível”.⁵ Tal afirmação é particularmente verdadeira com respeito aos corpora de fala, uma vez que eles oferecem ao pesquisador características distintivas em linguagem natural que representam interações reais entre interlocutores.

Nos últimos anos, tem havido uma crescente conscientização em relação à importância do corpus de fala para os estudos linguísticos, uma vez que este tem o potencial de revelar e destacar características linguísticas que são peculiares deste contexto comunicativo. Ademais, apesar de tanto a modalidade escrita quanto a oral proporcionarem ricos contextos linguísticos para análises de fenômenos pragmáticos, é na interação oral que se encontra um terreno ainda mais fértil para o surgimento destes fenômenos, visto que esta requer que os interlocutores negociem a construção de conteúdo e sentido em tempo real. Nesse contexto, é plausível argumentar que a LC é um complemento vantajoso tanto para o estudo da língua oral quanto da pragmática. Certamente a LC tem sido acolhida como um complemento valioso para muitas áreas dos estudos linguísticos, e uma área que parece ter encontrado um aliado perfeito na LC é exatamente a do campo da Pragmática, que, juntas, provocaram o surgimento de um campo de estudo relativamente recente: a Pragmática de Corpus (doravante PC). Antes de introduzirmos tal fusão, contudo, faz-se necessária a apresentação de um breve panorama sobre a Pragmática.

3 Pragmática

Como mencionado na Seção 1, definir pragmática não é tarefa fácil. Em seu nível mais fundamental, podemos definir a pragmática como o estudo da língua em uso real, e que considera as relações entre contexto de uso e sentido intencionado. O termo pragmática, como conhecemos hoje, é atribuído a Morris (1938), a partir da obra *Fundamentos de uma*

⁵ Nossa tradução para: “the ability to examine large text corpora in a systematic manner allows access to a quality of evidence that has not been available before.”

teoria dos signos que dividiu a linguagem em três planos: o sintático, o semântico e o pragmático. Por um lado, Morris define a área da linguística que lida com o contexto e aponta seu vínculo com outras áreas (e.g. sintaxe e semântica), mas, por outro, apresenta uma definição limitada de contexto, ignorando fatores importantes como as relações sociais e situacionais (CULPEPER; HAUGH, 2014). Para Culpeper e Haugh (2014, p. 6), apesar de indicarem que uma divisão anglo-americana e continental da pragmática não deve ser enfatizada, Morris propõe uma visão micro do contexto, pois parte de uma visão anglo-americana, que também é o fundamento para outros trabalhos, como as implicaturas conversacionais de Grice (1975) e a Teoria da Relevância de Sperber e Wilson (1995).

A outra visão da pragmática, denominada continental, relaciona-se para além da linguística, com outras áreas cognitivas, sociais e culturais, através da linguagem em uso e do comportamento humano em contextos sociais (CULPEPER; HAUGH, 2014). Dessa forma, tendo em vista o exposto acima e o escopo deste trabalho, não se pretende aqui explorar a definição de pragmática de forma exaustiva. No entanto, a definição de Yule (1996) nos parece ir de encontro ao propósito da PC.⁶ Segundo o autor:

A pragmática é o estudo da relação entre as formas linguísticas e os usuários dessas formas. [...] A vantagem de estudar a linguagem por meio da pragmática é que se pode falar sobre as intenções das pessoas, suas suposições, seus propósitos ou objetivos, e os tipos de ações (por exemplo, pedidos) que realizam quando falam⁷. (YULE, 1996, p. 4).

Ademais, Yule (1996, p. 4) nota que uma das vantagens de considerar a pragmática como objeto de estudo é a possibilidade de investigar o que ele denomina *conceitos humanos*, que incluem os

⁶ Destacamos que o silêncio, a prosódia e a linguagem corporal também são fenômenos pragmáticos que auxiliam na interpretação do significado e, especialmente em corpora multimodais, são também relevantes para os propósitos da PC.

⁷ Nossa tradução para: “Pragmatics is the study of the relationship between linguistic forms and the users of those forms. [...] The advantage of studying language via pragmatics is that one can talk about people’s intended meanings, their assumptions, their purposes or goals, and the kinds of actions (for example, request) that they are performing when they speak”.

significados, as suposições, os propósitos, os objetivos e as ações dos indivíduos, mesmo apontando que as análises de tais conceitos são extremamente difíceis de serem conduzidas. Todavia, ao discorrer sobre a relação entre a LC e a Pragmática, Raso (2016) reforça que por meio da contribuição empírica da LC, todos os temas da Pragmática são passíveis de investigação. Dessa forma, Rühlemann e Clancy (2018) confirmam que se a pragmática geralmente lida com pequenas quantidades de textos em determinados contextos, através da LC pode-se aplicar tais análises a um volume maior de dados. Nesse sentido, a seção que segue tem como objetivo apresentar a origem da PC através dos princípios que a norteiam, destacando suas limitações, desenvolvimentos e as duas abordagens que a compõem: forma-função e função-forma.

4 Pragmática de Corpus

A PC pode ser definida como o campo linguístico de investigação da linguagem autêntica e em uso real com o auxílio de corpora, com vista à interpretação contextual da linguagem escrita ou falada. Tal campo linguístico dá-se pela intersecção entre os campos da LC e da Pragmática (RÜHLEMANN; AIJMER, 2015) e, embora o termo em si seja de cunhagem recente, esta é uma junção que tem evoluído consideravelmente na última década (cf. ROMERO-TRILLO, 2008b para uma introdução à trajetória científica que uniu as duas áreas de conhecimento).

Estudos de fenômenos pragmáticos baseados em corpora têm sido realizados desde os anos 90 (JUCKER; TAAVITSAINEN, 2014), com o interesse em tal abordagem aumentando gradativamente através dos anos e afirmado, posteriormente, com a publicação de um volume dedicado à LC no *'Journal of Pragmatics'* em 2004, com a edição da Conferência da IPrA (*International Pragmatics Association*) em 2007, também com foco na LC, e com a edição da Conferência ICAME (*International Computer Archive of Modern and Medieval English*) com foco em pragmática e discurso em 2008. Contudo, foi com a influente publicação de Romero-Trillo (2008a) que a atenção se voltou para o fato de que há um relacionamento de interesses mútuos entre a LC e a Pragmática, o que poderia ser proveitosamente explorado quando estas duas disciplinas eram fundidas. *Pragmatics and Corpus Linguistics: a mutualistic entente* (ROMERO-TRILLO, 2008a) é o primeiro livro

que agrega pesquisadores que combinam as metodologias de ambas as disciplinas para a investigação de diversas questões linguísticas. Desde então, a produção acadêmica na área da PC tem aumentado exponencialmente, e firmado a disciplina como uma abordagem eficiente para a investigação e compreensão do uso de diversos recursos linguísticos em contextos reais. Tal afirmação é evidenciada pela abundante literatura disponível, na qual inclui uma série anual de livros lançada em 2013 (*Yearbook of Corpus Linguistics and Pragmatics*), livros de contribuições organizadas (AIJMER; RÜHLEMANN, 2015; JUCKER *et al.*, 2009; TAAVITSAINEN *et al.*, 2014), um guia de estudo e pesquisa (RÜHLEMANN, 2019), e um periódico recentemente lançado e exclusivamente dedicado à PC (*Corpus Pragmatics*).

No que concerne aos tipos de estudos que são abordados pela PC, Aijmer e Rühlemann (2015) demonstram que essa disciplina não se limita à investigação de apenas alguns fenômenos pragmáticos como, por exemplo, os marcadores pragmáticos – provavelmente o fenômeno mais estudado na PC – mas também abrange temas centrais da Pragmática, como relevância, dêixis, processabilidade e atos da fala. O livro foi organizado para ser uma referência, assim como uma contribuição para o crescimento da disciplina, e é dividido em seis áreas nucleares da Pragmática: atos da fala; princípios pragmáticos; marcadores pragmáticos; avaliação; referência; e tomada de turno. Por sua vez, Clancy e O’Keeffe (2015) oferecem uma discussão crítica sobre os temas que são mais investigados dentro da PC, sendo eles: atos da fala; marcadores pragmáticos; linguagem, poder e ideologia; a organização de discurso; dêixis. Além de apresentarem uma revisão da literatura e exemplificação de estudos, os autores destacam que a PC é uma área que está “madura e carregada de oportunidades de pesquisa”⁸ (CLANCY; O’KEEFFE, 2015, p. 236)

Embora a LC e a Pragmática sejam hoje reconhecidas como mutuamente favoráveis (CLANCY; O’KEEFFE, 2015; ROMERO-TRILLO, 2008b), Romero-Trillo (2008b) observa que estes campos representaram, por muitos anos, duas linhas de pensamento paralelas, porém mutuamente exclusivas e excludentes. Isto porque a LC, de um lado, trabalha com grandes quantidades de textos e adota uma abordagem mecânica, matemática e mais estrita; enquanto, por outro lado, o campo da

⁸ Nossa tradução para: “[...] the area is ripe with research opportunities.”

Pragmática trabalha com estudos em menor escala (um pequeno número de textos), abordando a língua de uma perspectiva contextual que permite uma visão mais flexível, inferencial e interpretativa da linguagem em uso através de diferentes métodos (cf. JUCKER *et al.*, 2018). Em outras palavras, enquanto a LC é essencialmente quantitativa (embora permita também espaço para uma análise qualitativa) e inicialmente preocupada com número de frequência e medidas estatísticas, a Pragmática é fundamentalmente qualitativa e preocupada com a interpretação do significado no contexto de uso (RÜHLEMANN; AIJMER, 2015; RÜHLEMANN; CLANCY, 2018).

Contudo, pode-se dizer que, apesar de metodologicamente distintas e opostas, e, portanto, aparentemente impossíveis de se relacionar, a LC e a Pragmática compartilham ao menos duas características que permitiram a exploração de um terreno em comum, assim corroborando a fusão de tais disciplinas. A primeira característica diz respeito ao fato de que, historicamente, ambos os campos acolheram outras metodologias e quadros teóricos para substanciar e fortalecer suas próprias abordagens. O’Keeffe *et al.* (2011, p. 20) constatam que a Pragmática, como um quadro teórico de investigação da produção linguística e interação, não é em si uma metodologia, sendo então combinada muitas vezes com uma gama de métodos em pesquisas focadas em fenômenos pragmáticos. Em contraste, Rühlemann e Clancy (2018, p. 242) observam que apesar de os corpora (particularmente os megacorpora) terem sido originalmente compilados com propósitos lexicográficos em mente, seu crescimento e desenvolvimento os levaram à combinação de muitos quadros teóricos linguísticos com uma metodologia de LC. Em concordância com esta visão, para Biber *et al.* (1998, p. 9), uma investigação baseada em corpus não deveria ser considerada como uma abordagem exclusiva e distinta, mas como uma abordagem que complementa outras mais tradicionais. Por sua vez, a segunda característica compartilhada é a de que ambas as disciplinas têm como pressuposto básico o entendimento de que forma e função não estão intrinsecamente correlacionadas. Em outras palavras, o significado e a função de uma forma linguística são recuperados e interpretados através de sua relação com o cotexto (LC) e com o contexto (Pragmática).

O relacionamento complementar entre a LC e a Pragmática é essencialmente simbiótico. Isto porque a Pragmática oferece meios para a interpretação contextual da riqueza de descobertas resultantes

dos métodos da LC. Por outro lado, enquanto interessada em dados reais e no significado criado na interação comunicativa, a Pragmática também se beneficia da LC através da disponibilidade de uma grande quantidade de produção linguística de ocorrência natural. Além do mais, a metodologia da LC tem o potencial de recuperar com sucesso itens pragmáticos específicos e/ou destacar padrões pragmáticos que seriam, de outra forma, difíceis de iluminar sem o auxílio dos instrumentos da LC (O'KEEFFE *et al.*, 2011).

