Register Variation on the Searchable Web: A multi-dimensional analysis

Variação de registro na Internet: uma análise multidimensional

Douglas Biber e Jesse Egbert

Tradução para o português do Brasil:
Patrícia Pereira Bértoli (UERJ-GELC) e
Simone Vieira Resende (EBAC-GELC)
https://orcid.org/0000-0002-1439-3840

Revisão:
Flávia Azeredo-Cerqueira (Johns Hopkins University)

Apresentação

O artigo traduzido aqui, originalmente publicado no *Journal of English Linguistics*, em 2016, aborda uma série de questões relevantes para o campo da Linguística de *Corpus*, principalmente em relação à análise multidimensional e os conceitos e definições de registro. Mais especificamente, o artigo descreve a coleta de um *corpus* que se apresenta como representativo dos variados registros presentes em toda a Internet, a partir de amostras de textos publicados e pesquisáveis. A partir daí, por meio da metodologia de análise multidimensional, os registros são agrupados por suas semelhanças e diferenças na coocorrência de características linguísticas com base funcional compondo nove dimensões que abraçam esses registros.*

Biografias dos Autores

Douglas Biber é professor regente de inglês (Linguística Aplicada) na Northern Arizona University. O foco de sua pesquisa tem sido em linguística *corpus*, gramática da língua inglesa e variação de registro (em inglês e cross-linguística; sincrônica e diacrônica). Tem mais de 200 artigos de pesquisa publicados, 8 livros editados e 15 livros e monografias de autoria; estes incluem um livro sobre Registo, Gênero e Estilo (Cambridge, 2009), a coautoria da Gramática da Longman de Inglês Falado e Escrito (1999), e outros livros acadêmicos sobre complexidade gramatical em inglês acadêmico (Cambridge, 2016), registros universitários americanos (Benjamins, 2006), análise de discurso baseada em *corpus* (Benjamins, 2007) e Análises Multi-Dimensionais de variação de registro (Cambridge 1988, 1995).

Jesse Egbert é professor assistente do Departamento de Linguística e Língua Inglesa na Northern Arizona University. Sua pesquisa tem foco na variação linguística entre registros, particularmente escrita acadêmica e linguagem da Internet. Também está interessado em questões metodológicas em linguística *corpus*, incluindo desenho de *corpus*, métodos estatísticos e triangulação metodológica, que é o tema de um volume próximo, coeditado com Paul Baker, intitulado Abordagens em Triangulação Metodológicas em Pesquisas de Linguística do *Corpus*.

RESUMO

Grande parte das pesquisas linguísticas sobre a internet tem como base o estudo de características linguísticas específicas que ocorrem na linguagem da internet (por exemplo, o uso de *emoticons*, abreviaturas, contrações e acrônimos) e também os "novos" registros da internet, aqueles mais evidentes, como por exemplo, *blogs*, fóruns da Internet, mensagens instantâneas e *tweets*. A análise multidimensional (AMD) já foi utilizada para investigar registros da internet, principalmente na análise de características gramaticais fundamentais, como por exemplo, substantivos, verbos e preposições. Uma pesquisa de cunho multidimensional difere de forma teórica e metodológica da maioria das abordagens de pesquisa no campo da linguística na medida em que ela se constrói a partir da noção de coocorrência linguística, que defende a ideia de que as diferenças entre registros podem ser descritas de forma mais adequada quando consideramos os conjuntos de características linguísticas que possuem base funcional. Contudo, a maioria dos estudos multidimensionais já realizados anteriormente são semelhantes a outras pesquisas sobre os novos registros da internet, como por exemplo, *blogs*, *posts* do *Facebook*/*Twitter* e mensagens de e-mail. Estes são os registros que quase sempre associamos com a internet, e por isso faz sentido que eles sejam o foco da maioria das pesquisas já realizadas. No entanto, isso só mostra como sabemos pouco sobre a complexidade dos registros encontrados na web e os padrões de variação linguística entre eles. Este é o objetivo do presente estudo. Em vez de começar com o foco nos novos registros que são considerados interessantes por natureza, analisamos uma amostra representativa de toda a web. Os usuários finais codificaram as características situacionais e comunicativas de cada documento do *corpus*, levando a uma gama muito mais ampla de categorias de registro do que as utilizadas em qualquer outro estudo feito anteriormente: oito categorias gerais; várias categorias de registros híbridos; e vinte e sete categorias de registros específicos. Esta abordagem é capaz de gerar uma amostra muito mais inclusiva e diversificada de registros da web do que qualquer outro estudo já realizado em língua inglesa. O objetivo deste estudo é documentar os padrões de variação linguística que subjazem esses registros. Por meio da AMD, revelamos as dimensões da variação linguística da web e as semelhanças e diferenças entre os registros que compõem essas dimensões.

PALAVRAS-CHAVE: Registros da web; Linguagem da internet; Análise multidimensional; Registros híbridos; Variação de registro

Register Variation on the Searchable Web:
A multi-dimensional analysis

ABSTRACT

Most previous linguistic investigations of the web have focused on special linguistic features associated with Internet language (e.g., the use of emoticons, abbreviations, contractions, and acronyms) and the "new" Internet registers that are especially salient to observers (e.g., blogs, Internet forums, instant messages, *tweets*). Multi-Dimensional (MD) analysis has also been used to analyze Internet registers, focusing on core grammatical features (e.g., nouns, verbs, prepositional phrases). MD research differs theoretically and methodologically from most other research approaches in linguistics in that it is built on the notion of linguistic co-occurrence, with the claim that register differences are best described in terms of sets of co-occurring linguistic features that have a functional underpinning. At the same time, though, most previous MD studies are similar to other previous research in their focus on new Internet registers, such as blogs, *Facebook/ Twitter posts*, and email messages. These are the registers that we immediately think of in association with the Internet, and thus it makes sense that they should be the focus of most previous research. However, that emphasis means that we know surprisingly little at present about the full range of registers found on the web and the patterns of linguistic variation among those registers. This is the goal of the present study. Rather than beginning with a focus on new registers that are assumed to be interesting, we analyze a representative sample of the entire searchable web. End-users coded the situational and communicative characteristics of each document in our *corpus*, leading to a much wider range of register categories than that used in any previous linguistic study: eight general categories; several hybrid register categories; and twenty-seven specific register categories. This approach thus leads to a much more inclusive and diverse sample of web registers than that found in any previous study of English Internet language. The goal of the present study is to document the patterns of linguistic variation among those registers. Using MD analysis, we explore the dimensions of linguistic variation on the searchable web, and the similarities and differences among web registers with respect to those dimensions.

