



Cognição e variação linguística de gêneros/registros jornalísticos: um estudo baseado em *corpus*

Linguistic variation and cognition of press genres/register: a corpus-based study

Carlos Henrique KAUFFMANN*

RESUMO: As páginas de um jornal congregam diversos gêneros/registros linguísticos especializados, seja qual for o meio de acesso para a sua leitura. Nem sempre, porém, percebe-se distinção clara entre gêneros/registros jornalísticos, mesmo entre os produtores desses textos, o que leva a especular sobre a conformação e estabilidade linguística de determinados gêneros/registros. O presente estudo investiga essa questão, ao analisar um *corpus* formado por textos de duas edições da "Folha de S.Paulo", classificados em termos de gêneros/registros por especialistas. Os resultados foram tabulados por grau de concordância entre classificadores. A maioria das classificações recaiu sobre as categorias reportagem e notícia. Em seguida, escores que refletem o consenso na determinação do gênero/registo foram mapeados segundo as dimensões de variação do texto jornalístico de Kauffmann (2005), verificando o quão separados linguisticamente estão os gêneros/registros menos consensuais.

ABSTRACT: The newspaper is a repository of specialized genres/register that are present in a daily basis on its pages, by any medium of access. However, as the clear distinction among press genres/register is sometimes blurry, even by the writers of those texts, some questions about the linguistic stability of some genres/register could be made. This study will analyze language variation from a corpus of the "Folha de S.Paulo" broadsheet newspaper, classified in terms of press genres/register by four expert classifiers. This task generated a database of texts classified by journalistic genres/register, in which the majority of the texts were labeled as reportage and/or news report. Scores that reflect the degree of consensus around press registers were then plotted along the dimensions of variation of the journalistic text identified in Kauffmann (2005). This allowed to verify to what extent the average partial agreement scores of the main news reports are linked to their respective agreement categories.

* Doutor em linguística aplicada e estudos da linguagem pela Pontifícia Universidade Católica de São Paulo (PUCSP). ORCID: <https://orcid.org/0000-0001-8792-7375>. chkauffmann@corpuslg.org

PALAVRAS-CHAVE: Linguística de *Corpus*. Análise multidimensional. Gênero jornalístico. Registro.

KEYWORDS: Corpus Linguistics. Multi-dimensional analysis. Press genre. Register.

1 Introdução

A língua é sistematicamente utilizada por meio de formas discursivas que coabitam os espaços de produção linguística de qualquer sociedade, a qualquer tempo. Modernamente, um desses espaços produtivos provém da veiculação de material noticioso. Seja em formato impresso ou digital, jornais, blogs e sites informativos reúnem em suas páginas diversos gêneros, ou registros – tais como reportagem, notícia, artigo, editorial etc. – convivendo cotidianamente, cada qual possuidor de características próprias capazes de o distinguir socialmente como um grupo singular.

O registro jornalístico, visto sob um prisma mais ampliado, representa uma das referências fundamentais do uso real da língua, ao lado dos discursos falado, acadêmico e literário (BIBER *et al.*, 1999). Esta visão permitiu o desenvolvimento de estudos comparativos e investigações de natureza diacrônica de registros com base em pesquisa empírica, a partir de *corpora* especializados – isto é, bases de textos coletados criteriosamente e armazenados digitalmente, na maioria das vezes etiquetados sob diversos níveis de análise linguística, de forma automática, semiautomática ou manual (BERBER SARDINHA, 2004). Por outro lado, quando observados em maior detalhe, entram em ação processos de identificação e reconhecimento de gêneros/registros jornalísticos, capazes de diferenciar um texto de outro sob esse aspecto. Tais processos são considerados habilidades importantes na educação, formação e conscientização de estudantes, uma vez que auxiliam no aprimoramento da cidadania e estimulam uma visão crítica de mundo, em conformidade com o que é recomendado pelos Parâmetros Curriculares Nacionais (BRASIL, 1998).

Não há, porém, um entendimento pacífico sobre a atribuição de um determinado texto da imprensa a um gênero ou registro – termos aqui empregados de

forma equivalente, pois as duas definem variedades linguísticas com certos propósitos comunicativos reconhecidos socialmente, passíveis de serem submetidas a uma análise linguística empírica (BIBER; CONRAD, 2009). A classificação em termos de gênero/registo é comumente feita após a publicação do texto jornalístico e independe do tema ou tópico tratado por ele. Tampouco existe delimitação clara de gêneros/registros jornalísticos, exceto por alguns explicitamente distintos, como o editorial, que por vezes é reconhecido por uma apresentação gráfica diferenciada.

Esta pesquisa buscou identificar em textos jornalísticos quais são os gêneros/registros jornalísticos considerados mais consensuais – e os mais conflitantes –, em termos de seu reconhecimento, entre observadores especialistas. Para tal, associa os recursos metodológicos da Linguística de *Corpus* (BERBER SARDINHA, 2004), abordagem que investiga padrões de língua com a utilização de *corpora* auxiliados por computador, a uma pesquisa qualitativa, que aferiu a cognição de um grupo em relação aos gêneros/registros existentes no jornal. O grau de concordância dessa classificação, tomado em conjunto, foi analisado em termos de sua chance em relação ao acaso com uma medida estatística – o kappa de Fleiss, explicado a seguir – adequada ao desenho da pesquisa. Os textos mais e menos concordes em termos de gêneros/registros foram analisados e comparados de acordo com a Análise Multidimensional (BIBER, 1988) efetuada por Kauffmann (2005), que identificou as dimensões atuantes na variação linguística do texto da imprensa no modo escrito em língua portuguesa. Desse modo, buscou-se interpretar em que medida as diferenças entre gêneros/registros jornalísticos, influenciadas pelos graus de concordância obtidos, podem ser explicadas por sua natureza linguística.

2 Pressupostos teóricos

Diversas áreas do conhecimento têm estudado a natureza e a tipologia dos gêneros/registros jornalísticos, como a de Comunicação Social (MARQUES DE MELO;

ASSIS, 2010), *Linguística Aplicada* (MARCUSCHI, 2002; AITCHISON; LEWIS, 2003), *Análise Crítica do Discurso* (VAN DIJK, 1988; FAIRCLOUGH, 1995), bem como manuais de jornalismo de caráter utilitário e profissional (FOLHA DE S. PAULO, 2018; EDITORA ABRIL, 1990; O GLOBO, 1998; O ESTADO DE S. PAULO, 1990).

Marques de Melo e Assis (2010) reúne estudos sobre diversos gêneros/registros jornalísticos na perspectiva comunicacional, tratando-os pela denominação de "formatos" (cf. COSTA, 2010), por sua vez inscritos em categorias de "gêneros" – opinativo, informativo, interpretativo, diversional ou utilitário. Marques de Melo e Assis (2016), revisitando esse modelo, considera que a imprensa escrita tenha originado tal estrutura de classificação e ainda influencia os gêneros/registros jornalísticos da atualidade, mesmo em suportes eletrônicos ou digitais.