Portanto, por essas razões, podemos argumentar que a LC e a Pragmática têm uma tendência natural a se gravitarem. Como assinalado por Romero-Trillo (2008b, p. 5, *itálico dos autores*), “a linguística de corpus e a pragmática são *duas versões do mesmo fenômeno*: a mecânica – material – (estudos baseados em corpus), e a sua interpretação e explicação (pragmática).”⁹

Como um quadro teórico-metodológico, a PC une os princípios fundamentais e as metodologias das áreas do conhecimento que a deram origem. Segundo Rühlemann e Aijmer (2015, p. 12), a PC tem uma metodologia de leitura integrada (FIGURA 1), no sentido de que integra a metodologia de análise tipicamente qualitativa e horizontal da pragmática com a metodologia predominantemente quantitativa e vertical da LC.

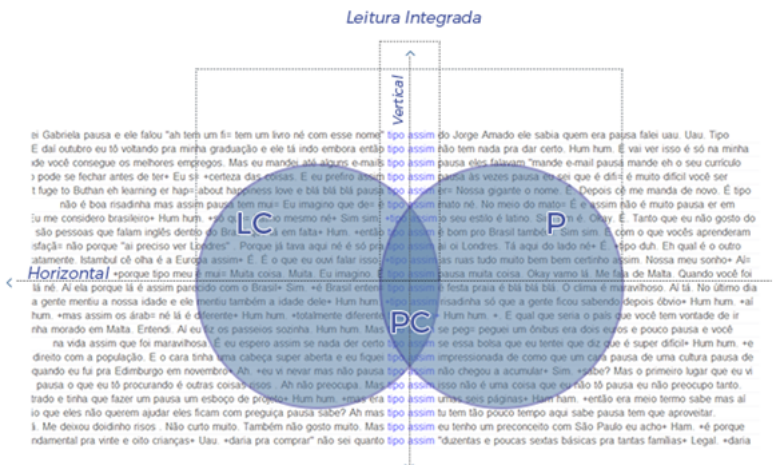
Em relação à metodologia vertical, os autores referem-se ao nóculo (ou item de busca), que toma a posição central e vertical nas listas de concordância, e sua relação colocacional com as palavras à sua direita e esquerda. A leitura vertical do corpus oferece ao analista uma hipótese inicial com base na posição do item investigado e seu ambiente linguístico, como também na preferência lexical (ou padrões de colocação e coligação) do mesmo. Ademais, é durante esse processo de leitura que aquele de posse de um corpus simples (um corpus que não foi etiquetado) filtra os dados para que usos pragmáticos de um item específico sejam separados daqueles não pragmáticos.

Em contrapartida, a metodologia de leitura horizontal permite uma investigação mais aprofundada e minuciosa pela qual o analista leva em consideração contextos maiores do que apenas as palavras e frases imediatamente ao redor do nóculo. Assim, não apenas o cotexto imediato

⁹ Nossa tradução para: “corpus linguistics and pragmatics are two versions of the same phenomenon: the mechanics – the subject-matter – (corpus studies), and its interpretation and explanation (pragmatics).”

é levado em consideração, como também a leitura do contexto expandido, para além da linha visualizada no concordanciador de uma dada busca.

FIGURA 1 – Metodologia integrada de leitura da Pragmática de Corpus



Fonte: Adaptada de Rühlemann e Clancy (2018, p. 245)

A integração das leituras vertical e horizontal na PC cria um processo analítico de dupla direcionalidade que tem um grande potencial para a investigação de material linguístico autêntico. Em síntese, esta sinergia mutualista entre a LC e a Pragmática faz da PC um quadro teórico-metodológico significativo para a análise fidedigna e empírica da linguagem usada e desenvolvida (e.g. desenvolvimento de L2) em contextos reais. Dentro deste quadro teórico-metodológico, o analista pode conduzir análises estatísticas enquanto, concomitantemente, realiza uma interpretação detalhada dos dados com referência a informações contextuais importantes que podem referir-se não apenas ao contexto, como também à situação na qual a comunicação ocorre e ao conhecimento prévio entre interlocutores.

4.1. Limitações, oportunidades e desenvolvimento da Pragmática de Corpus

Não é sem limitações e desafios que a PC, como uma disciplina emergente, posiciona-se como uma subárea da Linguística. Romero-Trillo (2008b), já no princípio, notara que uma das desvantagens em conduzir estudos pragmáticos com uma metodologia de LC é que, muitas

vezes, os corpora eram desprovidos de informações de contextualização, como o status socioeconômico e cultural dos participantes e o tipo de relacionamento e contexto situacional das interações comunicativas. Tal incapacidade, contudo, já está solucionada com a adição de metadados durante o processo de compilação e construção dos corpora. O acesso ao contexto dá-se ainda com maior precisão em casos de corpora menores,¹⁰ em que o analista e o compilador são um e o mesmo, o que permite uma perspectiva interna de análise, sustentando as condições de interpretação de dados. Este relacionamento próximo entre o analista e os dados, como também entre linguagem e contexto, faz dos pequenos corpora a opção perfeita para os estudos pragmáticos. De fato, pesquisadores que têm se debruçado sobre os estudos e discussões que permeiam o desenvolvimento de um modelo científico para a PC defendem e promovem o uso de pequenos corpora para a investigação de fenômenos pragmáticos (CLANCY; O'KEEFFE, 2015; RÜHLEMANN; CLANCY, 2018; VAUGHAN; CLANCY, 2013).

Isso não significa, contudo, que estudos baseados em PC podem apenas ser realizados quando pequenos corpora são utilizados. Grandes e megacorpora podem ser, e já foram, utilizados em pesquisas de PC (cf. VAUGHAN; CLANCY, 2013; O'KEEFFE *et al.*, 2020 para a descrição de alguns exemplos e soluções). Entretanto, quando se trata de análise qualitativa, é comum que os pesquisadores tenham de reduzir a amostra a fim de avaliar e analisar as ocorrências linguísticas com a perspectiva pragmática.

O'Keeffe (2018) aponta para um outro desafio que, por sua vez, afeta diretamente um dos temas mais espinhosos no desenvolvimento de um modelo de PC: a questão da forma e função. Vejamos que tal desafio se dá pela oposição metodológica entre a LC e a Pragmática. De um lado, temos a LC que possui uma metodologia de análise ascendente, ou seja, de baixo para cima (*bottom-up*) a partir da forma. Para isso,

¹⁰ O conceito de pequeno corpus é abstrato, uma vez que ainda não há uma concordância entre os linguistas de corpus a esse respeito. Pode-se dizer que um pequeno corpus é um que contenha menos que um milhão de palavras. Contudo, no âmbito da Pragmática de Corpus, um pequeno corpus que permita etiquetamento manual, processamento e análise de dados, varia entre 50 e 500 mil palavras. Há ainda estudos pragmáticos de corpus que utilizam corpora menores que 50 mil palavras, como Vaughan e Clancy (2013) e McAllister (2015).

a investigação ocorre a partir da observação do corpus por meio de percursos metodológicos específicos, como a extração de listas de palavras classificadas por ordem de frequência, listas de palavras-chave com base em um corpus de referência, agrupamentos lexicais e linhas de concordância. Nesse sentido, “os estudos pragmáticos baseados em corpora são geralmente baseados na forma e começam mapeando palavras ou construções em uma série de funções”¹¹ (AIJMER, 2018, p. 555). No entanto, existem aspectos negativos desta abordagem para a análise pragmática, atrelados à diversidade funcional e ambiguidade (O’KEEFFE, 2018). Isto porque em uma abordagem que toma a forma como o ponto de partida para, então, conduzir uma análise das funções de dado item linguístico (abordagem forma-função), nos deparamos com o desafio de filtrar as funções pragmáticas e não pragmáticas que este performa no discurso. Durante o processo de filtragem, há ainda a possibilidade de os resultados apresentarem ocorrências ambíguas, como também aquelas que demonstram uma diversa seleção de funções pragmáticas para o mesmo item.

Por outro lado, temos a Pragmática com uma metodologia tradicionalmente descendente, ou seja, de cima para baixo (*top-down*) a partir da função. Para tanto, a investigação inicia-se por uma determinada função (*e.g.* atos de fala) e caminha em direção às formas que a desempenham. Este modelo de abordagem efetua-se através de instrumentos mais diretos, partindo da coleta de produção elicitada, em contexto controlado, e por meio de técnicas como entrevistas, testes de complementação discursiva (TCDs), *role-plays*, diários, entre outros. Como ponderam O’Keeffe *et al.* (2020), é inegável que o campo da Pragmática possui abordagens tradicionais de investigação já estabelecidas e, conseqüentemente, a transposição da metodologia da LC para a pragmática apresenta desafios. Isto porque uma função pragmática pode apresentar inúmeras formas linguísticas. Ou seja, em uma abordagem que toma a função como ponto de partida para, então, recuperar as formas linguísticas que a desempenham (abordagem função-forma), o pesquisador enfrentará a dificuldade de não extrair todas as formas de uma função, à exceção de casos em que o corpus é manual e pragmaticamente anotado.

¹¹ Nossa tradução para: “Corpus-based pragmatic studies are generally form-based and they start by mapping words or constructions onto a range of functions.”

Podemos, assim, considerar que as observações descritas acima são limitações metodológicas em ambas as áreas. Nesse panorama, como já antecipado, dentro de um modelo para a PC, precisamos encontrar um equilíbrio entre as duas abordagens, de maneira que o melhor que cada uma tem a oferecer para os estudos pragmáticos baseados em corpora possa ser utilizado na investigação de quaisquer perguntas norteadoras a serem abordadas dentro de um quadro teórico-metodológico de PC (O'KEEFFE *et al.*, 2020).

Adiante, apresentamos ambas as abordagens para os estudos da PC, forma-função e função-forma, bem como seus modelos metodológicos que, junto ao modelo de leitura integrada de Aijmer e Rühlemann (2015) previamente apresentado, constroem o quadro metodológico da nova disciplina.

4.1.1 Abordagem forma-função

A abordagem forma-função na PC é herança da LC, e tem servido bem aos estudos pragmáticos com base em corpora, cujos interesses são o mapeamento das funções pragmáticas desempenhadas por específicas formas linguísticas. Dentro de um modelo metodológico de PC, a abordagem forma-função é ideal para a investigação de fenômenos pragmáticos cujas formas são bem definidas e menos variáveis do que, por exemplo, os atos da fala. Fenômenos pragmáticos que se adaptam bem a esta abordagem incluem dêixis, marcadores pragmáticos (abrangendo todas as subcategorias deste grupo, como marcadores discursivos, marcadores de hesitação, tokens de resposta, mitigadores, etc.) e itens lexicogramaticais de vagueza.

O início da PC deu-se, majoritariamente, com a utilização da abordagem forma-função e, ainda hoje, estudos pragmáticos com base em corpora são geralmente abordados tendo a forma como o ponto de partida (AIJMER, 2018; JUCKER 2013). Isto se justifica pelo fato de a LC oferecer uma metodologia de busca e análise linguística rigorosa e já bem definida, portanto, mais proeminente e dominante, como também pelo fato de a LC encontrar certas dificuldades quando a abordagem inicia-se de maneira inversa, i.e. função-forma (Seção 4.1.2 abaixo). Aijmer (2018, p. 555) argumenta, contudo, que uma grande vantagem em conduzir uma análise forma-função dentro de um modelo de PC é a de que uma dada forma sob investigação pode ser estudada em detalhe e precisão, levando-se em consideração aspectos como frequência,

posição sintática, prosódia semântica, colocação e coligação. Ademais, o fato de que múltiplas funções pragmáticas de uma determinada forma linguística podem ser evidenciadas, avaliadas e categorizadas, faz da PC uma ferramenta que oferece uma nova e significativa perspectiva de investigação aos estudos de fenômenos pragmáticos.