KEYWORDS: Web register; Internet language; Multi-dimensional analysis; Hybrid registers; Register variation

1. Introdução

O interesse linguístico pela internet tem duas preocupações principais: 1) o desejo de usar a web como um *corpus* capaz de representar a língua inglesa de uma forma geral (ou qualquer outro idioma) e 2) desenvolver uma pesquisa capaz de descrever os padrões de variação linguística que subjazem esse tipo de discurso. A pesquisa que se encaixa no primeiro grupo é geralmente chamada de WAC, da sigla em inglês *Web as Corpus*, que significa a internet como *corpus*. A base do pressuposto de tal pesquisa é a crença de que a web é tão grande e tão diversificada que pode ser usada como um *corpus* substituto capaz de representar a língua como um todo (GATTO, 2014).

Naturalmente, essa suposição é apenas uma crença esperançosa, a menos que esteja apoiada por evidências empíricas. Tais evidências só costumam aparecer em pesquisas feitas pelo segundo grupo: investigações empíricas a respeito da composição da web e dos padrões de variação encontrados. A maioria das pesquisas linguísticas desse tipo tem como foco características linguísticas especiais associadas à linguagem da internet (por exemplo, uso de *emoticons*, abreviaturas, contrações e acrônimos) e os "novos" registros da internet que se mostram mais evidentes para os estudiosos (por exemplo, *blogs*, fóruns da internet, mensagens instantâneas, *tweets*; como apresentados nos estudos feitos por Crystal (2001) e Baron (2008).

A análise multidimensional (AMD) também já foi usada para analisar registros da internet (BIBER; KURJIAN, 2007; GRIEVE et al. 2011; HARDY; FRIGINAL, 2012; TITAK, ROBERSON, 2013; BERBER-SARDINHA, 2014). Esses estudos diferem dos outros estudos na medida em que focam no uso das características gramaticais que são realmente relevantes e não apenas em algumas características linguísticas específicas que podem ser associadas à linguagem da internet. Assim, é possível afirmar que a AMD leva em consideração o uso de características como pronomes, substantivos, tempos verbais, advérbios, etc. A principal inovação teórica e metodológica da AMD é sua capacidade de analisar coocorrências linguísticas, uma vez que as características lexicogramaticais são analisadas estatisticamente para assim determinar a forma como elas coocorrem nos textos (ver Seção 3). Essas AMDs, portanto, diferem de outros estudos a respeito da linguagem da internet devido ao foco linguístico.

Por outro lado, a maioria desses estudos feitos por meio de uma AMD é semelhante a outras pesquisas já feitas e que focaram nos novos registros da internet, como *blogs,* postagens do *Facebook*/*Twitter* e mensagens de e-mail. Esses são os registros que logo nos vêm à cabeça quando pensamos na internet, e por isso faz sentido que sejam o foco da maioria das pesquisas anteriores. No entanto, tal ênfase nos mostra que, atualmente, sabemos muito pouco sobre a totalidade de registros que encontramos na web e quase nada sobre os padrões linguísticos desses registros.[1]

Este é o objetivo do presente estudo. Em vez de começar com um conjunto de registros que são considerados interessantes, tentamos construir uma amostra representativa de toda a web. Usuários finais codificaram as características situacionais e comunicativas de cada documento do *corpus* de estudo, levando a uma gama muito mais ampla de categorias de registros do que já utilizada em qualquer estudo linguístico anterior: oito categorias gerais; várias categorias de registros híbridos; e vinte e sete categorias de registros específicos. Essa abordagem apresenta uma amostra muito mais inclusiva e diversificada de registros da web do que aquelas feitas em estudos anteriores sobre a linguagem da internet em língua inglesa. O objetivo deste estudo é documentar os padrões de variação linguística que subjazem esses registros. Por meio da AMD, revelamos as dimensões da variação linguística da web e as semelhanças e diferenças entre os registros que compõem essas dimensões.

2. Representação do conjunto de registros encontrados na *web*: os pré-requisitos para a descrição linguística

Estudos sobre variação linguística de um tipo específico de discurso costumam seguir três etapas metodológicas principais: 1) a pesquisa é realizada para identificar os principais registros que existem em um determinado tipo de discurso; 2) procedimentos de amostragem são utilizados para a coleta de um subcorpus representativo para cada um dos registros estudados;

[1] Biber e Kurjian (2007) têm um trabalho excepcional a esse respeito. Em vez de começar com um conjunto de registros especiais da internet, este estudo investigou todos os documentos advindos de dois domínios do Google (Ciência e Casa). Usando a AMD associada à Análise de Cluster, o estudo documentou os "tipos de texto" linguisticamente definidos que ocorreram nesses domínios temáticos, incluindo vários tipos de texto que não são exemplos de novos registros da internet.

3) técnicas de análises de *corpus* são utilizadas para descrever os padrões de variação linguística entre os registros e dentro de cada registro. Muitos são os pesquisadores que se interessam em testar a eficácia da terceira etapa, como por exemplo, a comprovação da precisão e da análise automática de características linguísticas específicas. Os pesquisadores também reconhecem a importância da segunda etapa devido a sua capacidade de fomentar discussões sobre como o *corpus* foi coletado e questionar até que ponto o *corpus* é capaz de representar os registros investigados. Em contrapartida, a primeira etapa é raramente reconhecida. Os pesquisadores passam grande parte do tempo pensando sobre os registros que existem dentro de um determinado tipo de discurso (ou seja, os registros que devem ser incluídos no *corpus*), porém, quase nunca documentam os métodos utilizados durante o processo. Como consequência, essa etapa da pesquisa é geralmente avaliada a partir da validação nominal, pressuposto de que nós, pesquisadores, incluímos no *corpus* os principais registros referentes àquela determinada área.