Na área linguística, Marcuschi (2002) pôs em perspectiva uma visão de cunho bakhtiniano para os gêneros textuais, definidos como "entidades sociodiscursivas e formas de ação social incontornáveis em qualquer situação comunicativa" (MARCUSCHI, 2002, p. 19), capazes de assumir diversidade de formas e uma dinâmica histórica. O autor inscreve os gêneros textuais em grupos maiores, os domínios discursivos, que designariam "uma esfera ou instância de produção discursiva ou de atividade humana" (MARCUSCHI, 2002, p. 22) – por exemplo, o domínio jornalístico incluiria todos os gêneros/registros tratados no presente estudo. Nesse sentido, o conceito de domínio discursivo aproxima-se da definição mais ampla de registro, no sentido empregado por Biber (BIBER; CONRAD, 2009). Já os tipos de texto expressariam formas reconhecíveis no aspecto linguístico, que seriam utilizados em proporções diversas na produção de gêneros textuais, numa "construção teórica definida pela natureza linguística de sua composição" (MARCUSCHI, 2002, p. 21). São exemplos de tipos de texto a narração, a argumentação, a exposição, a descrição e a injunção (ou instrução).

Outras perspectivas, na área de Comunicação, são propostas por Chaparro (1997) e Seixas (2009). Com base em uma análise dos papéis identitários discursivos dos textos, Seixas (2009) propõe uma classificação alternativa de duas categorias gerais, embora tenha alocado nelas, em sua maioria, gêneros/registros jornalísticos já conhecidos. Chaparro (1997), por sua vez, analisa os gêneros/registros do jornal segundo esquemas de superestruturas e macroestruturas (VAN DIJK, 1988), que resultam em uma tipologia que diferencia gêneros/registros agrupados sob esquemas narrativos ou esquemas argumentativos.

Os manuais de redação publicados na imprensa brasileira descrevem características de gêneros/registros jornalísticos olhando-os sob o viés prático da produção textual especializada. Na maioria das publicações técnicas, há a distinção entre o jargão profissional que designa o texto ("matéria", entre outros termos) e a tipologia de gêneros/registros jornalísticos estabelecidos. As diversas edições do manual do jornal "Folha de S. Paulo" (FOLHA DE S.PAULO, 1984; 1987; 1992; 2001; 2018) apresentam os gêneros/registros mais conhecidos, com maior ou menor visibilidade. Por exemplo, apesar de uma edição mais recente do manual (FOLHA DE S.PAULO, 2018) não mais destacar o gênero/registro notícia como um item de um verbete, seja dedicado aos gêneros jornalísticos (FOLHA DE S. PAULO, 2001), ou como verbete próprio (FOLHA DE S.PAULO, 1992), sua presença ainda é efetiva no capítulo voltado à prática jornalística, conforme a publicação qualifica situações como o uso de "notícia exclusiva" ou, ainda, de *fake news* ("notícia falsa").

Estudos fundamentais sobre gêneros/registros jornalísticos na Linguística Aplicada originaram-se da área de Análise Crítica do Discurso (VAN DIJK, 2010; FAIRCLOUGH, 1995; BONINI, 2012), embora outras abordagens (BELL, 1991; AITCHINSON; LEWIS, 2003) tenham também expandido a discussão.

Com um viés cognitivo, no propósito de investigar o conhecimento de gêneros/registros jornalísticos entre uma comunidade linguística que lida com o texto

de imprensa profissionalmente, Bonini (2002) apresentou trechos e textos de alguns gêneros/registros a jornalistas, para efetuar uma classificação, sem que tivessem prévio conhecimento metalinguístico sobre os gêneros/registros. O tipo de avaliação utilizada levou a uma lista aberta de gêneros/registros, e ao uso de jargão por alguns observadores. Outros trabalhos do mesmo autor investigaram as designações ou rótulos dos gêneros/registros, entre eles Bonini (2003). Por sua vez, Bonini (2009) analisou a distinção existente entre os gêneros/registros notícia e reportagem com a metodologia de Swales (1990), que envolve a análise de movimentos retóricos, concluindo que há um *continuum* fluido interligando os dois gêneros/registros.

A Análise Multidimensional (BIBER, 1988; BERBER SARDINHA, 2000) foi a metodologia escolhida para mensurar a variação linguística entre gêneros/registros efetuada pela pesquisa de classificação por observadores especialistas, grupo composto por profissionais da área jornalística com formação acadêmica de nível superior. Originalmente, a Análise Multidimensional (AMD) centrou-se no estudo de variação entre os vários gêneros/registros que compõem a língua, nos modos escrito e oral, em inglês e outras línguas (BIBER, 1995), e sempre incluiu gêneros/registros ligados ao jornalismo. Na AMD realizada para o português do Brasil (BERBER SARDINHA et al., 2014b), por exemplo, entre os 48 gêneros/registros estudados, sete estão ligados à mídia: reportagem, revista de notícias, revista de celebridades, editoriais, crônicas (modo escrito), e entrevistas publicadas e notícias de TV (modo oral). Com base no estudo foi possível concluir que os gêneros/registros de mídia são letrados, não procedurais e com foco no passado; tomados individualmente, os gêneros/registros editorial, crônica e entrevista são mais argumentativos que os demais (KAUFFMANN, 2015).

Por meio da AMD, é possível também estudar subgrupos de gêneros/registros, como os de cunho jornalístico (KAUFFMANN, 2005) ou acadêmico (BIBER, 2006), além de gêneros/registros isolados, como por exemplo filmes norte-americanos

(VEIRANO PINTO, 2013), músicas pop (BÉRTOLI-DUTRA, 2010) e reportagens de capa da revista "Time" (SOUZA, 2012). Outros estudos utilizaram corpora de textos jornalísticos para análise de variação da linguagem (BIBER *et al.*, 1999; BEDNAREK, 2006; BIBER; CONRAD, 2009, p. 116-7).

As perguntas de pesquisa que motivaram o presente trabalho poderiam ser elencadas da seguinte forma:

1. Qual é a medida de concordância geral obtida na classificação de gêneros/registros jornalísticos por meio de observadores especialistas?
2. Quais são os gêneros/registros jornalísticos considerados em maior e menor grau de concordância na classificação de gêneros/registros jornalísticos por meio de observadores especialistas?
3. Como os gêneros/registros jornalísticos e seus diversos subgrupos de concordância estão mapeados segundo as dimensões de variação do texto jornalístico de Kauffmann (2005)?

Para respondê-las, serão apresentados previamente detalhes da metodologia da AMD, do coeficiente utilizado nas medidas de concordância da classificação e do desenho do *corpus*. A seguir, serão apresentados os resultados da pesquisa e, concomitantemente, sua discussão em relação a outros estudos da área.