Neste contexto, O’Keeffe *et al.* (2020) apresentam quatro modelos de estratégias investigativas que compõem a abordagem forma-função dentro do quadro metodológico da PC. Os modelos foram adaptados daqueles propostos por Ädel e Reppen (2008) para os estudos baseados em corpus de maneira geral. São eles: (a) *busca direta (one-to-one searching)*, (b) *amostragem (sampling)*, (c) *filtragem (sifting)*, e (d) *listagem baseada em frequência (frequency-based listing)*. Como é apontado pelos autores na proposta original, e reiterado pelos autores da recente adaptação para a PC, estes modelos mesclam-se com frequência, ou seja, uma pesquisa pode implementar apenas um modelo como também uma combinação de dois ou mais modelos.

A **busca direta** é aquela na qual o pesquisador insere uma forma para busca e recupera cem por cento de ocorrências relevantes. Contudo, tal abordagem só é possível uma vez que o corpus tenha sido previamente etiquetado. Por exemplo, consideremos um pesquisador que se interesse pelo vocativo *querido(a)* (e.g. *querido, não faça isso*). Para a recuperação total de ocorrências relevantes, o corpus necessita de anotação pragmática para que ocorrências do adjetivo ou substantivo *querido(a)* (e.g. *amigo querido; aproveitaram para cantar parabéns para o querido*) sejam automaticamente excluídas da busca e, assim, o pesquisador tenha em mãos apenas uma lista de ocorrências do fenômeno de seu interesse. Com todas as ocorrências pragmáticas da forma *querido(a)* recuperadas e isoladas automaticamente, o pesquisador segue para a análise qualitativa e contextual das suas funções.

Por sua vez, o modelo de **amostragem** diz respeito àquela abordagem em que o pesquisador criteriosamente seleciona um ou mais itens de busca que representam um certo fenômeno pragmático. Ädel e Reppen (2008, p. 3) alertam para o fato de que há uma desvantagem na utilização deste modelo investigativo: apenas um subconjunto de casos de determinado fenômeno é investigado, e não o fenômeno em sua totalidade. Ainda assim, a partir de uma cuidadosa avaliação e desenho de um quadro de amostragem, o emprego de itens de busca representativos tem grande potencial para a recuperação de ocorrências relevantes. Os

itens de busca representativos de um fenômeno pragmático podem ou não ser linguísticos. No caso de atos de fala, por exemplo, itens de busca que podem representar o agradecimento incluem dispositivos indicadores de força ilocutória (DIFIs) como *(muito) obrigado(a)*, *eu agradeço*, *estou (muito) agradecido(a)* e *Deus lhe pague*. Por outro lado, itens de busca não linguísticos variam de informações extralinguísticas anotadas no corpus ao uso de códigos de transcrição (cf. VAUGHAN, 2008, para um estudo em que *laughs* e *laughter* são usados como itens de busca para o resgate de ocorrências de humor; CLANCY; MCCARTHY, 2015, para um estudo em que etiquetas de falantes, e.g. <\$1>, são utilizadas como itens de busca para resgatar casos em que *if* e *when* são empregados como iniciadores de turno). Estudos que lançam mão deste modelo investigativo certamente fazem também o uso da estratégia de filtragem, uma vez que o relacionamento de uma forma linguística com uma função pragmática específica não é garantido.

Filtragem é o modelo de estratégia investigativa no qual o pesquisador necessita de uma leitura manual das ocorrências obtidas em uma busca para que, desta maneira, a separação entre ocorrências relevantes e não relevantes seja manualmente efetuada. Uma vez que uma primeira leitura superficial de exclusão é concluída, o pesquisador pode seguir para uma segunda leitura, mais detalhada e minuciosa, das ocorrências relevantes para a análise, determinação e descrição das funções de dada forma. A filtragem é especialmente útil para estudos que utilizam corpora não etiquetados, iniciando a análise a partir da forma. É também empregada em estudos que se baseiam em amostras, principalmente aqueles que empregam itens de busca não linguísticos. Contudo, estudos que não se baseiam em amostra, mas cujo foco está em determinada e específica forma linguística a fim de mapear suas funções pragmáticas, também se enquadram bem dentro deste modelo. Por exemplo, consideremos um estudo em que a análise foca nas funções que são performadas pelo marcador pragmático *tipo* na língua portuguesa oral do Brasil. A fim de analisar contextualmente o marcador pragmático e descrever as suas funções no discurso, o pesquisador precisa, primeiramente, ‘separar o joio do trigo’; ou seja, eliminar ocorrências nas quais *tipo* não atua em uma função pragmática (e.g. *eu prefiro outro tipo de comida*).

Por fim, temos a **listagem baseada em frequência**. Neste modelo investigativo, que mais se assemelha a uma metodologia típica

de LC, o pesquisador inicia a pesquisa a partir de listas de frequência. Ou seja, listas de frequência, sejam elas de um corpus singular ou de comparações entre corpora, são analisadas com o objetivo de identificar formas salientes (palavras ou colocações) que possam indicar fenômenos pragmáticos. Neste sentido, a pesquisa e seus itens de busca são delimitados pelo corpus e seu contexto discursivo em particular e, de acordo com Ádel e Reppen (2008, p. 3), esta é uma estratégia efetiva para destacar padrões linguísticos de um banco de dados específico. O’Keeffe *et al.* (2020, p.53) afirmam que variações em frequência, distribuição e padrões linguísticos, são comumente justificadas através de conclusões pragmáticas em estudos com base em listas de frequência. De fato, itens linguísticos de natureza pragmática são omnipresentes no discurso oral, e mesmo abordagens guiadas por corpus (diferentes das baseadas em corpus) têm uma alta probabilidade de apresentar resultados que apontam para o sistema pragmático da linguagem em uso (cf. O’KEEFFE *et al.*, 2007, capítulos 2 e 8).

Um exemplo de estudo guiado por corpus é o de Santos (2019). Neste estudo, o pesquisador usa uma amostra do *Spoken Corpus of Brazilian Portuguese and L2-English* (SCoPE²) (SANTOS, 2020) e compara uma lista das 15 palavras mais frequentes com uma mesma lista feita a partir do *Limerick Corpus of Irish English* (LCIE) (FARR *et al.*, 2004), a fim de determinar diferenças e similaridades entre nativos do inglês irlandês e brasileiros universitários na Irlanda. Nesta etapa, *like* foi identificada como uma palavra merecedora de maior atenção, estando colocada na décima quarta posição da lista SCoPE², enquanto na terceira posição na lista LCIE. Uma segunda etapa de listagem incluiu uma lista de palavras-chave provenientes do SCoPE² quando este corpus foi contrastado e comparado ao LCIE como o corpus referência para o estudo. Neste momento, *like* ocupou a posição de terceiro lugar, confirmando sua saliência no SCoPE². Após uma filtragem manual de ocorrências pragmáticas e não pragmáticas, Santos (2019) pôde confirmar um total de 85% de ocorrências de *like* funcionando como um marcador pragmático. Em sua etapa qualitativa, o estudo demonstra que *like* apresenta as mesmas sete funções em ambos os corpora (mitigador, dispositivo de aproximação, exemplificador, marcador de hesitação, marcador de foco e dispositivo de discurso relatado), com diferenças concernentes apenas à quantidade de ocorrências em cada função, como também à posição sintática do marcador pragmático. Este estudo exemplifica e demonstra

a adequabilidade da estratégia de listagem baseada em frequência como um ponto de partida inicial para a investigação de formas linguísticas que se destacam em corpora de contextos linguísticos específicos devido às suas funções pragmáticas.

É inegável que a abordagem forma-função, com seus quatro modelos investigativos apresentados aqui, oferece ferramentas eficazes para o estudo de uma grande variedade de fenômenos pragmáticos baseados em corpora. Contudo, há também fenômenos que são mais efetivamente abordados dentro de um quadro metodológico reverso (função-forma), devido à dificuldade de se recuperar e identificar toda a abrangência de formas que os caracterizam. Esta abordagem reversa, embora mais complexa e justificavelmente mais trabalhosa, tem ganhado a atenção dos pesquisadores de PC nos últimos anos, e estes (especialmente os da escola europeia) têm se debruçado sobre o assunto a fim de desenvolverem um modelo metodológico inclusivo para a PC; ou seja, que inclua ambas as abordagens. Abaixo, apresentamos uma proposta de modelos investigativos que constituem a abordagem função-forma.

4.1.2 Abordagem função-forma

A abordagem função-forma espelha-se na metodologia tradicional pragmática na qual o ponto de partida acontece através de uma determinada função e investiga as formas que emergem a partir de contextos controlados por mecanismos de elicitación. Se, por um lado, a LC mostra-se uma perfeita aliada à Pragmática pela abordagem forma-função, como observamos na seção anterior, por outro, um dos desafios metodológicos da PC é identificar, em corpora de estudo, fenômenos pela abordagem função-forma que são variáveis e difíceis de ser identificados. Isso se dá porque as buscas pelas funções não são realizadas de forma automática e precisam ser bem delimitadas.

A metodologia pragmática partindo da função ainda é incipiente no campo da PC. Isso se justifica pela dificuldade em extrair as ocorrências de dada função, uma vez que para a maioria dos fenômenos pragmáticos, a correlação forma e função é inexistente (RÜHLEMANN; AIJMER, 2015). Ao mesmo tempo em que temos a técnica tradicional de elicitación de dados com controle do contexto situacional dos estudos pragmáticos a partir da função, acessamos uma menor quantidade de

dados. Em contrapartida, temos uma grande variedade de formas nos corpora disponíveis e pouca riqueza contextual, a menos que o corpus seja compilado pelo próprio pesquisador (O'KEEFFE, 2018).

É nesse desafio para equilibrar função, contexto, forma e dados linguísticos em grandes quantidades, que apresentamos e discutimos a seguir os modelos de estratégias investigativas, a partir de O'Keeffe *et al.* (2020), que compõem a abordagem função-forma dentro do quadro metodológico da PC. São eles: (a) *busca direta (one-to-one searching)*, (b) *amostragem, busca e filtragem (sampling, searching and sifting)*, (c) *sementes (seeds)*, e (d) *mapeamento pragmático indireto* (originalmente nomeado pelos autores como *solutions for larger corpora*, mas aqui apresentado com a terminologia adaptada).

A **busca direta** apresenta maiores dificuldades na abordagem função forma. Tomando como base os atos de fala em um corpus, McAllister (2015, p. 29) adverte-nos que neste tipo de busca o pesquisador limita-se às formas linguísticas que considera atuar de forma pragmática, não sendo suficiente para destacar as formas de tal fenômeno em sua totalidade, o que descarta, principalmente, as formas não presumidas como tendo valor pragmático. Desta forma, a recuperação de todas as ocorrências relevantes da busca direta nessa perspectiva, que parte da função, só é possível por meio de corpora pragmaticamente anotados, considerados os 'cálices sagrados' para a área da PC (O'KEEFFE *et al.*, 2020).

A construção de corpora orais não é tarefa fácil, e isso se reflete no grande número de corpora escritos que, presumivelmente, ainda dominam a área da LC. Isso sucede porque a compilação envolve uma série de procedimentos, como o desenho do corpus, a disponibilidade dos informantes, o processo de gravação, a elaboração dos critérios de transcrição, o processo transcrito em si, entre outros. Nesse sentido, o número de corpora orais anotados pragmaticamente é ainda menor, mas se torna necessário dependendo do tópico a ser investigado. Apesar de a anotação de corpora ser um processo manual maçante, esta realidade está gradativamente mudando nos últimos anos com o advento de novas ferramentas e sistemas de anotação que, posteriormente, nos permitem a extração automática de um determinado fenômeno pragmático, como o *Dialogue Annotation and Research Tool* (DART) (WEISSER, 2015). Segundo Weisser (2019, p. 131), a ferramenta possibilita pesquisas com grandes corpora não apenas para investigações de aspectos sintáticos,

mas também pragmáticos, como DIFIs, atos de fala, entre outras características da interação.