As pesquisas sobre a linguagem da internet que são baseadas em *corpus* são quase sempre realizadas da mesma forma, primeiramente, identificam-se os registros mais importantes de acordo com suas características mais salientes e depois organiza-se um *corpus* que seja capaz de representar esses registros. Por exemplo, os *blogs* são facilmente associados à web, e a maioria dos estudos baseados em *corpus* que investigam a linguagem da internet considera o blog como sendo um registro da web (HERRING; PAOLILLO, 2006; GRIEVE et al., 2010; TITAK; ROBERSON, 2013). As mensagens de e-mail e as postagens do *Facebook* – apesar de não serem uma página de pesquisa da web – estão intensamente associados com a internet e são frequentemente também incluídas em tais estudos. As reportagens de jornais – originalmente parte da mídia impressa — são também muito associadas à *web* e, por isso, são frequentemente estudadas.

No que tange à linguagem que não pertence à internet (*non-internet language*), não há de fato uma metodologia alternativa para a construção de um *corpus* para o estudo da variação de registro. Não há como especificar ou delimitar a linguagem falada e escrita produzida em sua totalidade em um dia específico (nem mesmo totalizar o discurso escrito de uma determinada época), portanto, não há como estabelecer de forma empírica qual é o conjunto de registros que existem em uma determinada língua, e certamente não há como construir uma amostra aleatória que seja capaz de representar com precisão a proporção de cada registro na totalidade de um idioma.

A *web* é fundamentalmente muito diferente nesse sentido: a população de documentos que podem ser pesquisados na web é finita, pormenorizada e indexada. Assim, pelo menos em teoria, é possível obter uma amostra aleatória de toda a web capaz de representar de forma eficaz e proporcional toda a gama de registros encontrados na web em um determinado momento. Isso é exatamente o que tentamos realizar nos estudos desenvolvidos por Egbert, Biber e Davies (2015) e Biber, Egbert e Davies (2015). Por meio de buscas feitas em língua inglesa no Google em que se buscou pelos trigramas mais frequentes (por exemplo, *is not the, and from the*), conseguimos explorar a estrutura de amostragem indexada da web ao mesmo tempo que minimizamos os vieses de conteúdo que são tipicamente gerados pelas buscas feitas no Google (ver Seção 2.1). O resultado foi um *corpus* de c. 48.000 documentos da web coletados de forma aleatória e capaz de representar um espectro completo de registros da web.

No entanto, essa metodologia para a construção de um *corpus* representativo não identifica por si só o registro de cada documento. Acontece que essa tarefa está longe de ser trivial, levando à iniciativa de Identificação Automática de Registro (Gênero) entre os pesquisadores da WAC (ver KILGARRIFF; GREFENSTETTE 2003; FLETCHER 2012; SANTINI; SHAROFF 2009; CROWSTON; KWASNIK; RUBLESKE 2010; ROSSO; HAAS 2010; SHAROFF; WU HA; MARKERT 2010). Obviamente, o pré-requisito para a identificação automática de registro é ter uma taxonomia do que seriam possíveis categorias de registros encontrados na internet. Contudo, mesmo esse processo mostrou-se não ser tão simples quanto aparenta. Como resultado, cada pesquisador tem utilizado seu próprio conjunto de categorias, que podem variar de forma ampla (ver REHM et al., 2008). Um estudo baseado na *Wikipedia*, desenvolvido de forma colaborativa pelos pesquisadores da WAC, resultou numa lista de setenta e oito registros/gêneros distintos (http://www.webgenrewiki.org/), mas não há consenso sobre as categorias.

Em nossa própria pesquisa anterior (EGBERT; BIBER; DAVIES, 2015; BIBER; EGBERT; DAVIES, 2015) usamos uma abordagem alternativa para lidar com este tipo de problema. Primeiro, baseamos nossa investigação em um *corpus* mais representativo e muito maior do que o utilizado na pesquisa anterior (48.571 documentos da web), obtidos por meio de amostragem aleatória de documentos disponíveis publicamente na web (consulte a Seção 2.1). Em segundo lugar, em vez de depender de codificadores especializados, recrutamos usuários finais da web para realizar nossa codificação de registro, com cada documento codificado por quatro avaliadores diferentes, de modo que poderíamos assim avaliar o grau de concordância entre os usuários. Na terceira e última etapa, a mais fundamental de todas, não forçamos os usuários a escolher diretamente de um conjunto predefinido de categorias de registros específicos. Em vez disso, pedimos aos usuários que identificassem as características situacionais básicas de cada documento da web, codificados de forma hierárquica. Essas características situacionais nos guiaram até às categorias gerais de registro, que por sua vez nos ajudaram a gerar uma lista de sub-registros específicos. Trabalhando por meio de uma árvore de decisão hierárquica, os usuários foram capazes de identificar cada um dos registros da maioria dos documentos da web com um alto grau de confiabilidade.

Nas Seções 2.1 e 2.2, resumimos brevemente os resultados dessa pesquisa anterior, descrevendo as diferentes categorias de registros usadas para estruturar a pesquisa e a distribuição dos textos presentes no *corpus* entre essas categorias. Com essa base, passamos para a nossa pesquisa principal cujo objetivo é realizar uma análise linguística abrangente dos padrões de variação de registros encontrados na web.

2.1. Construindo e codificando o *corpus*

O *corpus* da pesquisa é uma amostra abrangente quase-aleatória de documentos da internet que são públicos e podem ser acessados por meio de pesquisa, ou seja, todo o universo de textos que foram indexados por meio de mecanismos de busca (por exemplo, Google, Bing, Yahoo!). No entanto, para interpretar os resultados com base na análise desse *corpus*, é importante estar atento aos tipos de textos que não estão incluídos. Obviamente, nosso estudo não inclui docu-

mentos privados encontrados na internet, como mensagens de e-mail ou postagens feitas em contas privadas nas mídias sociais (i.e., postagens no *Facebook*).[2] Estes textos não podem ser acessados por meio de mecanismos de busca na web e, portanto, não fazem parte do discurso que investigamos em nosso estudo. É menos óbvio, mas igualmente importante, mencionar que não fazem parte do nosso *corpus* os documentos armazenados na *deep web*, a internet oculta. Como por exemplo, os milhares de documentos informativos protegidos por senhas e armazenados na internet, incluindo a maioria dos artigos de pesquisa referenciados e publicados nas principais revistas acadêmicas, bem como os inúmeros relatórios técnicos e documentos encontrados em páginas institucionais. Acontece que os documentos informativos são muito importantes para a parte pública e pesquisável da web (ver Seção 2.2). No entanto, se fosse possível obter amostras de documentos privados armazenados na *deep web*, a proporção de documentos informativos seria muito maior.