3 Metodologia

A consecução de uma AMD pode ser total (BIBER, 1988; BERBER SARDINHA *et al.*, 2014b) ou parcial, como no caso deste estudo, em que é realizada uma adição de gêneros/registros a uma AMD total (BERBER SARDINHA *et al.*, 2019). A AMD envolve algumas etapas básicas (BERBER SARDINHA; VEIRANO PINTO, 2019), entre elas: 1. a coleta de um *corpus* extenso e balanceado da língua ou variedade de língua em uso, de modo a criar uma amostra de textos que representam os vários gêneros/registros que a compõem; 2. o uso de software capaz de anotar cada palavra em termos de suas características morfosintáticas, e de algoritmos que permitem contar e processar os dados resultantes do primeiro procedimento, para tabulá-los em matrizes em que são

levadas em conta algumas variáveis selecionadas, a partir de análise da anotação inicial; 3. o alinhamento dos tamanhos dos textos que compõem o *corpus* de estudo, por meio da normalização das frequências absolutas para uma frequência relativa, por mil palavras; 4. a análise fatorial, técnica estatística multivariada que busca investigar no *corpus* a existência de fatores latentes – grupos de variáveis coocorrentes que indicam uma influência não identificada originalmente na pesquisa –, ou dimensões; 5. o cálculo da padronização das variáveis e dos escores médios de gênero/registo nas dimensões, medidas que permitem a análise e a interpretação da variação linguística encontrada nas dimensões. Neste estudo, apenas a última etapa da AMD será executada, de modo a incorporar os dados relativos à concordância de gêneros/registros jornalísticos à pesquisa original de AMD efetuada em Kauffmann (2005).

O *corpus* coletado para representar a variedade de gêneros/registros jornalísticos no modo escrito foi composto por sete edições integrais da edição impressa do jornal "Folha de S.Paulo" (idêntica à versão digital, oferecida no formato pdf, e muito semelhante à edição online), de modo a possibilitar a construção de uma semana construída (KENNEDY, 1998, p. 75; KAUFFMANN, 2005) com um total de 1.431 textos. Todos foram processados pelo etiquetador on-line VISL, para a língua portuguesa (BICK, 2005), uma versão simplificada do PALAVRAS (BICK, 2014), posteriormente desenvolvido. Das 19 variáveis de ordem lexicogramatical que inicialmente compuseram a AMD, foram selecionadas por fim 13 variáveis na extração final da análise fatorial, com peso significativo (acima do valor de 0,3). A matriz padrão expressa na Tabela 1 reúne as variáveis coocorrentes da solução de dois fatores em Kauffmann (2005), interpretados como dimensões: Narrativo versus Expositivo (Dimensão 1) e Argumentativo versus Informativo (Dimensão 2). Quando uma variável tem um valor negativo, a relação de co-ocorrência é inversa: enquanto as

variáveis positivas do fator se manifestam em um texto, as negativas tendem a ficar ausentes.

Tabela 1 – Composição e peso das variáveis nas dimensões da AMD.

Variável	Dimensão 1	Dimensão 2
Pretérito perfeito	0,72	-0,36
Verbos / pronomes 3ª pessoa singular	0,64	
Conjunções subordinativas	0,52	0,40
Verbos públicos (falar, afirmar, dizer, etc.)	0,44	
Pretérito imperfeito	0,37	
Advérbios	0,35	
Verbos / pronomes 1ª pessoa singular	0,33	
Substantivos	-0,35	
Presente do indicativo		0,55
Pronomes demonstrativos		0,48
Quantidade de palavras		0,34
Nomes próprios		-0,58
Números cardinais		-0,35

Fonte: Kauffmann (2005, p. 92).

O peso que cada variável teve no cálculo de escores das dimensões é indireto. O escore é produto da soma das frequências relativas de cada texto nas variáveis de maior valor entre as dimensões (BIBER, 1988, p. 93), e não do peso relativo de cada variável na composição da dimensão. Os escores médios por gênero/registo foram resultantes da média das frequências padronizadas das observações (textos) agrupadas nos gêneros/registros. As análises de variância ANOVA (Tabela 2) realizadas *a posteriori* mostraram que as dimensões encontradas são significativamente relevantes para explicar a variação encontrada no *corpus* em relação aos diversos gêneros/registros jornalísticos.

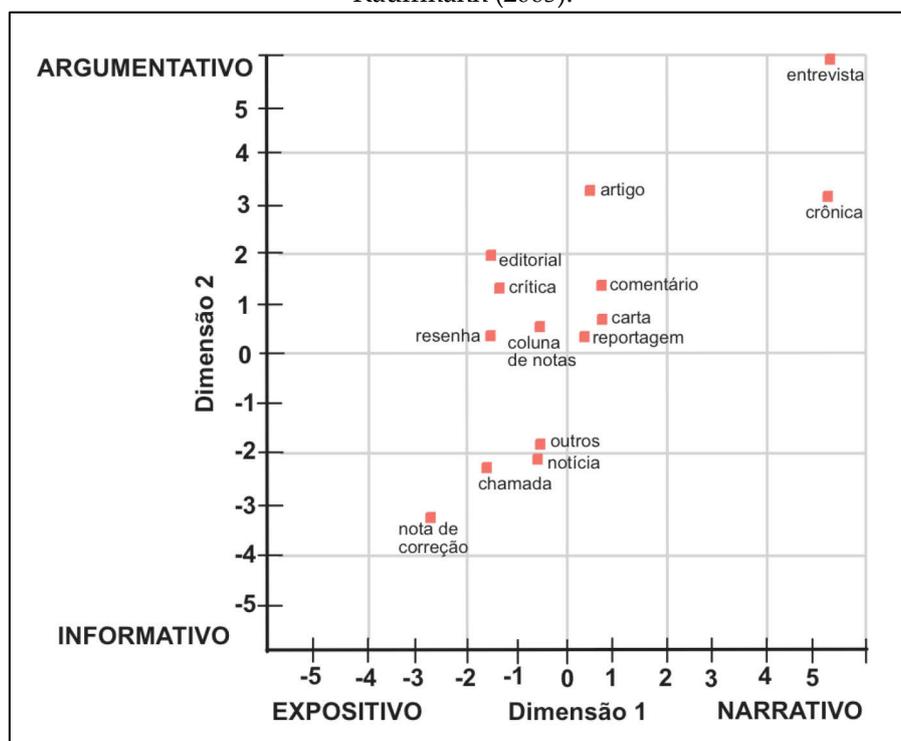
Tabela 2 – ANOVA entre gêneros/registros, por dimensão (valores de F, p e R²).

Dimensão	F	p	R ²
Dimensão 1	6,7	0,00000	5,8%
Dimensão 2	50,4	0,00000	31,6%

Fonte: Kauffmann (2005, p. 109).