No Brasil, destacamos o C-ORAL-BRASIL (RASO; MELLO, 2012), corpus de fala espontânea informal do português brasileiro que surgiu a partir do C-ORAL-ROM (CRESTI; MONEGLIA, 2005), por sua vez composto por línguas românicas europeias (espanhol, francês, italiano e português europeu). A estrutura do C-ORAL-BRASIL permite o estudo da fala sob uma perspectiva pragmática devido ao seu design e ampla variedade situacional (RASO, 2012). Para isso, o corpus é segmentado em enunciados (atos de fala) e em unidades tonais, respectivamente com barra dupla (//) e barra simples (/), e símbolos para indicação de enunciados interrompidos (+) e retratações ([/n]). Ademais, as transcrições e o material acústico são alinhados. Todo o processo de segmentação do C-ORAL-BRASIL foi validado para a redução de desacordos entre os membros do projeto. Como exposto, devido à anotação de um corpus grande ser desafiadora, uma saída para o estudo de caráter pragmático que os pesquisadores têm encontrado é a anotação de corpora menores que possibilitam uma anotação mais precisa. Ainda assim, como aponta O’Keeffe (2018), o trabalho requer critérios de validação.

Exemplos de corpora internacionais anotados incluem o subcorpus do *Michigan Corpus of Academic Spoken English* (MICASE), manualmente anotado para atos de fala (MAYNARD; LEICHER, 2007) e o *Systems of Pragmatic annotation in ICE-Ireland* (SPICE-Ireland) Corpus, um projeto desenvolvido a partir do subcorpus do *International Corpus of English-Ireland* (ICE-Ireland) (KALLEN; KIRK, 2012), e são considerados avanços para a área. O SPICE-Ireland é anotado pragmática e prosodicamente e possui aproximadamente 600,000 palavras. Dessa forma, por exemplo, o status do ato de fala de cada enunciado foi identificado com etiquetas, tendo Searle (1976) como base: representativas <rep> ... </rep>; diretivas <dir> ... </dir>; comissivas <com> ... </com>; expressivas <exp> ... </exp> e declarativas <decl> ... </decl>. Além disso, outras marcações foram usadas, como (*) para marcadores discursivos, (+) para indicar discursos relatados, (%) em final de unidade entonacionais, entre outras combinações. Portanto, num modelo de busca direta, dentro da abordagem função-forma, o uso de um corpus pragmaticamente anotado como o SPICE-Ireland possibilita a extração total de formas que representam um determinado fenômeno pragmático. Ou seja, um pesquisador interessado em atos da fala com

foco, por exemplo, em expressivas do inglês irlandês, tem a possibilidade de estudar todas as formas desta função naquela variedade linguística com o auxílio das etiquetas que foram previamente anotadas no corpus. Em outras palavras, todas as formas de uma determinada função pragmática são recuperadas através do uso de itens de busca com base em etiquetas e códigos anotados.

Corpora pequenos, apesar de não serem ideias para certos estudos como os lexicais e fraseológicos, podem ser adequados a outros (KOESTER, 2010). Nesse sentido, especialmente a partir de perguntas de pesquisa e contextos bem delimitados, conseguimos anotar e extrair fenômenos pragmáticos específicos para análise. Já considerando a abordagem função-forma para investigar corpora maiores, precisaremos combinar modelos investigativos a fim de delimitar os dados.

A **amostragem, busca e filtragem** é a combinação de modelos investigativos, aplicada a corpora maiores, que nos possibilita uma análise pragmática partindo da função. A transformação de um corpus grande em uma amostra menor é ideal para a PC. Com o objetivo de tornar determinado corpus mais facilmente manejável e a investigação completa de determinada função pragmática, este modelo inicia-se através da escolha de uma amostragem aleatória de determinado corpus. A partir desta amostra, estabelecemos os parâmetros para identificar os itens e conduzimos uma filtragem manual, ao mesmo tempo em que conduzimos o processo de anotação de todas as ocorrências relevantes nos dados. Dessa forma, podemos, através da anotação por uma leitura qualitativa, categorizar todos os fenômenos restantes. Um dos principais benefícios deste modelo investigativo, é que, se o pesquisador ainda tiver acesso aos metadados, este terá em mãos um número maior de ocorrências significativas do que um estudo tradicional pragmático possibilitaria, além das informações contextuais detalhadas das situações.

Por sua vez, a busca a partir de **sementes**, isto é, resultados de pesquisas anteriores, é outro método que pode ser aplicado a corpora maiores. Os resultados provenientes de pesquisas pragmáticas, através dos variados instrumentos de coleta ao longo dos anos, constituem um rico banco de dados neste modelo de investigação. Dessa forma, o pesquisador não desconsidera o que foi produzido e tem as sementes como ponto de partida para análise das funções pragmáticas em corpora. Considerando as pesquisas pragmáticas com TCDs, por exemplo, podemos considerar resultados de estudos como os de Beebe *et al.* (1990), que investigaram

a transferência pragmática em recusas de inglês como segunda língua em um grupo de japoneses nativos de japonês, japoneses falantes de inglês como segunda língua e americanos nativos de inglês. Os atos de fala de recusa foram analisados em pedidos, convites, ofertas e sugestões, considerando ainda variáveis sociais. Segundo os autores (p. 57), se um participante, na recusa de um convite para um jantar realizado por amigos, proferir um ato de recusa como *'I'm sorry, I have theater tickets that night. Maybe I could come by later for a drink'*, teremos uma fórmula como [expression of regret] [excuse] [offer of alternative]. Assim, neste modelo de análise da função-forma, podemos considerar formas específicas e fórmulas oferecidas pelo estudo de Beebe *et al.* (1990) como sementes para a busca em outros corpora, considerando ainda contextos situacionais similares, obtendo um conjunto de dados comparáveis. Tem-se, nesse caso, modelos de formas existentes para extrair fenômenos específicos em corpora e gerar um subcorpus comparável pelas mesmas condições (O'KEEFFE, 2018).

4.1.2.1 Mapeamento pragmático indireto na função-forma

Como vimos até aqui, as formas de refinamento na abordagem função-forma exigem que o pesquisador busque mecanismos diferentes para extrair dados relevantes. Outros métodos que podem ser usados como forma de investigação, como apontado por O'Keeffe *et al.* (2020), consistem no uso de (a) *DIFIs*, (b) *busca por léxico ou características gramaticais associadas a um ato de fala*, e (c) *busca por expressões metacomunicativas*. Destacamos que estes métodos são apresentados pelos autores como constituintes de um conjunto de ações a serem aplicadas como soluções metodológicas para o uso de corpora de maior escala. Contudo, consideramos tais modelos investigativos como parte de um mapeamento pragmático indireto que consiste em mecanismos de buscas mediadas por itens que objetivam abranger determinadas funções pragmáticas ao máximo. Como apresentado anteriormente, soluções metodológicas para grandes corpora que não são pragmaticamente anotados relacionam-se, frequentemente, com estratégias como amostragem e filtragem, que, por sua vez, são invariavelmente aplicadas aos modelos apresentados dentro do escopo de mapeamento pragmático indireto, assim justificando uma adaptação de terminologia para o último modelo que constitui a abordagem função-forma.

Nesse sentido, a partir do refinamento de dados por meio do uso dos DIFIs, considerando o ato de fala expressivo na língua inglesa, tomamos como exemplo o ato de desculpar-se e sua realização pelo DIFI *sorry*. Dessa forma, teremos várias ocorrências para esta busca, mas observamos que o uso de DIFIs na PC está relacionado à precisão e busca limitados. Isso se dá porque quando buscamos por *sorry*, deparamo-nos com ocorrências não relacionadas ao ato de desculpar-se, com o item de busca aparecendo, por exemplo, como atributivo (JUCKER, 2013). Da mesma forma, a busca é limitada, pois nem todos os pedidos de desculpas incluem *sorry*, e tal DIFI pode não ser representativo deste ato. Assim, para um mapeamento mais abrangente, precisaremos incluir outros DIFIs como *my mistake*, *pardon* ou *excuse me*. No entanto, o processo de filtragem neste processo é indispensável.

Como extensão da busca por DIFIs, consideramos o uso do léxico ou expressões gramaticais associadas a determinados atos de fala (JUCKER, 2013). Em um estudo sobre elogios através de diferentes culturas, Wolfson (1981, p. 122) mostra que estes são altamente convencionalizados no inglês americano em estruturas sintáticas como *NP [is/looks] (really) ADJ* ou *I (really) [like/love] NP*. Destacamos aqui a importância de termos um corpus etiquetado morfossintaticamente que ajudará nesse tipo de busca. A não extração de todas as formas, principalmente as mais incomuns, é uma limitação, além de não ser uma metodologia direta, já que precisamos transformar expressões em fórmulas específicas (JUCKER, 2013).

Por fim, expressões metacomunicativas são ‘palavras e frases que podem ser usadas para falar sobre aspectos da comunicação, no sentido de que nomeiam um determinado ato de fala, como elogio, saudação, insulto ou agradecimento, ou um tipo específico de comportamento, como a polidez e a impolidez¹² (JUCKER *et al.*, 2012). Portanto, podemos examinar como as pessoas concebem e avaliam o uso da linguagem (HAUGH, 2018). No entanto, a análise da metalinguagem ou expressões metacomunicativas necessita do suporte de megacorpora para a obtenção de ocorrências significativas, e, ainda que o acesso a evidências de

¹² Nossa tradução para: “[metacommunicative expressions are] words and phrases that can be used to talk about aspects of communication, in the sense that they name a particular speech act, such as compliment, greet, insult or thank, or a particular type of behaviour, such as polite or impolite”

determinados elementos a partir de como as pessoas os avaliam seja indireto, temos em mãos dados importantes de natureza etnográfica (JUCKER, 2013). Nesse sentido, o estudo de Culpeper (2009) demonstra isso ao investigar a metalinguagem da impolidez por acadêmicos e não acadêmicos no *Oxford English Corpus* com aproximadamente 2 bilhões de palavras. O autor teve como ponto de partida uma lista de rótulos usados na literatura sobre polidez e impolidez, como *impolite(ness)*, *rude(ness)*, *aggravation*, *aggravated/aggravating*, *language/facework (aggravated impoliteness)*, *aggressive facework*, *face attack* e *verbal aggression*. Culpeper revela que *rude* é frequente na linguagem de não acadêmicos e, na maioria das vezes, ocorre em contextos públicos, além de que entre os sujeitos tidos como *rude*, estão porteiros, garçons/garçonetes, equipe. Curiosamente, as pessoas também fazem menção à ‘Yorker’ e ‘French’, reforçando estereótipos de lugares.

Discorreremos, nas duas últimas seções, os oito modelos investigativos de análise que integram as abordagens forma-função e função-forma, essenciais para as investigações de fenômenos pragmáticos em corpora. Na próxima seção do artigo, escolhemos um modelo investigativo, a **filtragem**, pela abordagem forma-função, para ilustrar em um estudo de caso como a análise do marcador pragmático *kind of* pode ser conduzida.

5 Estudo de caso: *kind of* no inglês de brasileiros universitários

Segundo Jucker *et al.* (2018, p. ix), PC é uma das três abordagens metodológicas empíricas para a análise de fenômenos pragmáticos (sendo as outras duas a pragmática experimental e a pragmática observacional). Ademais, dentre os fenômenos pragmáticos, os marcadores pragmáticos (MPs) são considerados como umas das áreas-chave da pesquisa pragmática de corpus (CLANCY; O’KEEFFE, 2015). MPs são uma extensa e eclética classe de itens linguísticos que se manifestam em formas de palavras (*e.g. like* no inglês, e *tipo* no português) ou fórmulas linguísticas padronizadas (*e.g. you know* e *you know what I mean* no inglês, e *olha só* e *e por aí vai* no português), muitas vezes também manifestando-se em formas que possam não ser consideradas como palavras (*e.g. interjeições: ah, oh; tokens responsivos: mmhm/hum hum*). Estes itens linguísticos marcam as atitudes e posicionamentos dos falantes, e atuam de maneira interpessoal no ato da conversação, além

de funcionarem como um auxílio na organização estrutural do discurso (cf. BRINTON, 1996; CARTER; McCARTHY, 2006).