O *corpus* utilizado para o estudo foi extraído do extrato "Geral" do *Corpus of Global Web-based English* (GloWbE; disponível em <http://corpus2.byu.edu/glowbe/>). O *corpus* GloWbE contém c. 1,9 bilhões de palavras em c. 1,8 milhões de documentos da web, coletados entre os meses de novembro e dezembro de 2012, usando os resultados das pesquisas feita no Google, sobre os trigramas mais frequentes em língua inglesa (ou seja, os trigramas mais comuns que ocorrem no COCA, *Corpus of Contemporary American English*, por exemplo, *is not the, and from the*). Cerca de 800 a 1.000 links foram salvos para cada trigrama (ou seja, oitenta a cem páginas de resultados do Google), minimizando assim o viés das preferências já incorporadas às pesquisas feitas no Google. Muitos estudos do tipo WAC anteriores usaram métodos semelhantes com n-gramas que funcionavam como sementes de mecanismo de pesquisa (i.e., BARONI; BERNARDINI, 2004; BARONI et al., 2009; SHAROFF, 2005; 2006). É importante reconhecer que nenhuma pesquisa do Google é verdadeiramente aleatória. Assim, mesmo pesquisas feitas com trigramas de base funcional (por exemplo, *is not the*) são, até certo ponto, processadas com base em escolhas e previsões construídas pelo mecanismo e busca do Google. No entanto, quando selecionamos centenas de documentos para cada um desses n-gramas de bases funcionais em vez de palavras de conteúdo, minimizamos essa influência.

Para criar uma amostra representativa de documentos da web a serem analisados na pesquisa, extraímos aleatoriamente 53.424 URLs do GloWbE *Corpus*. Essa amostra, que compreende documentos da web de cinco regiões geográficas (Estados Unidos, Reino Unido, Canadá, Austrália e Nova Zelândia), representa uma grande amostra de documentos da web coletados a partir de todo o espectro da web pesquisável. Como o objetivo final do projeto é descrever as características léxico-gramaticais de documentos da web (ver Seção 4), qualquer documento com menos de setenta e cinco palavras foi excluído da amostra.

Para criar o *corpus* real de documentos usados para nosso estudo, baixamos os documentos da web associados a esses URLs usando HTTrack (http://www.httrack.com). No entanto, por-

[2] Na primavera de 2015, o *Twitter* chegou a um acordo com o Google para tornar todos os *tweets* públicos pesquisáveis. Entretanto, no momento em que nosso *corpus* foi construído, os *tweets* ainda não haviam sido indexados pelo Google (ou outros mecanismos de busca), e, portanto, não foram incluídos em nossa amostra de documentos da web.

que houve um intervalo de sete meses entre a identificação inicial de URLs e a classificação de registro real dos documentos, c. 8% dos documentos (N = 3713) não estavam mais disponíveis (ou seja, eles estavam vinculados a sites que não existiam mais). Essa alta taxa de desgaste reflete a natureza extremamente dinâmica do universo de textos na web.

Como o objetivo do projeto era realizar análises linguísticas de documentos da web a partir da gama de registros, foram excluídas 1.140 URLs, que consistiam principalmente de fotos e gráficos. Assim, o *corpus* final do projeto foi constituído por 48.571 documentos; o material não textual foi removido de todos os documentos da web (limpeza de HTML e remoção de clichês) usando JusText (http://code.google.com/p/justext).

A presente análise da variação do registro baseia-se em uma amostra de 90% da totalidade desse *corpus*. Essa decisão prevê a próxima etapa deste projeto: desenvolver métodos para identificar automaticamente a categoria de registo de um documento web, com base nas suas características linguísticas (ver Seção 5). Para tanto, dividimos o *corpus* em dois subcorpora: o primeiro utilizado para desenvolver nossos modelos linguísticos (descritos no presente trabalho), e o segundo para testar a precisão preditiva desses modelos. Assim, extraímos aleatoriamente uma amostra de 10% dos documentos do *corpus*[3], que usaremos para futuras pesquisas para avaliação do poder preditivo de nossa descrição linguística. As análises linguísticas aqui relatadas baseiam-se na amostra de 90%, compreendendo 43.685 documentos e c. 52.665.000 palavras.

2.2. Categorias de registros do *corpus*

Como observado na introdução, foi utilizada uma abordagem *bottom-up* para identificar o conjunto de possíveis registros encontrados em nosso *corpus*, bem como a categoria de registro de cada documento individual. Egbert, Biber, Davies (2015) e Biber, Egbert e Davies (2015) fornecem detalhes completos desse procedimento, das categorias de registro resultantes, e da distribuição de registros e sub-registros no *corpus*.

Recrutamos usuários finais típicos da web para a codificação dos registros (através do *Mechanical Turk*[4]), com cada documento no *corpus* codificado por quatro avaliadores diferentes. Os usuários finais identificaram características situacionais básicas de cada documento da web, codificado de forma hierárquica: modo falado ou escrito; único autor ou múltiplos participantes interativos (para documentos escritos); e, finalmente, identificando o propósito comunicativo principal de documentos escritos, não interativos. Essa codificação levou à identificação de vinte categorias principais de registro geral e híbrido, bem como listas de sub-registros específicos pertencentes a essas categorias. Ao trabalhar por meio de uma árvore de decisão hierárquica, os usuários conseguiram identificar a categoria de registro da maioria dos textos da internet com alto grau de confia-

[3] A amostra de 10% é realmente uma amostra aleatória estratificada. Ou seja, escolhemos aleatoriamente 10% dos documentos dentro de cada categoria de registro para garantir a cobertura adequada das categorias menores. O método específico utilizado envolveu a geração de números aleatórios para cada documento em uma determinada categoria de registro e, em seguida, a extração de todos os documentos que tiveram números inferiores a 10% do número total de documentos na categoria. Como resultado, o subcorpus de 10% não constitui exatamente 10% do número total de documentos.