As dimensões resultantes em Kauffmann (2005) assemelham-se às dimensões de variação D2 e D5 do português (BERBER SARDINHA *et al.*, 2014b). A Dimensão 1 de Kauffmann (2005) tem, no polo narrativo, variáveis de tempos verbais que coincidem com o posicionamento para o passado da Dimensão 5 de Berber Sardinha *et al.* (2014b) – pretéritos perfeito e imperfeito; já a Dimensão 2 de Kauffmann (2005) é semelhante no propósito argumentativo, embora ambas não compartilhem as variáveis que caracterizam essa dimensão (cf. discussão em BERBER SARDINHA *et al.*, 2014a).

Gráfico 1: Distribuição dos escores médios de gêneros/registros jornalísticos nas dimensões de Kauffmann (2005).



Fonte: adaptado de Kauffmann (2005, p. 115).

Com o cálculo dos escores médios por gênero/registo, foi possível posicionar graficamente cada gênero/registo ao longo dos eixos representados pelas dimensões resultantes (Gráfico 1). Os gêneros/registros entrevista, artigo, crônica, comentário, carta e reportagem pertencem aos quadrantes "Narrativo-Argumentativo"; em quadrante oposto, os registros chamada, notícia e nota de correção estão presentes no quadrante "Expositivo-Informativo". O quadrante "Expositivo-Argumentativo" concentra, em uma terceira região, os gêneros/registros editorial, resenha, crítica e coluna de notas (KAUFFMANN, 2005, p. 115).

A presente análise da classificação de gêneros/registros jornalísticos por especialistas, aqui relatada, originalmente apoiou a pesquisa conduzida em Kauffmann (2005), para confirmar a hipótese de que uma classificação efetuada segundo uma criteriosa revisão na literatura a respeito dos gêneros/registros na imprensa diária escrita do Brasil, por um só especialista (no caso, o autor), seria largamente coincidente com uma classificação de gêneros/registros jornalísticos feita por outros observadores especialistas, de forma independente. Foi logo percebido, porém, que os resultados tiveram a capacidade de iluminar alguns aspectos relativos à cognição de gêneros/registros jornalísticos por observadores especialistas, ultrapassando os propósitos iniciais do levantamento, motivo pelo qual justificou-se uma análise mais detida sobre os dados gerados.

A pesquisa desenvolvida lançou ao escrutínio de mais três especialistas uma amostra significativa do *corpus* de pesquisa de Kauffmann (2005), gerando dados inéditos relativos ao grau de concordância entre os gêneros/registros jornalísticos. O *corpus* abrangido pela pesquisa é composto por duas edições completas da "Folha", escolhidas aleatoriamente, perfazendo um total de 425 textos (136.330 palavras) por classificar – equivalentes a 29,6% do total de textos estudados em Kauffmann (2005). Cada um deles realizou a sua classificação individualmente, com instruções, planilha de preenchimento, cópias impressas de cada texto e a edição impressa onde foram

originalmente publicados, acompanhado de uma lista de gêneros/registros jornalísticos com 14 itens previamente estabelecida, e material de apoio com suas respectivas definições e referências, provenientes de fontes da literatura, como livros acadêmicos da área de comunicação e jornalismo, manuais de redação e estilo e dicionários especializados. Em comum, os observadores especialistas apresentavam experiência profissional jornalística acumulada por vários anos nas áreas editorial e de documentação em empresas de comunicação ligadas à produção de jornais. Nenhum deles teve contato com os demais para discutir assuntos ligados à classificação efetuada. Como resultado, chegou-se a um *corpus* de textos jornalísticos classificado de acordo com a tipologia usual de gêneros/registros jornalísticos identificada anteriormente pela literatura, porém graduada pela convergência – total ou parcial – ou divergência na atribuição de determinado gênero/registo a cada texto, entre os quatro especialistas.

O cômputo das menções de gênero/registo dadas aos textos do *corpus*, mostrado adiante, foi tabulado para refletir o grau de concordância/discordância entre gêneros/registros. A fim de interpretar os resultados derivados da pesquisa, um modo encontrado de analisar e sintetizar a opinião de um grupo de quatro classificadores, com um número razoável de alternativas à escolha, é proposto no Quadro 1.

Quadro 1 – Graus de concordância entre quatro observadores.

Grau de concordância	Menções de gênero(s)/registo(s) dadas a um texto do <i>corpus</i>
Concordância total	Quatro menções de um mesmo gênero/registo
Concordância predominante	Três menções de um mesmo gênero/registo e uma menção única de gênero/registo diverso
Concordância parcial	Duas menções duplas de um mesmo gênero/registo (Divisão 2 – 2) ou Uma menção dupla e duas menções únicas de gêneros/registros diversos (Divisão 2 – 1 – 1)
Discordância	Quatro menções únicas de gêneros/registros diversos

Fonte: elaborado pelo autor.

A distinção permite uma primeira análise dos casos concordes (aqueles com graus de concordância total e predominante) para a identificação de dimensões de linguagem capazes de distinguir entre os gêneros/registros jornalísticos mais frequentes do *corpus*, de forma avalizada.

Para avaliar de forma mais consistente os resultados, o recurso de mensuração da concordância interobservadores utilizado nesta pesquisa é o coeficiente kappa de Fleiss, uma medida de associação mais precisa que o percentual de concordância, pois ele verifica em que medida há concordância considerando as chances de acaso entre vários observadores (BIBER; EGBERT; DAVIES, 2015, p. 20). Foi utilizada uma variação do kappa de Fleiss, que considera o fato de que os classificadores não têm um número de casos fixado para cada categoria (RANDOLPH, 2005). Um valor de kappa igual ou acima de 0,7 é normalmente utilizado para indicar que houve concordância entre os observadores. Um valor a partir de 0,8 é considerado uma concordância "quase completa" (MIOT, 2016, p. 91).

4 Resultados

A classificação realizada pelo grupo de observadores especialistas resultou em vários conjuntos de textos, que apresentam níveis de concordância quanto ao gênero/registo, em maior ou menor grau: grupos de gêneros/registros com textos consensualmente percebidos pelos observadores de forma plena ou parcial, grupos de textos com divergência de opinião sobre gênero/registo etc. Todavia, tomados em conjunto, a classificação efetuada pelos observadores especialistas apresentou um coeficiente de concordância de 0,76 (kappa de Fleiss), indicando uma significativa concordância do grupo em relação às designações dos gêneros/registros jornalísticos atribuídos aos textos.

A Tabela 3 apresenta os resultados da pesquisa, organizados pelo número de textos atribuídos pelos classificadores de acordo com o grau de concordância, envolvendo todos os gêneros/registros jornalísticos mencionados na pesquisa. Mais de

86% dos textos do *corpus*, correspondentes à soma dos textos em que houve concordância total ou predominante, foram considerados pelo grupo de especialistas como portadores de características patentes que os fazem integrantes de um determinado gênero/registo jornalístico reconhecido socialmente.

Tabela 3 – Concordância para gêneros/registros jornalísticos de 425 textos.