Assim, em consideração à natureza pervasiva dos MPs na linguagem, como também atentos à sua importância pragmática no processo de interpretação de sentidos na comunicação, apresentamos agora um estudo de caso comparativo, no qual analisamos o MP *kind of* no inglês falado por brasileiros universitários no Brasil e na Irlanda. O estudo de caso apresentado tem objetivo duplo: (a) exemplificar o modelo da PC em um estudo que considera um fenômeno pragmático com base em corpora, e (b) contribuir com o campo de estudos que visa a melhor compreensão do uso de MPs em língua adicional/estrangeira (L2). Apesar de a produção acadêmica com foco em MPs na língua materna (L1) não ser escassa, muito ainda precisa ser estudado e compreendido com respeito ao uso de MPs na L2. Em sua maioria, os estudos de MPs na L2 tendem a comparar a produção linguística do não-nativo com aquela do nativo. Embora tal comparação seja positiva, por destacar características particulares da L2, alertamos para o fato de que tal comparação não deve atrelar-se à perspectiva de linguagem deficiente com a qual muitos destes estudos são relacionados quando descrevem a linguagem do não-nativo que não atinge uma norma nativa (cf. PRODRUMOU, 2008). Neste sentido, o presente estudo não só aborda L2 de uma perspectiva de competência, como também compara L2 com L2, a fim de investigar possíveis diferenças no uso do MP *kind of* entre dois grupos distintos de brasileiros universitários falantes do inglês.

O sentido nuclear de *kind of/sort of* é o de aproximação e imprecisão. Ao fazerem uso deste MP, os falantes indicam aos seus interlocutores que o material linguístico que o sucede deve ser interpretado de maneira vaga e imprecisa. Ou seja, *kind of* (como também sua versão mais usada no inglês britânico, *sort of*) sinaliza uma informação de imprecisão, e convida o interlocutor a aproximar e ajustar o material linguístico para que a interpretação intencionada seja feita. Estudos que investigaram *kind of/sort of* em L1 destacam que estes MPs operam nos três principais domínios funcionais da classe: atitudinal, interpessoal e textual (cf. AIJMER, 2002; KIRK, 2015). No domínio atitudinal (ou de posicionamento), o falante marca que há uma discrepância entre o que este tem em mente e o material linguístico falado. Neste domínio, *kind of/sort of* apresenta um status especial quando comparado com a maioria dos MPs, uma vez que sua presença afeta diretamente o sentido do conteúdo proposicional (AIJMER, 2002). No domínio interpessoal

(ou interativo), *kind of/sort of* sinaliza que o material linguístico marcado deve ser interpretado de maneira não assertiva e final. Em outras palavras, neste domínio, o falante leva em consideração a maneira que sua escolha de palavras será interpretada pelo seu interlocutor e, portanto, *kind of/sort of* é usado para mitigar a força ilocutória do material linguístico que o sucede. Finalmente, no domínio textual, o MP é utilizado para sinalizar uma necessidade de ajustamento por parte do próprio falante com relação à estrutura do seu discurso. Neste domínio, *kind of/sort of* auxilia o falante com a reestruturação semântica e sintática do conteúdo comunicado, indicando também, a nível interpessoal, que o falante deseja manter o turno e continuar seu discurso.

No que diz respeito aos estudos investigativos do MP *kind of/sort of* em L2, destacamos três que lançaram mão do *Louvain International Database of Spoken English Interlanguage* (LINDSEI) a fim de apontar diferenças no uso deste MP entre nativos e não-nativos. O primeiro, Gilquin (2008), investiga marcadores de hesitação no discurso de aprendizes em nível avançado de inglês com língua materna francesa. Para tanto, a autora compara o LINDSEI-FR com o *Louvain Corpus of Native English Conversation* (LOCNEC). Dentre os marcadores de hesitação analisados, a autora inclui uma subcategoria de MPs, da qual *kind of/sort of* é pertencente. Os resultados de sua pesquisa demonstram que este coorte de aprendizes faz grande uso de pausas, sejam elas preenchidas ou silenciosas, mas não conseguem explorar a variedade de MPs que desempenham a mesma função. A autora também sugere, com base em seus resultados, que essa má representação de MPs pode ser devido a uma superdependência do MP *well*, que apresentou grande frequência no LINDSEI-FR. Com referência ao MP *kind of/sort of*, a autora destaca que *sort of* é usado, funcionalmente, de maneira diferente quando comparado com nativos, uma vez que estes aprendizes fazem um maior uso da sua função textual devido a limitações de vocabulário. Apesar de a maioria dos exemplos apresentados pela autora serem discutivelmente casos de hesitação, no sentido de reformulação e ganho de tempo para recuperação cognitiva de vocabulário, o uso de *kind of/sort of* nesta função por não-nativos é previsto. Tal função já foi evidenciada no discurso nativo do inglês irlandês (KIRK, 2015), e do inglês britânico (AIJMER, 2002), sendo uma estratégia conversacional crucial. Sendo assim, é de se esperar que não-nativos, especialmente aprendizes, façam uso da função textual do MP *kind of/sort of* em sua busca por formulação linguística adequada em uma L2.

Buyse (2010), por sua vez, analisa o inglês falado por aprendizes avançados e nativos da língua holandesa. O autor investiga o uso de uma seleção de alguns dos MPs mais comuns da língua inglesa: *so*, *well*, *you know*, *like*, *kind of/sort of*, e *I mean*. Este é um estudo quantitativo e apresenta resultados de frequências de ocorrências de cada um dos MPs investigados de maneira comparativa entre os corpora LINDSEI-DU e LOCNEC. Os resultados apresentados mostram que os aprendizes do LINDSEI-DU raramente fazem uso de MPs com funções interpessoais (*you know*, *like*, *kind of/sort of*, *I mean*), enquanto aqueles com funções textuais são usados em demasia. Apesar da importância funcional dos MPs de funções interpessoais, o autor argumenta que um motivo para sua baixa frequência no LINDSEI-DU, quando comparado ao LOCNEC, é o fato de tais MPs serem relacionados à linguagem informal e, muitas vezes, serem estigmatizados.

Finalmente, Miranda (2020) investiga o uso de marcadores de linguagem vaga por brasileiros aprendizes de inglês em nível avançado e americanos nativos da língua inglesa. Comparando os corpora LINDSEI-BR e SBCSAE (*Santa Barbara Corpus of Spoken American English*), o autor faz uma análise quantitativa de vários marcadores de linguagem vaga, destacando o MP *kind of/sort of* por ser o mais frequente em ambos os corpora. Contudo, e interessantemente, apesar de os aprendizes brasileiros não fazerem uso da forma *sort of*, enquanto os falantes americanos o fazem com ambas as formas, LINDSEI-BR apresenta uma frequência maior do que no SBCSAE quando somados *kind of* e *sort of* juntos. Na etapa qualitativa de seu estudo, Miranda (2020) identifica as funções pragmáticas exercidas pelo MP *kind of* em ambos os corpora, e nota que a função mais usada pelos aprendizes brasileiros é atitudinal, marcando inexistência do material linguístico que o sucede, seguida da função interpessoal de mitigação. No SBCSAE, contudo, a mitigação marcada pelo *kind of* é mais frequente que no LINDSEI-BR, e casos de *kind of* marcando vocabulário técnico ou complexo, como também linguagem vulgar, são encontrados apenas no SBCSAE.

O que estes estudos sobre o uso de MPs na linguagem do aprendiz revelam é que, apesar de algumas limitações, MPs são também presentes no discurso oral do aprendiz e têm um papel importante na L2. Por equiparem falantes de L2 em um nível interpessoal quando usados com sucesso, e por limitarem a contribuição destes falantes em conversação quando ausentes, MPs são indispensáveis para uma comunicação bem

sucedida. Considerando este panorama apresentado, o presente estudo contribui para o corpo de estudos em MPs na L2 ao analisar *kind of* no discurso oral de brasileiros falantes de inglês em dois contextos diversos.

5.1 Corpora e metodologia

Os dados sob investigação no presente estudo são provenientes de dois corpora: o SCoPE² e o *Brazilian Spoken English Learner Corpus* (BraSEL), e apenas um subcorpus de cada é utilizado. O SCoPE² e o BraSEL foram desenvolvidos para projetos de doutorado com o objetivo de analisar a pragmática do inglês falado por brasileiros, porém em contextos linguísticos distintos. O primeiro é um corpus bilíngue composto pelo inglês falado por universitários na Irlanda e pelo português nativo destes mesmos participantes (cf. SANTOS, 2020 para uma descrição detalhada do processo de desenho, compilação e transcrição do corpus). O material linguístico do SCoPE² representa a linguagem informal em uso real, e foi coletado através de encontros informais de bate-papo entre o pesquisador e os participantes. É importante ressaltar que este não é um corpus de aprendizes, mas sim de inglês como língua adicional, uma vez que todos os participantes terminaram os seus cursos de língua inglesa, obtiveram com sucesso um certificado internacional de proficiência em nível avançado ou proficiente, e se comunicam eficientemente dentro de um ambiente internacional com ambos nativos e não-nativos da língua inglesa a nível pessoal, profissional e cultural. Nesse sentido, os participantes do SCoPE² são considerados *successful users of English* (SUEs; PRODROMOU, 2008).

Em contrapartida, o BraSEL é um corpus que está em fase de compilação, e é composto por material linguístico coletado de brasileiros universitários no Brasil que, seguindo a tabela de níveis de proficiência do Quadro Europeu Comum de Referência para as Línguas (CEFR, no acrônimo em inglês), classificam-se entre os níveis A1-C2. Este é um corpus de aprendizes de inglês como língua estrangeira e os dados são coletados através de um encontro entre um entrevistador e o participante para uma conversa informal que envolve três atividades comunicativas distintas. Na primeira parte, os aprendizes têm a oportunidade de brevemente apresentarem-se, enquanto a segunda parte decorre através de uma conversa informal sobre a vida em geral dos aprendizes (e.g. interesses e hobbies). Por fim, a terceira e última etapa é baseada em uma série de imagens provocantes, das quais o aprendiz deve selecionar uma e

descrevê-la, também respondendo perguntas sobre as imagens. Tabela 1, abaixo, resume as características de cada subcorpus utilizado neste estudo.

TABELA 1 – Características dos subcorpora provenientes do SCoPE² e do BraSEL Corpus

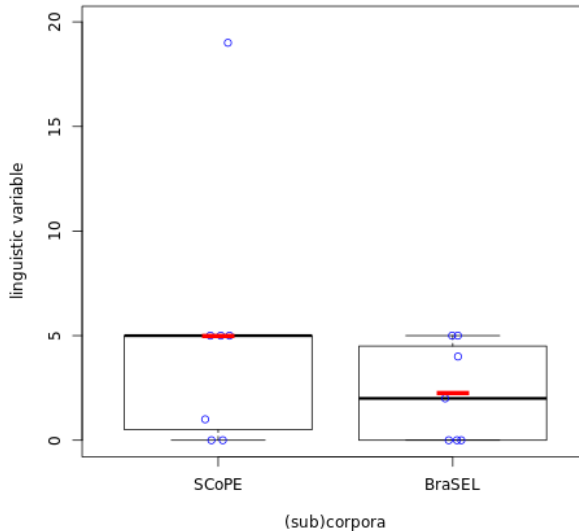
	SCoPE ²	BraSEL
Número de participantes	7	7
Gênero	5 mulheres, 2 homens	5 mulheres, 2 homens
Nacionalidade	Brasileiros	Brasileiros
Contexto de L2	Universitários na Irlanda	Universitários no Brasil
Nível de proficiência	C1-C2 CEFR; SUEs (Prodromou, 2008)	B2 CEFR
Tipo de dados	Gravações de áudio de bate-papos informais	Gravações de áudio de tarefas comunicativas informais
Tipo de interação	Um a um (pesquisador e participantes)	Um a um (entrevistador e participantes)
Número de tokens	43,596	20,890

A metodologia de análise aplicada no presente estudo refere-se a um dos modelos de estratégias investigativas da abordagem de PC forma-função: a filtragem (cf. Seção 4.1.1). O método de análise também se beneficia do modelo de leitura integrada proposto por Rühlemann e Aijmer (2015) e descrito na Seção 4. Assim sendo, primeiramente o número total de ocorrências do MP *kind of* em cada subcorpus foi contabilizado e normalizado a fim de comparação quantitativa. Apenas as ocorrências referentes aos participantes foram consideradas, uma vez que as participações do pesquisador e do entrevistador foram excluídas. As ocorrências foram, então, manualmente verificadas e filtradas para que casos não pragmáticos de *kind of* fossem também excluídos. Na etapa qualitativa, o texto e o contexto foram analisados a fim de informar e auxiliar na determinação de funções pragmáticas desempenhadas pelo MP no discurso oral de L2 dos brasileiros universitários participantes de ambos os corpora. Finalmente, cada caso pragmático de *kind of* foi alocado em um dos três domínios funcionais do MP (atitudinal, interpessoal e textual) a motivo de comparação de preferência funcional demonstrada por cada grupo de brasileiros ao fazerem uso desta forma linguística.