[4] Nota das tradutoras: aplicativo de terceirização pessoal, disponível em < https://www.mturk.com/>

bilidade: 95% dos documentos do *corpus* foram categorizados com sucesso em um entre os vinte registros principais (ou seja, oito registros gerais e doze registros híbridos; ver abaixo).

A Tabela 1 apresenta a divisão de documentos entre as principais categorias de registros. Oito registros gerais emergiram da análise: NARRATIVA; DESCRIÇÃO INFORMACIONAL (ou EXPLICAÇÃO),

TABELA 1. Composição do *Corpus* de acordo com as Categorias de Registros

	nº de documentos com maioria de concordância (i.e., pelo menos 3 dos 4 avaliadores concordaram na codificação)	% do total de documentos no *corpus*	nº de palavras
Registros Gerais			
NARRATIVA	13.688	31,3	13.797.504
DESCRIÇÃO INFORMACIONAL	6338	14,5	8.664.046
OPINIÃO	4936	11,3	7.754.456
DISCUSSÃO INTERATIVA	2835	6,5	2.690.415
COMO-FAZER / INSTRUCIONAL	1030	2,4	1.027.940
PERSUASÃO INFORMACIONAL	751	1,7	684.912
LÍRICO	571	1,3	250.669
FALADO	414	0,9	829.656
Subtotal para todos os registros gerais (i.e., pelo menos 3 dos 4 avaliadores concordaram)	**30.563**	**70,0%**	**35.699.598**
Principais Registros Híbridos com avaliação (2+2)			
NARRATIVA + DESCRIÇÃO INFORMACIONAL	1625	3,7	1.800.500
NARRATIVA + OPINIÃO	1511	3,5	2.166.774
DESCRIÇÃO INFORMACIONAL + OPINIÃO	665	1,5	826.595
PERSUASÃO INFORMACIONAL + DESCRIÇÃO INFORMACIONAL	386	0,9	329.644
COMO-FAZER / INSTRUCIONAL + DESCRIÇÃO INFORMACIONAL	334	0,8	519.370
PERSUASÃO INFORMACIONAL + OPINIÃO	174	0,4	270.396
COMO-FAZER / INSTRUCIONAL + OPINIÃO	152	0,3	140.752
Principais Registros Híbridos com avaliação (2+1+1)			
NARRATIVA + DESCRIÇÃO INFORMACIONAL + OPINIÃO	3186	7,3	3.835.944
PERSUASÃO INFORMACIONAL + DESCRIÇÃO INFORMACIONAL + OPINIÃO	882	2,0	1.010.772
PERSUASÃO INFORMACIONAL + NARRATIVA + OPINIÃO	848	1,9	1.488.240
PERSUASÃO INFORMACIONAL + DESCRIÇÃO INFORMACIONAL + NARRATIVA	686	1,6	646.212
COMO-FAZER / INSTRUCIONAL + DESCRIÇÃO INFORMACIONAL + OPINIÃO	585	1,3	646.425
Todos os outros registros híbridos	2088	4,8	3.284.157
TOTAL	**43.685**	**100,0%**	**52.665.379**

OPINIÃO, DISCUSSÃO INTERATIVA, COMO-FAZER, PERSUASÃO INFORMACIONAL, LÍRICO E DISCURSO FALADO. Setenta por cento dos documentos do *corpus* (ou seja, 30.563 documentos) foram atribuídos a um desses oito registros gerais; esses documentos são os que os quatro avaliadores concordaram ou pelo menos três dos quatro avaliadores concordaram sobre a categorização. NARRATIVA é o mais importante desses registros gerais (c. 31% de todos os documentos do *corpus*), enquanto DESCRIÇÃO INFORMACIONAL e OPINIÃO também são registros predominantes na web. Juntos, 57% de todos os documentos do *corpus* pertencem a um desses três registros gerais.

Os avaliadores concordaram menos em suas categorizações dos outros 13.122 documentos (30%) do *corpus*. No entanto, uma análise mais cuidadosa mostrou que essas discordâncias foram surpreendentemente sistemáticas, e poderiam realmente ser exploradas para identificar documentos com características intermediárias do registro.

Muitos documentos da web foram codificados com uma divisão de opinião de 2-2 ou 2-1-1. Por exemplo, dois avaliadores podem ter codificado uma determinada página como NARRATIVA, enquanto outros dois avaliadores classificaram a mesma página como DESCRIÇÃO INFORMACIONAL. Uma possível interpretação dessas divisões é que elas simplesmente mostram falta de concordância entre os avaliadores, refletindo falta de confiança no quadro de categorização dos registros. No entanto, a distribuição real desses pares sugere uma interpretação diferente. Em teoria, existem vinte e oito combinações diferentes de 2-2 que poderiam ser formadas a partir das oito categorias gerais de registro em nosso quadro. Diante desse potencial, é surpreendente que apenas sete combinações de registros gerais tenha ocorrido comumente em divisões de 2-2 (ver Tabela 1). Da mesma forma, existem dezenas de possíveis combinações 2-1-1, mas apenas cinco dessas ocorrem com alta frequência. Esse conjunto restrito de combinações de registros recorrentes sugere uma explicação alternativa para a falta de concordância entre os avaliadores: ao invés de refletir um problema com a rubrica de codificação, essas combinações comuns de 2-2 e 2-1-1 podem ser interpretadas como evidência de que esses documentos pertencem a registros híbridos — registros que combinam os propósitos comunicativos e outras características situacionais de dois ou mais registros gerais (ver também SANTINI, 2007, 2008; VIDULIN; LUŠTREK; GAMS, 2009). Três dessas combinações híbridas são especialmente importantes: NARRATIVA + DESCRIÇÃO INFORMACIONAL, NARRATIVA + OPINIÃO, e NARRATIVA + DESCRIÇÃO INFORMACIONAL + OPINIÃO. Conjuntamente, essas três combinações representam cerca de 14,5% de todo o *corpus* (ver discussão na Seção 4.4).