Concordância total	Concordância predominante	Concordância parcial		Discordância
		Divisão 2 – 2	Divisão 2 – 1 – 1	
267	100	30	26	2
62,8%	23,5%	7,1%	6,1%	0,5%

Fonte: elaborada pelo autor.

Para caracterizar em maior detalhe como se dá a distribuição de gêneros/registros concordes no meio jornalístico, a Tabela 4 apresenta a proporção que possui cada gênero/registo concorde em relação ao total de textos, com base na soma das avaliações de concordância total e predominante. A mesma tabela exhibe também, abaixo da linha destacada, o total de textos e a proporção atingida pelos textos que obtiveram classificação de concordância parcial ou discordância.

Tabela 4 – Frequência de gêneros/registros jornalísticos concordes no *corpus*.

Gênero/Registro	Número de textos	Percentual
Reportagem	201	47,3%
Notícia	52	12,2%
Crítica	22	5,2%
Chamada	20	4,7%
Artigo	19	4,5%
Carta	18	4,2%
Coluna de notas	15	3,5%
Editorial	7	1,6%
Entrevista	5	1,2%
Outros	4	0,9%
Crônica	2	0,5%
Nota de correção	2	0,5%
Resenha	0	0,0%
Comentário	0	0,0%

Concordância parcial de gêneros/registros	56	13,2%
Discordância	2	0,5%
Total	425	100%

Fonte: elaborada pelo autor.

Em relação à proporção de gêneros/registros encontrada em Kauffmann (2005, p. 105) e Chaparro (1997, p. 139), o percentual atribuído à notícia discrepa à primeira vista – 20% nos estudos anteriores, em contraste com os 12,2% desta pesquisa –, enquanto a proporção atribuída ao gênero/registro reportagem, de 50%, é semelhante, e o mesmo ocorre em relação aos demais gêneros/registros. No entanto, como será demonstrado a seguir, a proporção do gênero/registro notícia alcança valores próximos aos dos estudos anteriores, se for levada em consideração sua presença hegemônica na categoria de "Concordância parcial de registros".

Cabe analisar em detalhe, portanto, os casos em que houve dissenso na classificação. Biber, Egbert e Davies (2015) decompuseram a concordância parcial sobre gêneros/registros quando analisaram a classificação de quatro observadores a respeito de gêneros/registros em textos da internet, dividindo-a em dois grupos: o de duas menções duplas de dois gêneros/registros; e o de uma menção dupla de um gênero/registro, com duas menções únicas de gêneros/registros diversos. No primeiro caso, os autores aventam a possibilidade de haver gêneros/registros híbridos que possuem características comuns dos dois gêneros/registros para conformar um tipo de gênero/registro híbrido. Nesta pesquisa, igualmente foram mantidos os subgrupos da categoria de concordância parcial, no intuito de investigar em que medida as diferenças de concordância parcial se refletem na mensuração dos grupos de textos, em relação às dimensões de variação do texto jornalístico de Kauffmann (2005), e se é possível admitir uma ideia de gênero/registro híbrido com as evidências disponíveis.

Os textos que apresentaram uma classificação de concordância parcial do tipo Divisão 2 – 2 totalizam 30 textos. Seus gêneros/registros componentes estão

apresentados na Tabela 5. Evidencia-se a tensão terminológica entre os gêneros/registros reportagem e notícia, com a presença de 22 textos no *corpus* com menções duplas a esses gêneros/registros, somando dois terços dos casos nessa situação de concordância. Destaca-se secundariamente a divisão de opiniões entre reportagem e crítica, existente em quatro textos.

Tabela 5 – Gêneros/registros com concordância parcial (Divisão 2 – 2).

Gêneros/registros com concordância parcial (Divisão 2– 2)		Nº de textos
Reportagem	Notícia	22
Reportagem	Crítica	4
Notícia	Chamada	1
Artigo	Crítica	1
Artigo	Comentário	1
Artigo	Crônica	1
Total		30

Fonte: elaborada pelo autor.

Por sua vez, a análise dos resultados de concordância parcial que obedecem à Divisão 2 – 1 – 1, exibida na Tabela 6, revela um número relevante de textos em que os gêneros/registros reportagem e notícia estão presentes, como gênero/registo majoritário ou não, na classificação composta em que estão sempre presentes três gêneros/registros. Nessa categoria de baixa concordância, notícia e reportagem estão associados a crítica, "outros", coluna de notas, artigo, carta e chamada.

Tabela 6 – Gêneros/registros com concordância parcial (Divisão 2 – 1 – 1).

Gêneros/registros com concordância parcial (Divisão 2 – 1 – 1)			Nº de textos
Menção dupla	Menção única	Menção única	
Notícia	Reportagem	Outros	8
Artigo	Comentário	Coluna de notas	5
Reportagem	Notícia	Crítica	4
Reportagem	Resenha	Crítica	2
Reportagem	Notícia	Artigo	1

Reportagem	Notícia	Coluna de notas	1
Notícia	Reportagem	Carta	1
Notícia	Crítica	Outros	1
Notícia	Comentário	Outros	1
Chamada	Reportagem	Notícia	1
Artigo	Reportagem	Comentário	1
Total			26

Fonte: elaborada pelo autor.

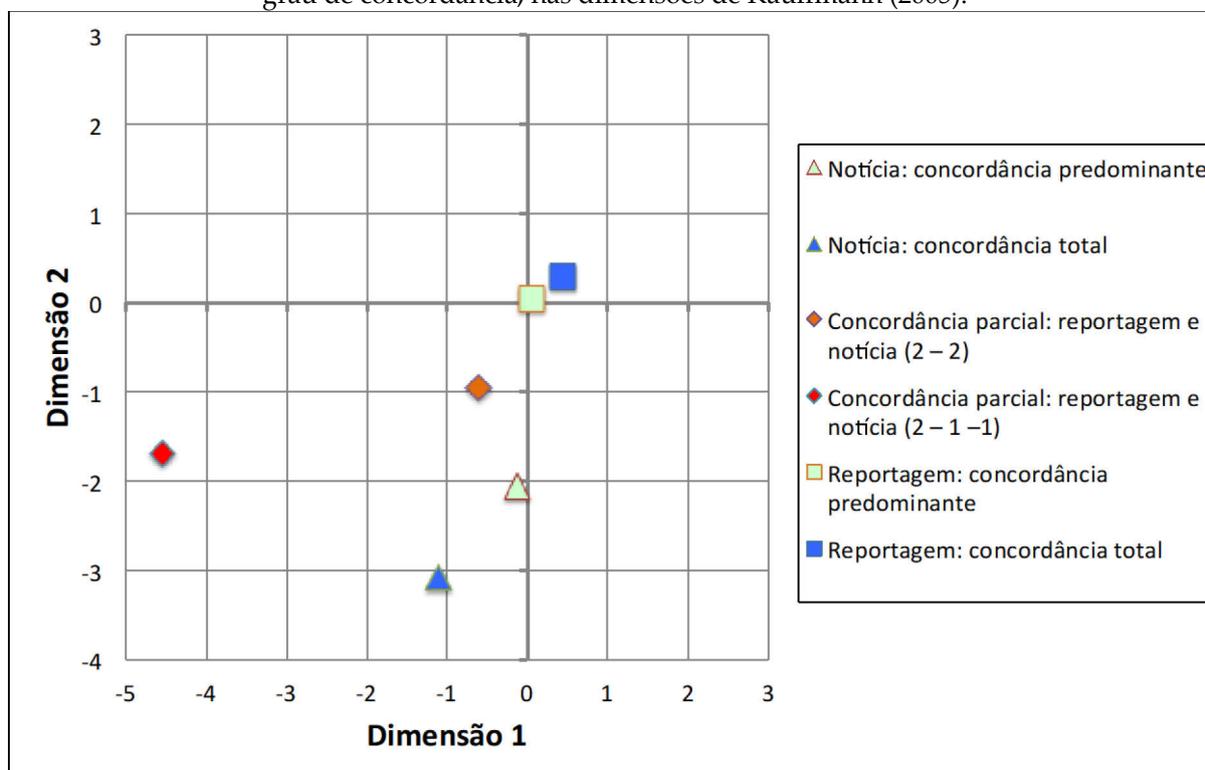
Quanto aos casos de discordância, conforme visto na Tabela 4, eles só ocorreram duas vezes no *corpus*. Devido à interpretação complexa e ao baixo número apresentado, não serão comentados.