5.2 Resultados

Apesar de *kind of* e *sort of* performarem as mesmas funções pragmáticas como MPs, apenas o primeiro é utilizado pelos brasileiros nas amostras analisadas. Considerando que *kind of* é mais comumente relacionado com o inglês americano, e *sort of* com o inglês britânico, a ausência do segundo pode ser justificada pelo fato de brasileiros terem mais contato com o inglês americano, seja através de instrução formal em sistemas de ensino de língua inglesa, ou através de mídias de entretenimento. Quando o número de ocorrências é quantitativamente comparado entre SCoPE² e BraSEL, *kind of* ocorre 16 vezes no BraSEL (frequência normalizada de 11.6 por cada 10,000 palavras) e 35 vezes no SCoPE² (13 por cada 10,000 palavras). Contudo, como apresentado na Figura 2 abaixo, a dispersão do MP em ambos os subcorpora não é nivelada. No SCoPE², dois participantes não usam o MP de maneira alguma, enquanto um participante sozinho faz uso do MP 19 vezes (cada participante é identificado por um círculo na figura, ou *boxplot*, abaixo). No BraSEL, por sua vez, três participantes não fazem uso do MP, enquanto outros dois fazem uso do MP 5 vezes, e outro o utiliza 4 vezes, não apresentando nenhum caso fora da curva (ou *outlier*; BREZINA, 2018).

FIGURA 2 – Dispersão de *kind of* nos subcorpora SCoPE² e BraSEL



Fonte: Boxplot gerado pela plataforma Lancaster Stats Tools online (BREZINA, 2018)

Estes resultados demonstram que, embora a frequência do uso de *kind of* entre os dois subcorpora analisados seja similar em sua superfície, uma investigação mais aprofundada de dispersão do item linguístico em ambos os subcorpora revela que nem todos os participantes fazem uso deste MP, e que um *outlier* em particular no SCoPE² apresenta uma possível maior dependência de *kind of* em seu discurso oral de inglês L2. Na próxima subseção, apresentamos resultados qualitativos a fim de descrever e exemplificar as funções performadas por *kind of* nos subcorpora analisados, bem como sugerir uma possível explicação para o caso fora da curva apresentado no SCoPE².

5.2.1 Resultados qualitativos

Como observado acima, o MP *kind of* atua nos três domínios funcionais da classe: o atitudinal (de posicionamento), o interpessoal (interativo) e o textual (organização do discurso). Contudo, *kind of* foi também identificado como atuando pragmaticamente em fórmulas padronizadas que constituem MPs de linguagem vaga (e.g. marcadores de categorias vagas como *and this kind of things, or this kind of stuff, etc.*), e também em casos onde o falante usa *kind of* coocorrendo com *things* para marcar vagueza no discurso. Estes dois casos, que não se classificam como MPs *kind of* independentes, são aqui classificados como *Outros*. Exemplos de todas as funções citadas são apresentados a seguir. Contudo, Tabelas 2 e 3, abaixo, primeiramente expõem a distribuição funcional de *kind of* no SCoPE² e no BraSEL, respectivamente.

TABELA 2 – Distribuição funcional de *kind of* no SCoPE²

	Atitudinal	Interpessoal	Textual	Outros	TOTAL
SUE_1	1	2	1	1	5
SUE_2	0	0	0	0	0
SUE_3	8	0	1	10	19
SUE_4	1	0	0	0	1
SUE_5	5	0	0	0	5
SUE_6	0	0	0	0	0
SUE_7	2	3	0	0	5
TOTAL	17	5	2	11	35

TABELA 3 – Distribuição funcional de *kind of* no BraSEL

	Atitudinal	Interpessoal	Textual	Outros	TOTAL
Aprendiz_1	1	0	1	0	2
Aprendiz_2	0	0	0	0	0
Aprendiz_3	0	0	0	0	0
Aprendiz_4	1	3	1	0	5
Aprendiz_5	2	3	0	0	5
Aprendiz_6	4	0	0	0	4
Aprendiz_7	0	0	0	0	0
TOTAL	8	6	2	0	16

Como visto nas Tabelas 2 e 3 acima, todas as funções pragmáticas de *kind of* são encontradas em ambos os subcorpora, apesar de não utilizadas na mesma frequência por todos os brasileiros participantes ou, às vezes, sequer usadas por alguns, com exceção apenas de *kind of* marcando linguagem vaga na estrutura interna da oração ou constituindo MPs de categorias vagas no BraSEL.

No domínio funcional atitudinal, o falante marca o material linguístico que sucede *kind of* como impreciso e inexato, e sinaliza que tal material deve ser ajustado de maneira a aproximar-se o máximo possível do sentido intencionado. Desta forma, o MP pode ser usado para (a) ajustar uma diferença entre um pensamento e uma representação linguística; (b) marcar palavras técnicas, jargões, uso metafórico da linguagem e coloquialismos do falante; (c) como um aproximador de numerais; e (d) para compensar a ausência de vocabulário no repertório linguístico de um falante (AIJMER, 2002). Nos subcorpora analisados, a função mais usada do *kind of* atitudinal é (a), seguida por (d); a função (b) não está presente nos subcorpora, enquanto a função (c) é usada apenas uma vez. É importante ressaltar que ambas as funções (a) e (d) caminham lado a lado e, muitas vezes, há uma linha tênue entre *kind of* marcando a necessidade de ajuste entre um pensamento e sua representação linguística e o mesmo marcando explicitamente a ausência de vocabulário no repertório linguístico de um falante, especialmente em contextos de L2. Considere os exemplos abaixo extraídos do SCoPE² e do BraSEL, que representam as funções (a) e (d). Note que, como destacado nas Tabelas 2 e 3 acima, participantes SUE fazem parte do SCoPE², enquanto participantes aprendizes fazem parte do BraSEL:

- (1) SUE_3: *I went to Whistler. It's **kind of** Campos do Jordão in Brazil.*
- (2) Aprendiz_6: *Actually, I have started one book at vacation <nv> laugh </nv> but when when, em, our classes returned I **kind of** abandoned it.*
- (3) SUE_3: *... students who live like Sligo <\$E> pause </\$E> they receive two amounts <\$E> pause </\$E> one for food <\$E> pause </\$E> three hundred+*
 Pesquisador: *Mhmm.*
 SUE_3: *+in a card that is **kind of** uhm vale-alimentação+*
 Pesquisador: *Okay.*
 SUE_3: *+and another one is the money to use with things, four hundred.*
- (4) Aprendiz_5: *... I thought it was **kind of** a a lash, I think that's the name.*
 Entrevistador: *A leash.*
 Aprendiz_5: *A leash yes now yes uh-huh.*

Exemplos (1) e (2), acima, são casos de marcação de necessidade de ajuste lexical pelo MP *kind of*. No trecho (1), SUE_3 faz uma referência a uma cidade brasileira conhecida por ambos os participantes da conversa, e convida o seu interlocutor a inferir características reconhecidas de Campos do Jordão para a conceptualização mental da cidade de Whistler, no Canadá. Em outras palavras, *kind of* não só marca Campos do Jordão como uma representação linguística aproximada do que SUE_3 tem em mente, como também ajuda na interpretação de sentido da informação oferecida. Em (2), similarmente, Aprendiz_6 sinaliza que o verbo *abandon* pode não ser exatamente a melhor opção para o sentido intencionado, convidando, assim, o interlocutor a participar de um processo de inferência para a interpretação do sentido proposto.

Em contrapartida, exemplos (3) e (4) demonstram *kind of* marcando uma carência de vocabulário. Em (3), SUE_3 mais uma vez marca a troca de idiomas com *kind of*, mas, desta vez, claramente demonstrando que o termo em inglês para *vale-alimentação* não faz parte de seu repertório linguístico. Similarmente, Aprendiz_5 demonstra

dificuldade com a palavra *leash* e marca a necessidade de aproximação interpretativa por parte de seu interlocutor com o MP *kind of*.

No que diz respeito ao domínio funcional interpessoal, *kind of* também é utilizado para sinalizar uma discrepância entre um pensamento e uma representação linguística. Contudo, neste domínio funcional, o MP atua como um mitigador, suavizando a força ilocutória do material linguístico que o sucede. Na maioria dos casos analisados neste estudo, *kind of* interpessoal é usado para minimizar o impacto da escolha de palavras que são usadas, tendo em consideração a maneira como o interlocutor pode interpretá-las. Exemplo (5), abaixo, demonstra tal função interpessoal de *kind of*, onde Aprendiz_4 fala sobre a forma com a qual moradores de BH encaram um determinado bairro como um lugar perigoso. Aprendiz_4 usa a expressão *common sense*, mas a suaviza por não saber se a opinião é compartilhada pelo entrevistador.

- (5) Aprendiz_4: ... *how people talk about wh= what what stuff happens in BH. Everybody just mention oh the north area it's dangerous, don't don't go there. It's **kind of** common sense.*

Por sua vez, no domínio funcional textual, *kind of* auxilia o falante com a construção do discurso, no sentido de autocorreção por parte do falante que sinaliza que há uma discrepância entre a estrutura ou vocábulo usado e o que este tem em mente. Nesta função, *kind of* também é interativo, uma vez que sinaliza que o falante deseja manter o turno e completar sua fala. Apenas quatro ocorrências desta função foram encontradas nos subcorpora analisados, e o trecho (6), abaixo, exemplifica esta função:

- (6) Aprendiz_4: *Because it, like, when you live in BH you, the society **kinda** mm not the society but how people talk about wh= what what stuff happens in BH.*

No trecho (6), que precede o trecho (5), Aprendiz_4 reestrutura a sua fala, e o faz com o auxílio de diversos marcadores de hesitação e autocorreção, como *like*, *mm*, e palavras cortadas e repetidas (*wh= what*), incluindo a versão mais informal de *kind of*, *kinda*.