Em resumo, 70% dos documentos do *corpus* foram categorizados em um dos oito registros gerais, enquanto outros 25% dos documentos foram categorizados em um dos doze registros híbridos. Em contrapartida, menos de 5% dos documentos do *corpus* são idiossincráticos (2.088 documentos); os avaliadores discordaram da classificação desses documentos, e, portanto, não puderam ser agrupados em nenhuma das categorias principais.

Por fim, após cada usuário ter identificado essas características gerais do registro, pedimos que selecionasse um sub-registro específico a partir de listas de possíveis registros fornecidos na categoria geral. Por exemplo, ENTREVISTAS e TRANSCRIÇÕES DE TV foram possíveis escolhas sob a categoria geral DISCURSO FALADO, enquanto RELATÓRIOS DE NOTÍCIAS e BLOGS de viagem foram possíveis escolhas de registros específicos sob a categoria geral NARRATIVA.

Desenvolvemos as listas de possíveis sub-registros através de dez rodadas de testes piloto, que incluíram quanto os avaliadores foram capazes de reconhecer várias distinções de registro (ver EGBERT; BIBER; DAVIES, 2015). A Tabela 2 resume os principais sub-registros identificados em nosso *corpus*. Os avaliadores conseguiram concordar (pelo menos três dos quatro avaliadores) na categoria de sub-registro de c. 53% de todos os documentos do *corpus* (ou seja, 23.017 do total de 43.685 documentos).

Na maioria dos casos, esses sub-registros podem ser considerados como mais bem definidos do que os registros gerais principais, pois estão associados a propósitos comunicativos mais específicos. Por exemplo, os usuários finais identificaram de forma confiável milhares de documentos como tendo um propósito comunicativo geral de NARRATIVA (ver Tabela 1). No entanto, há uma grande gama de variação situacional dentro dessa categoria geral, incluindo tudo, desde narrativas ficcionais e narrativas pessoais em *blogs* (que são altamente pessoais, envolvidos e "orais") até reportagem de notícias e artigos de pesquisa histórico-acadêmica (que têm um foco mais informativo, menos envolvimento pessoal e, em geral, características "letradas").

A AMD abaixo mostra que os sub-registros também são mais claramente distintos do que os registros gerais em relação às suas características linguísticas. Nem o nível geral, nem o específico de análise deve ser considerado como o "correto", porque os registros podem ser definidos em qualquer nível de especificidade (ver BIBER; CONRAD, 2009, p.10). Assim, nas descrições abaixo, apresentamos resultados tanto para as categorias gerais de registro quanto para as principais categorias de sub-registro.

3. A análise multidimensional de registros da Internet

O presente estudo emprega a AMD a fim de oferecer uma descrição linguística relativamente abrangente da variação de registro na web pesquisável. A AMD foi desenvolvida como uma abordagem metodológica baseada em *corpus* para identificar os padrões de coocorrência linguística subjacentes em um domínio do discurso (em termos empíricos/quantitativos), possibilitando uma comparação de registros, no espaço linguístico definido por esses padrões de coocorrência. Essa abordagem de pesquisa foi primeiramente utilizada em Biber (1985; 1986) e depois desenvolvida mais plenamente em Biber (1988); desde então, a abordagem tem sido aplicada a inúmeros estudos de variação de registro (ver, por exemplo, o levantamento de estudos de MD em BIBER, 2014 e as recentes coleções de estudos de MD em FRIGINAL, 2013 e BERBER-SARDINHA; VEIRANO-PINTo, 2014).

Os procedimentos metodológicos para a AMD foram documentados em várias publicações anteriores (por exemplo, BIBER, 1988; 1995; CONRAD; BIBER, 2001). Em resumo, a noção de coocorrência linguística é dada ao *status* formal na abordagem MD por meio da Análise Fatorial Estatística (ou Análise de Componentes Principais), que identifica quantitativamente os conjuntos de características linguísticas que frequentemente coocorrem em textos. Os fatores são chamados de "dimensões" linguísticas de variação. É ainda possível calcular escores da dimensão

para cada texto, bem como médias de escores da dimensão para cada registro. Os gráficos dessas médias de escores da dimensão permitem a caracterização linguística de qualquer registro, a comparação das relações entre os dois registros e a interpretação funcional mais completa da dimensão subjacente. Por fim, é necessária a análise qualitativa para interpretar as funções comunicativas associadas a cada dimensão, a partir da composição linguística da dimensão junto às semelhanças e diferenças entre os registros em relação à dimensão.

TABELA 2. Principais Sub-Registros no *Corpus*

	nº de documentos (i.e., pelo menos 3 dos 4 avaliadores concordaram na codificação)	% de documentos
NARRATIVA		
REPORTAGEM DE NOTÍCIAS/ BLOG DE NOTÍCIAS	7168	52,4
REPORTAGEM ESPORTIVA	2202	16,1
BLOG PESSOAL	1534	11,2
ARTIGO HISTÓRICO	181	1,3
FICÇÃO	162	1,2
BLOG DE VIAGEM	113	0,8
Sem a maioria de concordância no sub-registro	2328	17,0
TOTAL	13.688	100,0
DESCRIÇÃO INFORMACIONAL (OU EXPLICAÇÃO)		
DESCRIÇÃO DE COISA	1425	22,5
ARTIGO DE ENCICLOPÉDIA	428	6,8
ARTIGO DE PESQUISA	371	5,9
DESCRIÇÃO DE PESSOA	325	5,1
BLOG INFORMACIONAL	303	4,8
PERGUNTAS FREQUENTES (FAQS)	95	1,5
Sem a maioria de concordância no sub-registro	3391	53,5
TOTAL	6338	100,0
OPINIÃO		
CRÍTICA	1024	20,7
BLOG DE OPINIÃO PESSOAL	1857	37,6
BLOG RELIGIOSO/SERMÃO	421	8,5
CONSELHO	289	5,9
Sem a maioria de concordância no sub-registro	1345	27.2
TOTAL	4936	100,0
DISCUSSÃO INTERATIVA		
FÓRUM DE DISCUSSÃO	1626	57,4
FÓRUM DE PERGUNTA E RESPOSTA	821	29,0
Sem a maioria de concordância no sub-registro	388	13,7
TOTAL	2835	100,0

(continua)