Observou-se que a principal divisão de opiniões entre os observadores especialistas está concentrada na oposição reportagem *versus* notícia. É possível argumentar que a razão para ocorrer essa dicotomia é a proximidade formal entre os dois gêneros/registros, emblemáticos da atividade jornalística. Desse modo, o conflito mostrado nas Tabelas 5 e 6 influenciou significativamente a baixa proporção de textos concordes classificados como notícia, conforme mostrado anteriormente. Por sua vez, observou-se que o gênero/registo reportagem sofreu impacto atenuado de variação de proporção devido à concordância parcial, pelo fato de ser o gênero/registo de maior expressão no *corpus*.

Para examinar de uma forma empírica as diferenças encontradas nos gêneros/registros jornalísticos reportagem e notícia em termos de sua concordância, recorreu-se ao modelo de mensuração de variação linguística proporcionado pela Análise Multidimensional aditiva (BERBER SARDINHA *et al.*, 2019). A Análise Multidimensional permite que se disponham nas suas dimensões as respectivas frequências manifestadas nos grupos de textos, em uma escala. Aqui, os grupos de textos submetidos à análise são compostos pelos gêneros/registros concordes de reportagem e notícia – ou seja, textos que obtiveram concordância total ou predominante na classificação por observadores especialistas – e por uma seleção dos grupos de textos que obtiveram concordância parcial, especificamente aqueles textos

classificados como notícia ou reportagem. Ou seja, dos 30 textos com concordância parcial (Divisão 2 – 2), foi selecionado o grupo de 22 textos que possui menções duplas simultaneamente de gêneros/registros reportagem e notícia; enquanto dos 26 textos na situação de concordância parcial (Divisão 2 – 1 – 1), foi selecionado o grupo de 15 textos com menções duplas em reportagem ou notícia, com uma menção única do outro gênero/registo, em complementação. Os resultados estão mostrados no Gráfico 2.

Gráfico 2 – Mapeamento dos escores médios dos gêneros/registros reportagem e notícia, conforme grau de concordância, nas dimensões de Kauffmann (2005).



Fonte: elaborado pelo autor.

Os escores médios dos gêneros/registros notícia e reportagem que têm concordância total/predominante estão bem próximos dos resultados obtidos em Kauffmann (2005). Apesar de não apresentados graficamente, as contagens dos demais grupos de gêneros/registros jornalísticos de concordância total/predominante apresentaram posicionamento semelhante à disposição dos gêneros/registros nos quadrantes formados pelas dimensões de Kauffmann (2005), exceto talvez pelo

gênero/registro artigo, que mudou de quadrante, conforme mostram os dados apresentados no Apêndice.

Os escores médios dos grupos de gêneros/registros notícia e reportagem com concordância predominante estão próximos dos escores de concordância total de seus respectivos gêneros/registros, indicando uma afinidade de características compartilhadas. Esses escores, em alguma medida, têm pequenas diferenças que os afastam do grupo de concordância total, e seus pontos correspondentes no Gráfico 2 indicam que essas diferenças os aproximam na direção do gênero/registro oposto.

No grupo de concordância parcial em que notícia e reportagem têm opiniões divididas (modelo 2 – 2), o ponto que o representa no Gráfico 2 está em uma região praticamente equidistante das regiões ocupadas pelos gêneros/registros concordes de reportagem e notícia, nas Dimensões 1 e 2. Esse posicionamento sugere que as características linguísticas que os textos jornalísticos noticiosos carregam têm uma considerável participação na distinção de gêneros/registros efetuada por observadores especialistas. A configuração linguística se equilibra em uma zona intermediária entre os dois gêneros/registros. Um exemplo de texto nessa situação é mostrado a seguir:

Exemplo 1 – Concordância parcial entre reportagem e notícia (Divisão 2 – 2)

Emenda do BC é promulgada por Sarney
DA SUCURSAL DE BRASÍLIA

Em rápida sessão do Congresso, realizada no plenário do Senado com a presença de apenas cinco parlamentares, o presidente do Senado, José Sarney (PMDB-AP), promulgou a emenda constitucional que permite a regulamentação do sistema financeiro nacional por várias leis complementares – e não mais por apenas uma lei, como a Constituição determinava, em seu artigo 192.

A emenda, que abre caminho para a aprovação de uma proposta de autonomia do Banco Central, revogou todos os incisos e parágrafos desse dispositivo constitucional – que trata do sistema financeiro –, inclusive o que estabelecia o teto de 12% ao ano para as taxas de juros reais.

A proposta original de alteração do artigo 192 foi apresentada em 1997 pelo então senador José Serra (PSDB-SP). Na CCJ (Comissão de Constituição e Justiça), recebeu substitutivo do senador Jefferson Péres (PDT-AM), que foi enviado à Câmara e aprovado pelos deputados neste mês de maio em segundo turno.