Finalmente, onze casos de *kind of* categorizados como *Outros* foram identificados, todos no SCoPE², e dos quais dez são utilizados por

um SUE apenas. Trechos (7) e (8) exemplificam duas subcategorias que compõem *Outros*:

- (7) SUE_3: ... *actually, not now but before during the the year* <\$E> pause </\$E> *uhm every Tuesday a woman came to see if everything was clean **this kind of things**.*
- (8) SUE_3: *And because we know we have **this kind of things** in Brazil and I travel a lot with my parents.*

No exemplo (7), *kind of* é componente de uma fórmula fixa que compõe uma subclasse de MPs, denominada marcadores de categorias vagas. A fórmula ocorre sempre como uma etiqueta de acréscimo ao final de uma contribuição linguística e marca aproximação e necessidade de ajuste por parte do interlocutor de uma categoria inteira. Em (7), SUE_3 fala sobre ter uma funcionária que, semanalmente, vai à hospedagem estudantil para fazer a limpeza. Desta maneira, *this kind of things* é retroativo e convida o interlocutor a inferir os tipos de coisas que se classificariam em uma categoria vaga de ações que envolvem limpeza. Em outras palavras, SUE_3 não precisa listar uma extensiva lista de atividades, uma vez que tal lista pode ser inferida pelo interlocutor através de uma interpretação de *this kind of things* no contexto em que este é usado. Diferentemente, exemplo (8) demonstra a mesma forma linguística, mas usada de maneira distinta. Em (8), *this kind of things* não é um acréscimo extra, mas parte interna da oração. Juntos, *kind of e things* atuam para implicar um sentido de linguagem vaga, e convidam o interlocutor a interpretar tal forma com referência ao contexto previamente explicitado. No caso, SUE_3 fala sobre o fato de haver no Brasil muitos destinos de viagem com paisagens, justificando seu interesse por viajar para lugares urbanos que destacam a arquitetura na Europa. Ao usar *this kind of things* dentro do contexto apresentado, o interlocutor é capaz de inferir o sentido intencionado pelo uso de tal linguagem vaga. Interessantemente, SUE_3 é a participante do subcorpus SCoPE² que apresentou menor performance linguística, apesar de seu certificado de proficiência C1. E é exatamente esta participante que se mostra como um *outlier* (fora da curva da média) na contagem de frequência de *kind of*. Das 19 ocorrências verificadas no discurso da SUE_3, 18 foram casos de uso do MP para indicar discrepância de vocabulário (atitudinal) ou para evitar a provisão explícita de vocabulário específico (*Outros*), o que

indica que SUE_3 pode manter uma dependência, ou preferência, por tais funções de *kind of* para compensar possíveis insuficiências linguísticas.

Este estudo confirma o valor pragmático do MP *kind of* em conversação e aponta para o fato de que, mesmo em contextos linguísticos distintos, os dois grupos de brasileiros fazem uso das funções de *kind of* em seus três domínios funcionais. Contudo, nem todos os participantes de ambos os subcorpora utilizam *kind of* pragmaticamente. Seria, contudo, equivocado concluir que os participantes que não fazem uso do MP *kind of* sofram de uma deficiência pragmática em seus processos comunicativos, uma vez que estes participantes podem lançar mão de outras formas que performam as mesmas funções pragmáticas apresentadas neste estudo. De fato, em outro estudo de caso preliminar, Santos (2019) nota que os participantes sob investigação, também de uma amostra do SCoPE², apoiam-se significativamente no MP *like* e em suas funções textuais de reestruturação do discurso.

6 Considerações finais

As considerações finais aqui apresentadas se remetem ao título deste trabalho. Propusemos, como objetivo principal, expor o estado da arte da Pragmática de Corpus, ponto de contato que se originou a partir de duas áreas que, apesar de lidarem com investigações sobre a linguagem em uso, possuem metodologias distintas – a Linguística de Corpus e a Pragmática. Ao fazer isso, inicialmente, contextualizamos a Linguística de Corpus e a Pragmática. Em seguida, apresentamos o percurso da Pragmática de Corpus a partir de seu contexto histórico, discursando sobre sua metodologia de dupla direcionalidade, além de suas limitações e desenvolvimentos. Nesta mesma senda, introduzimos duas abordagens profícuas da nova área, forma-função e função-forma. Finalizamos, então, com um estudo de caso ilustrativo sobre o marcador pragmático *kind of* em dois corpora orais, aplicando o método de filtragem pela abordagem forma-função.

É evidente que diferentes áreas da linguística buscam incorporar, com o passar dos anos, exemplos autênticos da linguagem às suas análises. Vimos que a Linguística de Corpus, por meio de seu desenvolvimento nas últimas décadas e rigor metodológico, dispõe de grande potencial para se comunicar com outros campos do conhecimento, considerando o que estes tradicionalmente já estabeleceram, introduzindo formas

inovadoras e eficazes para a manipulação de dados linguísticos reais em grandes ou pequenas quantidades. É assim que reiteramos o poder analítico da Pragmática de Corpus para a investigação e avanço dos estudos da linguagem, na qual duas disciplinas são beneficiadas em dependência mútua. Ressaltamos que, devido ao seu caráter jovem, a disciplina ainda está em desenvolvimento, principalmente no que diz respeito ao seu modelo metodológico. Sendo assim, novos estudos continuam a ser conduzidos, juntamente com o aperfeiçoamento de técnicas de análise e de anotação pragmática, e, por conseguinte, cada vez mais facilmente tomando a função como ponto de partida. Dessa forma, o que apresentamos aqui é base para o desenvolvimento futuro desta área emergente e promissora.

Agradecimentos

Agradecemos à Prof. Dra. Anne O’Keeffe, do Departamento de Língua Inglesa e Literatura da Mary Immaculate College, Irlanda, pelo contínuo apoio e pela disponibilidade e confiança em nos providenciar a literatura necessária para a escrita deste artigo.

Também agradecemos aos pareceristas por suas avaliações, sugestões e comentários que muito contribuíram para a versão final deste trabalho.

Contribuição dos autores

O artigo foi desenvolvido conjuntamente pelos autores, uma vez que ambos estão inseridos no contexto atual do desenvolvimento da Pragmática de Corpus por parte da escola europeia. Giovani Santos formulou o desenho geral do artigo, e Mateus Miranda auxiliou no afunilamento da proposta. As seções foram divididas entre os autores para serem redigidas, mas se beneficiaram da contribuição de ambos os autores. A análise dos dados também foi feita pelos dois autores de forma conjunta.

Referências

ÄDEL, A.; REPPEN, R. The Challenges of Different Settings: An Overview. In: ÄDEL, A.; REPPEN, R. (org.). *Corpora and Discourse: The Challenges of Different Settings*. Amsterdam; Philadelphia: John Benjamins, 2008. p. 1-6. DOI: <https://doi.org/10.1075/scl.31>

AIJMER, K. *English Discourse Particles: Evidence from a Corpus*. Amsterdam/Philadelphia: John Benjamins, 2002. DOI: <https://doi.org/10.1075/scl.10>

AIJMER, K. Corpus Pragmatics: from Form to Function. In: JUCKER, A. H.; SCHNEIDER, K. P.; BUBLITZ, W. (org.). *Methods in Pragmatics*. Berlin; Boston: De Gruyter Mouton, 2018. p. 555-586. DOI: <https://doi.org/10.1515/9783110424928-022>

AIJMER, K.; RÜHLEMANN, C. (org.). *Corpus Pragmatics: A Handbook*. Cambridge: Cambridge University Press, 2015. DOI: <https://doi.org/10.1017/CBO9781139057493>

BEEBE, L.; TAKAHASHI, T.; ULISS-WELTZ, R. Pragmatic Transfer in ESL Refusals. In: SCARCELLA, R. C.; ANDERSON, E.; KRASHEN, S. D. (org.). *Developing Communicative Competence in a Second Language*. New York: Newbury, 1990. p. 55-73.

BERBER SARDINHA, T. Beginning Portuguese Corpus Linguistics: Exploring a Corpus to Teach Portuguese as a Foreign Language. *DELTA*, São Paulo, v. 15, n. 2, p. 289-299, 1999. DOI: <https://doi.org/10.1590/S0102-44501999000200003>

BERBER SARDINHA, T. Linguística de corpus: histórico e problemática. *DELTA*, São Paulo, v. 16, n. 2, p. 323-367, 2000. DOI: <https://doi.org/10.1590/S0102-44502000000200005>

BIBER, D.; CONRAD, S.; REPPEN, R. *Corpus Linguistics: Investigating Language Structure in Use*. Cambridge: Cambridge University Press, 1998. DOI: <https://doi.org/10.1017/CBO9780511804489>

BREZINA, V. *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge: Cambridge University Press, 2018. DOI: <https://doi.org/10.1017/9781316410899>

BRINTON, L. *Pragmatic Markers in English: Grammaticalization and Discourse Functions*. Berlin: Mouton de Gruyter, 1996. DOI: <https://doi.org/10.1515/9783110907582>

BUYSSE, L. Discourse Markers in the English of Flemish University Students. In: WITCZAK-PLISIECKA, I. (org.). *Pragmatic Perspectives on Language and Linguistics: Speech Actions in Theory and Applied Studies*. Newcastle-upon-Tyne: Cambridge Scholars, 2010. p. 461-484.

- CAINES, A.; McCARTHY, M.; O'KEEFFE, A. Spoken Language Corpora and Pedagogical Applications. In: FARR, F.; MURRAY, L. (org.). *The Routledge Handbook of Language Learning and Technology*. Abingdon: Routledge, 2016. p. 348-361.
- CARTER, R.; McCARTHY, M. *Cambridge Grammar of English: A Comprehensive Guide*. Cambridge: Cambridge University Press, 2006.
- CLANCY, B.; McCARTHY, M. Co-Constructed Turn-Taking. In: AIJMER, K.; RÜHLEMANN, C. (org.). *Corpus Pragmatics: A Handbook*. Cambridge: Cambridge University Press, 2015. p. 430-453. DOI: <https://doi.org/10.1017/CBO9781139057493.023>
- CLANCY, B.; O'KEEFFE, A. Pragmatics. In: BIBER, D.; REPPEN, R. (org.). *The Cambridge Handbook of English Corpus Linguistics*. Cambridge: Cambridge University Press, 2015. p. 235-251. DOI: <https://doi.org/10.1017/CBO9781139764377.014>
- CRESTI, E.; MONEGLIA, M. (org.). *C-ORAL-ROM: Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam; Philadelphia: John Benjamins Publishing Company, 2005. DOI: <https://doi.org/10.1075/scl.15>
- CULPEPER, J. The Metalanguage of Impoliteness: Explorations in the Oxford English Corpus. In: BAKER, P. (org.). *Contemporary Corpus Linguistics*. London: Continuum, 2009. p. 64-86.
- CULPEPER, J.; HAUGH, M. *Pragmatics and the English Language*. Palgrave Macmillan: Basingstoke, 2014. DOI: <https://doi.org/10.1007/978-1-137-39391-3>
- FARR, F.; BRÓNA, M.; O'KEEFFE, A. The Limerick Corpus of Irish English: Design, Description and Application. *Teanga*, Dublin, v. 21, p. 5-29, 2004. DOI: <https://doi.org/10.35903/teanga.v21i0.172>
- FIRTH, J. R. *Papers in Linguistics 1934-1951*. London: Oxford University Press, 1957.
- GILQUIN, G. Hesitation Markers among EFL Learners: Pragmatic Deficiency or Difference? In: ROMERO-TRILLO, J. (org.). *Corpus Linguistics and Pragmatics: A Mutualistic Entente*. Berlin: De Gruyter Mouton, 2008. p. 119-149.

GOMES DE MATOS, F. Pós-graduação em lingüística no Brasil: orientações curriculares e output (dissertações). *Boletim da ABRALIN*, Recife, v. 3, p. 81-87, 1982.

GRICE, H. P. Logic and Conversation. In: COLE, P.; MORGAN, J. J. P. (org.). *Syntax and Semantics 3: Speech Acts*. New York: Academic Press, 1975. p. 41-58. DOI: https://doi.org/10.1163/9789004368811_003

HAUGH, M. Corpus-based metapragmatics, In: JUCKER, A. H.; SCHNEIDER, K. P.; BUBLITZ, W. (org.). *Methods in Pragmatics*. Berlin; Boston: De Gruyter Mouton, 2018. p. 619-644. DOI: <https://doi.org/10.1515/9783110424928-023>

JUCKER, A. H. Corpus Pragmatics. In: ÖSTMAN, J. O.; VERSCHUEREN, J. (org.). *Handbook of Pragmatics*. Amsterdam: John Benjamins, 2013. p. 1-18. DOI: <https://doi.org/10.1075/hop.17.cor3>

JUCKER, A.; TAAVITSAINEN, I.; SCHNEIDER, G. Semantic Corpus Trawling: Expressions of Courtesy and Politeness in the Helsinki Corpus. In: SUHR, C.; TAAVITSAINEN, I. (org.). *Developing Corpus Methodology for Historical Pragmatics*. Studies in Variation, Contacts and Change in English. Helsinki: Helsingin Yliopisto, 2012. v. 11. [s.p.]. Disponível em: http://www.helsinki.fi/varieng/series/volumes/11/jucker_tavitsainen_schneider/. Acesso em: 20 set. 2020.