TABELA 2. (Continuação)

	nº de documentos (i.e., pelo menos 3 dos 4 avaliadores concordaram na codificação)	% de documentos
COMO-FAZER		
COMO-FAZER/INSTRUCIONAL	757	73,5
RECEITA	110	10,7
Sem a maioria de concordância no sub-registro	163	15,8
TOTAL	1030	100,0
PERSUASÃO INFORMACIONAL		
DESCRIÇÃO COM INTENÇÃO DE VENDER	622	
EDITORIAL (incluindo BLOGS DE NOTÍCIAS + OPINIÃO)[a]	380	
Sem a maioria de concordância no sub-registro	109	
TOTAL	1111	100,0
LÍRICO		
MÚSICAS	476	83,4
POEMAS	45	7,9
Sem a maioria de concordância no sub-registro	50	8,8
TOTAL	571	100,0
FALADO		
ENTREVISTA	251	60,6
DISCURSO FORMAL	20	4,8
TRANSCRIÇÃO DE TV	11	2,7
Sem a maioria de concordância no sub-registro	132	31,9
TOTAL	414	100,0

[a] Uma das combinações híbridas frequentes dos sub-registros em nosso *corpus* é REPORTAGEM DE NOTÍCIAS/BLOG DE NOTÍCIAS + BLOG DE OPINIÃO PESSOAL. Após a leitura de diversos desses documentos, decidimos que eles possuem propósitos similares aos EDITORIAIS. Portanto, combinamos essas categorias de sub-registros para a AMD. Dessa forma, o resultado "total" para os documentos de PERSUASÃO INFORMACIONAL na Tabela 2 é maior que o total na Tabela 1.

Para o presente estudo, iniciamos com as mais de 150 características lexico-gramaticais específicas identificadas pelo etiquetador Biber Tagger (ver, por exemplo, BIBER et al., 1999). Há considerável sobreposição entre algumas dessas características, pois as características lexico-gramaticais podem ser medidas em muitos níveis diferentes de especificidade. Por exemplo, o etiquetador inclui a análise de características lexico-gramaticais específicas (por exemplo, verbos mentais que controlam uma oração em língua inglesa do tipo *that*-complemento), bem como construções sintáticas gerais (por exemplo, orações finitas complementares). Além disso, redundâncias foram eliminadas combinando algumas variáveis e abandonando variáveis que apresentaram baixas frequências no total. As variáveis com baixas comunalidades na análise fatorial preliminar (refletindo baixa variância compartilhada com a estrutura do fator) foram eliminadas. Cinquenta e sete variáveis linguísticas foram mantidas para a análise final (ver Apêndice1). Para descrições dessas características linguísticas individuais, ler Biber et al. (1999) e Biber (2006). A Análise de Componentes Principais (usando o procedimento FATOR

no SAS, com METHOD=PRIN) foi utilizada para extrair os fatores.[5] O número de dez fatores foi selecionado como ideal. A decisão de análise foi baseada na inspeção do resultado do gráfico de sedimentação (*scree-plot*) e autovalores (mantendo a maioria dos fatores com autovalores > 1,0), e na interpretabilidade dos fatores extraídos em diferentes soluções. No entanto, um desses fatores não foi prontamente interpretável, de modo que a discussão abaixo foca nos nove fatores restantes.[6] A solução fatorial representou 42,7% da variância compartilhada cumulativa.[7] Os fatores foram gerados usando uma rotação Promax, que resultou em correlações, em geral, pequenas entre as dimensões (Dimensões 1 e 2: r = .32; Dimensões 2 e 3: r = -.39; Dimensões 1 e 5: r = .42; Dimensões 2 e 5: r = .36; todas as outras correlações entre dimensões foram < .25).

O procedimento FATORIAL no SAS (especificando a opção "Score") foi usado para calcular os escores dos fatores para cada documento. Em seguida, calculamos as médias de escores de fatores (chamados de "escores de dimensão") e os desvios padrão para cada registro. A Tabela 3 apresenta os resultados de uma ANOVA (a partir do procedimento GLM no SAS), mostrando que todas as nove dimensões são capazes de prever significativamente a variação de registro. Os valores para R2 na Tabela 3 fornecem uma medida direta de importância, indicando a porcentagem da variância entre os escores das dimensões que podem ser previstos por reconhecerem as categorias do registro. As primeiras cinco dimensões são moderadamente fortes preditoras, enquanto as dimensões 6-9 têm valores R2 mais fracos. Essa análise estatística tenta prever as diferenças entre as vinte categorias de registro gerais e híbridos (e mais uma categoria "outra"),

[5] Análise de Componentes Principais e Análise Exploratória de Fatores são procedimentos estatísticos relacionados (ambas as opções encontram-se sob o procedimento FATOR no SAS). Optou-se por uma Análise de Componentes Principais no presente caso, pois é mais apropriado para grandes estudos, reduzindo um conjunto de variáveis altamente correlacionadas a um pequeno número de parâmetros subjacentes, explicando a quantidade máxima de variância nos dados. Recomenda-se a leitura de documentação online para o procedimento FATOR no programa SAS em https://support.sas.com/documentation/ cdl/en/statug/63347/HTML/default/viewer.htm#factor_toc.htm

[6] Em nossa discussão aqui (e na Tabela 4), reordenamos os fatores para agrupar dimensões que têm interpretações funcionais relacionadas. Assim, as Dimensões 1, 2 e 3 da Tabela 3 estão relacionadas à distinção oral-letrado; as Dimensões 3, 4 e 5 têm certa relação com os registros narrativos; e as Dimensões 6-9 refletem funções mais específicas do discurso.

Nova quantidade de Dimensões (usadas neste artigo)	Quantidade original de fatores
1	1
2	4
3 (invertido)	3
4	7
5	2
6	9
7	6
8	8
9	10
-- excluído --	5

O Fator 5 foi originalmente retirado da discussão aqui relatada, porque o fator não se apresentou facilmente interpretável. As características linguísticas primárias de carregamento no Fator 5 original são advérbios gerais, advérbios de ligação, advérbios de posicionamento, razão forma/palavra, coordenação e pronomes demonstrativos. Documentos representantes de DISCURSO FALADO e OPINIÃO estão marcados pela presença dessas características coocorrentes; documentos representantes de DESCRIÇÃO INFORMACIONAL e de LÍRICO estão marcados pela ausência dessas características.