A emenda mantém apenas o caput do artigo 192 com uma alteração, determinando que o sistema financeiro nacional será regulado em leis complementares. (CP5_0788.TXT)

Nos escores médios dos textos de concordância parcial no modelo de divisão 2 – 1 – 1, a Dimensão 2 parece indicar um papel mais importante que o da Dimensão 1 em relação à distinção dos gêneros/registros. A Dimensão 2, ligada a aspectos de argumentação *versus* informação, possivelmente seja um agente diferenciador mais eficaz entre os dois gêneros/registros, sendo a notícia mais vinculada com o aspecto informativo, enquanto a reportagem teria, por outro lado, um viés argumentativo. Em relação à Dimensão 1, esses grupos de concordância parcial de reportagem e notícia em composição com outros gêneros/registros se diferenciam dos demais grupos encontrados. Valores negativos altos na dimensão indicam que esse grupo possui textos com poucos verbos e complementos verbais, enquanto apresentam uma alta frequência de substantivos. O Exemplo 2 ilustra as características do grupo:

Exemplo 2 – Concordância parcial entre reportagem e notícia (Divisão 2 – 1 – 1)

Como são transmitidas as informações

ESPECIAL PARA A FOLHA

São três as principais formas de conectar periféricos sem fios: infravermelho, bluetooth e radiofrequência.

O infravermelho é comum em equipamentos portáteis. Basta apontar um equipamento para o outro a uma distância de até 40 cm para que eles se enxerguem. A velocidade de transmissão de dados chega a 4 Mbps.

A tecnologia bluetooth funciona com ondas de radiofrequência. Seu alcance é de aproximadamente 10 metros, e a velocidade de conexão pode atingir uma taxa de até 720 Kbps.

A radiofrequência (802.11) pode alcançar 11 Mbps, mas logo haverá placas que possibilitarão a transmissão de dados com velocidades superiores a 50 Mbps.

Alguns fabricantes já vendem produtos com a tecnologia 802.11g, que deve alcançar uma velocidade de 54 Mbps, ou seja, quase cinco vezes mais rápida. (JAR) (CP3_0533.TXT)

Os gêneros/registros reportagem e notícia mantêm uma certa ambiguidade discursiva. Por um lado, subsiste uma separação clara entre os grupos concordes dos gêneros/registros na percepção de especialistas, que reconhece a existência de traços discerníveis na produção de uma notícia ou de uma reportagem. Porém, de outra parte, pode-se admitir que se manifestam concomitantemente grupos intermediários de textos que estão a meio caminho dos dois gêneros/registros, gerando divergências na sua classificação – por exemplo quando, em um texto jornalístico, a carga informativa concentrada da notícia equilibra-se com os fatos narrados pela ação da reportagem.

5 Considerações finais

Os dados apresentados permitiram observar os gêneros/registros que atuam na imprensa diária escrita no Brasil sob um ponto de vista abrangente, tanto de ordem qualitativa como quantitativa. Através dos métodos empregados na pesquisa, foi possível concluir que 86,3% (367 textos, de um total de 425) dos textos classificados pelo grupo de observadores especialistas são concordes. Os gêneros/registros mais bem diferenciados pelo grupo foram editorial, nota de correção, chamada, carta, entrevista e reportagem. Já os gêneros/registros crônica, crítica, notícia, comentário e resenha demonstraram ser aqueles que apresentaram maior divergência de opinião.

Os achados confirmam a pesquisa de Bonini (2009), que por outros meios teóricos também detectou que existe uma interseção entre as áreas que delimitam os gêneros/registros reportagem e notícia. Nesse aspecto, a presente pesquisa não mostrou que essas zonas intermediárias teriam autonomia linguística capaz de produzir novas terminologias de registro. Assim, a hipótese de existência de gêneros/registros híbridos, como proposta por Biber, Egbert e Davies (2015), parece não se confirmar com os dados encontrados no ambiente do jornal, certamente mais restrito que o universo da *web*. Pode-se aventar que seja efeito de textos com

componentes dos dois gêneros/registros que teriam levado à indecisão na classificação, mas seria preciso efetuar novas pesquisas para investigar esse processo. Espera-se, por fim, que os resultados obtidos possam vir a contribuir com outros trabalhos que abordem aspectos relativos à cognição de gêneros/registros da imprensa.

Referências bibliográficas

AITCHISON, J.; LEWIS, D. M. (org.). **New media language**. Abingdon; New York: Routledge, 2003. DOI <https://doi.org/10.4324/9780203696965>

BEDNAREK, M. **Evaluation in media discourse**: analysis of a newspaper corpus. London; New York: Continuum, 2006.

BELL, A. **The Language of News Media**. Oxford: Blackwell, 1991.

BERBER SARDINHA, T. Análise Multidimensional. **D.E.L.T.A.**, v. 16, n. 1, p. 99-127, 2000. DOI <https://doi.org/10.1590/S0102-44502000000100005>

BERBER SARDINHA, T. **Linguística de Corpus**. Barueri: Manole, 2004.

BERBER SARDINHA, T. ET AL. A multi-dimensional analysis of register variation in Brazilian Portuguese. **Corpora**, v. 9, n. 2, p. 239-271, 2014 (a). DOI <https://doi.org/10.3366/cor.2014.0059>

BERBER SARDINHA, T. ET AL. Dimensions of register variation in Brazilian Portuguese. *In*: BERBER SARDINHA, T.; VEIRANO PINTO, M. (org.). **Multi-Dimensional Analysis, 25 years on**: a tribute to Douglas Biber. Amsterdam: John Benjamins, 2014 (b). p. 35-79. DOI <https://doi.org/10.1075/scl.60>

BERBER SARDINHA, T. *et al.* Adding registers to a previous Multi-Dimensional Analysis. *In*: BERBER SARDINHA, T.; VEIRANO PINTO, M. (org.). **Multi-Dimensional Analysis**: research methods and current issues. London: Bloomsbury, 2019. p. 165-186. DOI <https://doi.org/10.5040/9781350023857.0017>

BÉRTOLI-DUTRA, P. **Linguagem da música popular anglo-americana de 1940 a 2009**. 2010. 290 f. Tese (Doutorado em Linguística Aplicada e Estudos da Linguagem) – LAEL, Pontifícia Universidade Católica de São Paulo, São Paulo, 2010.

BIBER, D. **Variation across speech and writing**. Cambridge: Cambridge University Press, 1988. DOI <https://doi.org/10.1017/CBO9780511621024>

BIBER, D. **Dimensions of register variation: A Cross-linguistic Comparison**. Cambridge: Cambridge University Press, 1995. DOI <https://doi.org/10.1017/CBO9780511519871>

BIBER, D. **University language: a corpus-based study of spoken and written registers**. Amsterdam; Philadelphia: John Benjamins, 2006. DOI <https://doi.org/10.1075/scl.23>

BIBER, D.; CONRAD, S. **Register, genre and style**. Cambridge: Cambridge University Press, 2009. DOI <https://doi.org/10.1017/CBO9780511814358>

BIBER, D. *et al.* **Longman grammar of spoken and written English**. London: Longman, 1999.

BIBER, D.; EGBERT, J.; DAVIES, M.. Exploring the composition of the searchable web: a corpus-based taxonomy of web registers. *Corpora*, n. 101, p. 11-45, 2015. DOI <https://doi.org/10.3366/cor.2015.0065>

BICK, E. Gramática constritiva na análise automática da sintaxe portuguesa. *In*: BERBER SARDINHA, T. (org.). **A língua portuguesa no computador**. Campinas: Mercado de Letras, 2005. p. 91-112.