JUCKER, A. H.; SCHNEIDER, K. P.; BUBLITZ, W. (org.). *Methods in Pragmatics*. Berlin; Boston: De Gruyter Mouton, 2018. DOI: <https://doi.org/10.1515/9783110424928>

JUCKER, A. H.; SCHREIER, D.; HUNDT, M. (org.). *Corpora: Pragmatics and Discourse*. Amsterdam; New York: Rodopi, 2009. DOI: <https://doi.org/10.1163/9789042029101>

JUCKER, A. H.; TAAVITSAINEN, I. Diachronic Corpus Pragmatics: Intersections and Interactions. In: TAAVITSAINEN, I.; JUCKER, A. H.; TUOMINEN, J. (org.). *Diachronic Corpus Pragmatics*. Amsterdam; Philadelphia: John Benjamins, 2014. p. 3-26. DOI: <https://doi.org/10.1075/pbns.243.03juc>

KALLEN, J.; KIRK, J. *SPICE-Ireland: A User's Guide*. Belfast: Cló Ollscoil na Banríona, 2012.

KIRK, J. M. Kind of and Sort of: Pragmatic Discourse Markers in the SPICE-Ireland Corpus. In: AMADOR-MORENO, C. P.; McCAFFERTY, K.; VAUGHAN, E. (org.). *Pragmatic Markers in Irish English*. Amsterdam; Philadelphia: John Benjamins, 2015. p. 88-113. DOI: <https://doi.org/10.1075/pbns.258.04kir>

KOESTER, A. Building Small Specialised Corpora. In: O'KEEFFE, A.; McCARTHY (org.). *The Routledge Handbook of Corpus Linguistics*. London: Routledge, 2010. p. 66-79. DOI: <https://doi.org/10.4324/9780203856949-6>

LEVINSON, S. *Pragmatics*. Cambridge: Cambridge University Press, 1983.

LOVE, R. *Overcoming Challenges in Corpus Construction: The Spoken British National Corpus 2014*. New York: Routledge, 2020. DOI: <https://doi.org/10.4324/9780429429811>

MAYNARD, C.; LEICHER, S. Pragmatic Annotation of an Academic Spoken Corpus for Pedagogical Purposes. In: FITZPATRICK, E. (org.). *Corpus Linguistics beyond the Word: Corpus Research from Phrase to Discourse*. Amsterdam: Rodopi, 2007. p. 107-116. DOI: https://doi.org/10.1163/9789401203845_008

McALLISTER, P. G. Speech Acts: A Synchronic Perspective. In: AIJMER, K.; RÜHLEMANN, C. (org.). *Corpus Pragmatics: A Handbook*. Cambridge: Cambridge University Press, 2015. p. 29-51. DOI: <https://doi.org/10.1017/CBO9781139057493.003>

McCARTHY, M.; O'KEEFFE, A. Historical Perspective: What Are Corpora and How Have They Evolved? In: O'KEEFFE, A.; McCARTHY, M. (org.). *The Routledge Handbook of Corpus Linguistics*. Abingdon: Routledge, 2010. p. 3-13. DOI: <https://doi.org/10.4324/9780203856949-1>

McCARTHY, M.; O'KEEFFE, A. Spoken Grammar. In: CELCE-MURCIA, M.; BRINTON, D. M.; SNOW, M. A. (org.). *Teaching English as a Second or Foreign Language*. 4. ed. Boston: National Geographic Learning, 2014. p. 271-287.

McENERY, T.; HARDIE, A. *Corpus Linguistics: Method, Theory and Practice*. Cambridge: Cambridge University Press, 2012. DOI: <https://doi.org/10.1017/CBO9780511981395>

MIRANDA, M. 'Dreams Seem Kind of Utopic': Vague Category Markers in a Learner Corpus. *Revista Intercâmbio*, São Paulo, v. 44, p. 84-107, 2020.

MORRIS, C. W. *Foundations of the Theory of Signs*. Chicago: University of Chicago Press, 1938.

O'KEEFFE, A. Corpus-Based Function-to-Form Approaches. In: JUCKER, A. H.; SCHNEIDER, K. P.; BUBLITZ, W. (org.). *Methods in Pragmatics*. Berlin; Boston: De Gruyter Mouton, 2018. p. 587-618. DOI: <https://doi.org/10.1515/9783110424928-023>

O'KEEFFE, A.; CLANCY, B.; ADOLPHS, S. *Introducing Pragmatics in Use*. London: Routledge, 2011.

O'KEEFFE, A.; CLANCY, B.; ADOLPHS, S. *Introducing Pragmatics in Use*. 2. ed. rev. e aum. Abingdon; New York: Routledge, 2020. DOI: <https://doi.org/10.4324/9780429342950>

O'KEEFFE, A.; McCARTHY, M.; CARTER, R. *From Corpus to Classroom: Language Use and Language Teaching*. Cambridge: Cambridge University Press, 2007. DOI: <https://doi.org/10.1017/CBO9780511497650>

O'KEEFFE, A.; McCARTHY, M. (org.). *The Routledge Handbook of Corpus Linguistics*. Abingdon: Routledge, 2010.

PRODROMOU, L. *English as a Lingua Franca: A Corpus-Based Analysis*. London: Continuum, 2008. DOI: <https://doi.org/10.1093/elt/ccn064>

RAJAGOPALAN, K. Os caminhos da pragmática no Brasil. *DELTA*, São Paulo, v. 15, número especial, p. 323-338, 1999. DOI: <https://doi.org/10.1590/S0102-44501999000300013>

RAJAGOPALAN, K. Repensar o papel da linguística aplicada. In: LOPES, L. P. M. (org.). *Por uma linguística aplicada indisciplinar*. São Paulo: Parábola, 2006. p. 149-168.

RAJAGOPALAN, K. Pragmática. In: MOLLICA, M. C.; FERRAREZI, C. J. (org.). *Sociolinguística e sociolinguísticas*. São Paulo: Contexto, 2016. p. 197-204.

RASO, T. O corpus C-ORAL-BRASIL. In: RASO, T.; MELLO, H. (orgs.). *C-ORAL-BRASIL I: Corpus de referência do português brasileiro falado informal*. Belo Horizonte: Editora UFMG, 2012. p. 55-89.

RASO, T. Aspectos sociais e pragmáticos da linguística de corpora. In: MOLLICA, M. C.; FERRAREZI, C. J. (org.). *Sociolinguística e sociolinguísticas*. São Paulo: Contexto, 2016. p. 205-216.

RASO, T.; MELLO, H. (org.). *C-ORAL-BRASIL I: corpus de referência do português brasileiro falado informal*. Belo Horizonte: Editora UFMG, 2012.

ROMERO-TRILLO, J. (org.). *Corpus Linguistics and Pragmatics: A Mutualistic Entente*. Berlin: Walter de Gruyter, 2008a. DOI: <https://doi.org/10.1515/9783110199024>

ROMERO-TRILLO, J. Introduction. In: _____. (org.). *Corpus Linguistics and Pragmatics: A Mutualistic Entente*. Berlin: De Gruyter Mouton, 2008b. p. 1-10. DOI: <https://doi.org/10.1515/9783110199024>

RÜHLEMANN, C. *Corpus Linguistics for Pragmatics: A Guide for Research*. Abingdon; New York: Routledge, 2019. DOI: <https://doi.org/10.4324/9780429451072>

RÜHLEMANN, C.; AIJMER, K. Corpus Pragmatics: laying the foundations. In: AIJMER, K.; RÜHLEMANN, C. (org.). *Corpus Pragmatics: a handbook*. Cambridge: Cambridge University Press, 2015. p. 1-26. DOI: <https://doi.org/10.1017/CBO9781139057493.001>

RÜHLEMANN, C.; CLANCY, B. Corpus Linguistics and Pragmatics. In: ILIE, C.; NORRICK, N. R. (org.). *Pragmatics and Its Interfaces*. Amsterdam; Philadelphia: John Benjamins, 2018. p. 241-266. DOI: <https://doi.org/10.4324/9780429451072>

SANTOS, G. Second Language Pragmatics: A Corpus-Based Study of the Pragmatic Marker Like. *Letrônica*, Porto Alegre, v. 12, n. 4, p. 1-16, 2019. DOI: <https://doi.org/10.15448/1984-4301.2019.4.34002>

SANTOS, G. Designing and Building SCoPE²: A Spoken Corpus of Brazilian Portuguese and L2-English. *Research in Corpus Linguistics*, Murcia, v. 8, n. 1, p. 49-64, 2020. DOI: <https://doi.org/10.32714/ricl.08.01.04>

SINCLAIR, J. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press, 1991.

SINCLAIR, J. *Reading Concordances*. London: Pearson; Longman, 2003.

SINCLAIR, J. *Trust the Text: Language, Corpus and Discourse*. Abingdon: Routledge, 2004.

SINCLAIR, J. Corpus and Text: Basic Principles. In: WYNNE, M. (org.) *Developing Linguistic Corpora: A Guide to Good Practice*. Oxford: Oxbow Books, 2005. p. 1-16.

SINCLAIR, J.; FOX, G.; BULLEN, S.; MANNING, E. *Collins-COBUILD English Language Dictionary*. London: Collins, 1987.

SPERBER, D.; WILSON, D. *Relevance: Communication and Cognition*. Oxford: Blackwell, 1995.

TAAVITSAINEN, I.; JUCKER, A. H.; TUOMINEN, J. (org.). *Diachronic Corpus Pragmatics*. Amsterdam; Philadelphia: John Benjamins, 2014. DOI: <https://doi.org/10.1075/pbns.243>

TAGNIN, S. Corpora on-line. In: VIANA, V.; TAGNIN, S. E. O. (org.). *Corpora no ensino de línguas estrangeiras*. São Paulo: Hub Editorial, 2010. p. 354-361.

VAUGHAN, E. “Got a Date or Something?”: An Analysis of the Role of Humour and Laughter in the Workplace Meetings of English Language Teachers. In: ÄDEL, A.; REPPEN, R. (org.). *Corpora and Discourse: The Challenges of Different Settings*. Amsterdam; Philadelphia: John Benjamins, 2008. p. 95-115. DOI: <https://doi.org/10.1075/scl.31.07vau>

VAUGHAN, E.; CLANCY, B. Small Corpora and Pragmatics. In: ROMERO-TRILLO, J. (org.). *Yearbook of Corpus Linguistics and Pragmatics 2013: New Domains and Methodologies*. Dordrecht: Springer, 2013. p. 53-73. DOI: https://doi.org/10.1007/978-94-007-6250-3_4

WEISSER, M. Speech Act Annotation. In: AIJMER, K.; RÜHLEMANN, C (org.). *Corpus Pragmatics: A Handbook*. Cambridge: Cambridge University Press, 2015. p. 84-113. DOI: <https://doi.org/10.1017/CBO9781139057493.005>

WEISSER, M. The DART Annotation Scheme: Form, Applicability & Application. *Studia Neophilologica*, [S.l.], v. 91, p. 131-153, 2019. DOI: <https://doi.org/10.1080/00393274.2019.1616218>

WOLFSON, N. Compliments in Cross-Cultural Perspective. *TESOL Quarterly*, [S.l.], v. 15, p. 117-124, 1981. DOI: <https://doi.org/10.2307/3586403>

YULE, G. *Pragmatics*. Oxford: Oxford University Press, 1996.