BICK, E. PALAVRAS, a constraint grammar-based parsing system for Portuguese. *In*: BERBER SARDINHA, T.; FERREIRA, T. S. B. (org.). **Working with Portuguese corpora**. London; New York: Bloomsbury; Continuum, 2014.

BONINI, A. **Gêneros textuais e cognição**. Florianópolis: Insular, 2002.

BONINI, A. Os gêneros do jornal: o que aponta a literatura da área de comunicação do Brasil? **Linguagem em (Dis)curso**, v. 4, n. 1, jul./dez., 2003.

BONINI, A. The distinction between news and reportage in the brazilian journalistic context: a matter of degree. *In*: BAZERMAN, C.; BONINI, A.; FIGUEIREDO, D. (org.). **Genre in a changing world**. Perspectives on writing. Fort Collins, Colorado: The WAC Clearinghouse and Parlor Press, 2009. Disponível em: <https://wac.colostate.edu/books/genre/>. Acesso em: 23 out. 2021. DOI <https://doi.org/10.37514/PER-B.2009.2324.2.10>

BONINI, A. Análise crítica de gêneros jornalísticos. SBPJor – Associação Brasileira de Pesquisadores em Jornalismo. **10º Encontro Nacional de Pesquisadores em Jornalismo**. Pontifícia Universidade Católica do Paraná, Curitiba, 2012.

BRASIL. **Parâmetros Curriculares Nacionais** – 5ª a 8ª série do Ensino Fundamental. Brasília: SEF/MEC, 1998.

CHAPARRO, M. C. **Jornalismo, discurso em dois gêneros**. 1997. 261 p. Tese (Livre Docência) – Escola de Comunicações e Artes, USP, São Paulo, 1997.

COSTA, L.A. Gêneros jornalísticos. *In*: J. MARQUES DE MELO; F. de ASSIS (org.). **Gêneros jornalísticos no Brasil**. São Bernardo do Campo: Universidade Metodista de São Paulo, 2010. p. 43-83.

EDITORA ABRIL. **Manual de estilo**. Rio de Janeiro: Nova Fronteira, 1990.

FAIRCLOUGH, N. **Media discourse**. London: Edward Arnold, 1995.

FOLHA DE S.PAULO. **Manual geral da Redação**. São Paulo: Folha de S.Paulo, 1984.

FOLHA DE S.PAULO. **Manual geral da Redação**. São Paulo: Folha de S.Paulo, 1987.

FOLHA DE S.PAULO. **Novo manual da Redação**. São Paulo: Folha de S.Paulo, 1992.

FOLHA DE S.PAULO. **Manual da Redação**. São Paulo: Publifolha, 2001.

FOLHA DE S.PAULO. **Manual da Redação**. São Paulo: Publifolha, 2018.

KAUFFMANN, C. H. **O corpus do jornal**: variação linguística, gêneros e dimensões da imprensa diária escrita. 2005. 202 f. Dissertação (Mestrado em Linguística Aplicada e Estudos da Linguagem) – LAEL, Pontifícia Universidade Católica de São Paulo, São Paulo, 2005.

KAUFFMANN, C. H. Caracterização linguística de gêneros textuais do português brasileiro. **VIII Simpósio Internacional de Estudos de Gêneros Textuais (SIGET)**, 2015. Comunicação. São Paulo: 2015.

KENNEDY, G. **An Introduction to corpus linguistics**. London: Longman, 1998.

- MARCUSCHI, L. A. Gêneros textuais: definição e funcionalidade. *In*: DIONISIO, A. P.; MACHADO, A. R.; BEZERRA, M. A. (org.). **Gêneros textuais e ensino**. Rio de Janeiro: Lucerna, 2002. p. 19–36.
- MARQUES DE MELO, J.; ASSIS, F. (org.). **Gêneros jornalísticos no Brasil**. São Bernardo do Campo: Universidade Metodista de São Paulo, 2010.
- MARQUES DE MELO, J.; ASSIS, F. Gêneros e formatos jornalísticos: um modelo classificatório. *Intercom - RBCC São Paulo*, v. 39, n. 1, p. 39-56, jan./abr., 2016. DOI <https://doi.org/10.1590/1809-5844201613>
- MIOT, H. A. Análise de concordância em estudos clínicos e experimentais. *J. vasc. bras.*, v. 15, n. 2, p. 89-92, 2016. DOI <https://doi.org/10.1590/1677-5449.004216>
- O ESTADO DE S.PAULO. **Manual de redação e estilo**. MARTINS, E. (org.). São Paulo: O Estado de S.Paulo, 1990.
- O GLOBO. **Manual de redação e estilo**. GARCIA, L. (org.). São Paulo: Editora Globo, 1998.
- RANDOLPH, J. J. 2005. Free-marginal multirater kappa: An alternative to Fleiss' fixed-marginal multirater kappa. Comunicação. **Joensuu University Learning and Instruction Symposium 2005**, Joensuu, Finland, October 14-15th, 2005. (ERIC Document Reproduction Service No. ED490661).
- SEIXAS, L. **Redefinindo os gêneros jornalísticos**: proposta de novos critérios de classificação. Covilhã: LabCom, 2009.
- SOUZA, R. C. **A revista Time em uma perspectiva multidimensional**. 2012. 330 f. Tese (Doutorado em Linguística Aplicada e Estudos da Linguagem) – LAEL, Pontifícia Universidade Católica de São Paulo, São Paulo, 2012.
- SWALES, J. M. **Genre analysis**: English in academic and research settings. Cambridge: Cambridge University Press, 1990.
- VAN DIJK, T. A. **News as discourse**. Hillsdale, New Jersey: Erlbaum, 1988.
- VAN DIJK, T. A. **Discurso e poder**. São Paulo: Contexto, 2010.
- VEIRANO PINTO, M. **A linguagem dos filmes norte-americanos ao longo dos anos**: uma abordagem multidimensional. 2013. 467 f. Tese (Doutorado em Linguística

Aplicada e Estudos da Linguagem) – LAEL, Pontifícia Universidade Católica de São Paulo, São Paulo, 2013.

Apêndice

Tabela 7 – Registros jornalísticos concordes: posicionamento nas dimensões de Kauffmann (2005).

Registro	D1	D2
Reportagem	0,32	0,22
Notícia	-0,80	-2,77
Artigo	-0,39	1,97
Editorial	-2,10	1,14
Coluna de notas	-0,89	0,00
Crítica	-1,30	0,95
Chamada	-1,36	-2,39
Carta	0,63	0,20
Entrevista	3,86	3,74

Fonte: elaborada pelo autor.

Artigo recebido em: 31.10.2021

Artigo aprovado em: 30.05.2022