[7] Esse percentual é semelhante às taxas de outras análises fatoriais de registros; por exemplo, a solução de sete fatores em Biber (1988) representou 51,9% de variância compartilhada, enquanto a solução de quatro fatores em Biber, Egbert e Davies (2015) representou 44% de variância compartilhada.

TABELA 3. Dimensões como Preditivas da Variação de Registro
(20 categorias de Registros Gerais e Registros Híbridos + Outras)

Dimensão	Valor (escore) de F	Valor (escore) de *p*	% de variância explicada (R^2)
1	387.1	<.0001	15.1
2	278.8	<.0001	11.3
3	474.4	<.0001	17.9
4	722.6	<.0001	24.9
5	472.2	<.0001	17.8
6	166.2	<.0001	07.1
7	84.6	<.0001	03.7
8	61.7	<.0001	02.7
9	142.0	<.0001	06.1

por isso, não é surpresa que algumas das dimensões tenham relações relativamente fracas para a gama completa de registros. Em vários casos, no entanto, as dimensões servem para identificar características distintas de dois ou três registros específicos, embora não façam boa distinção entre os registros restantes; tais parâmetros linguísticos ainda são importantes para nosso propósito de entender o conjunto completo de diferenças linguísticas entre os registros da web. Além disso, como mostra a Seção 4, várias dimensões são fortes preditoras de diferenças entre sub-registros específicos, enquanto as categorias gerais de registro são menos distintas.

A Tabela 4 resume todos os principais resultados da análise MD. A Tabela 4 está organizada em cinco colunas:

- A Coluna 1 resume a interpretação funcional da dimensão.
- A Coluna 2 apresenta a lista de características linguísticas recorrentes importantes que compõem a dimensão (i.e., todas as características com cargas > ±.3; características com cargas > ±.2 estão entre parênteses).
- A Coluna 3 resume os escores da dimensão para cada um dos 8 registros gerais.
- A Coluna 4 resume os escores da dimensão para os sub-registros selecionados para ilustrar a gama de variação em relação aos registros gerais.
- A Coluna 5 contem os símbolos + ou –, para indicar o nível do escore de registros e sub-registros nas Colunas 3 e 4. Sete níveis foram encontrados para os escores da dimensão:

  > ± 1.5 representados por + + + + + + / – – – – – –
  > ± 1.2 representados por + + + + + / – – – – –
  > ± 0.9 representados por + + + + / – – – –
  > ± 0.6 representados por + + + / – – –
  > ± 0.3 representados por + + / – –
  > ± 0.15 representados por + / –
  Registros não marcados (com escores com valores próximos a 0.0) representados por 0

As características linguísticas com cargas positivas e negativas relevantes em cada dimensão estão listadas na coluna 2. Cada dimensão pode ter características "positivas" e "negativas". Em vez de refletir a importância, os sinais de positivo e de negativo identificam dois agrupamentos de características que ocorrem em um padrão complementar como parte da mesma dimensão. Ou seja, quando as características positivas ocorrem juntas com frequência em um texto, as características negativas são marcadamente menos frequentes nesse texto, e vice-versa. Para auxiliar na interpretação de cada dimensão, a coluna 2 agrupa características que são estruturalmente/semanticamente/funcionalmente semelhantes. No entanto, é importante enfatizar que esses são agrupamentos *post-hoc* que não tiveram influência na identificação dos padrões de coocorrência linguística pela análise fatorial.

A coluna 3 apresenta, em forma de um semi-gráfico, os oito registros gerais, indicando o escore da dimensão para cada registro. Os registros com escores positivos na dimensão estão listados no topo; os registros com escores negativos altos na dimensão estão listados no final. Esses escores das dimensões correspondem ao uso das características linguísticas listadas na coluna 2. Por exemplo, a Tabela 4 mostra que os documentos do registro LÍRICO têm escore positivo alto para a Dimensão 1, e que o escore corresponde ao uso frequente de características positivas da Dimensão 1 (e.g., verbos, pronomes, características de posicionamento), aliado a um uso infrequente de características negativas da Dimensão 1 (e.g., artigos definidos, frases preposicionais). DESCRIÇÃO INFORMACIONAL tem um grande escore negativo para a Dimensão 1, correspondendo às características linguísticas opostas: poucas características positivas da Dimensão 1, mas o uso frequente de características negativas da Dimensão 1.

A coluna 4 lista sub-registros específicos que são especialmente marcados para uma determinada dimensão, muitas vezes contrastando vários sub-registros dentro de uma única categoria geral. O formato da coluna 4 é o mesmo da coluna 3, com sub-registros contendo escores positivos altos na dimensão listados no topo e os sub-registros tendo escores negativos altos na dimensão com listados no final.

Por fim, a coluna 5 apresenta na forma de um semi-gráfico a magnitude do escore da dimensão para cada um dos registros e sub-registros listados nas colunas 3 e 4. Foi utilizado o formato de semi-gráfico porque os registros na coluna 3 alinham-se verticalmente com o Nível do Escore da Dimensão na coluna 5. Os escores da dimensão positiva refletem o uso frequente das características positivas em uma dimensão, juntamente com o raro uso das características negativas nessa dimensão. Os escores negativos da dimensão refletem características opostas. No entanto, é possível que um registro tenha escore negativo na dimensão, mesmo quando não há características negativas na dimensão, ou seja, no caso em que um registro raramente apresenta as características positivas nessa dimensão. Por exemplo, os documentos LÍRICO e os TRANSCRIÇÕES DE TV têm escores negativos elevados na Dimensão 7, porque os recursos positivos que definem essa dimensão (características da sintagmas nominais de posicionamento) são extremamente raros nesses registros.

DOI: 10.12957/matraga.2021.59614

Revista do Programa de Pós-Graduação em Letras da UERJ

TABELA 4. Resumo dos resultados da Análise Multidimensional

Interpretação da Dimensão	Características Linguísticas Coocorrentes na dimensão	Resumo dos escores da dimensão para os registros gerais	Resumo dos escores da dimensão para os sub-registros selecionados	Nível de escore da Dimensão
DIMENSÃO 1				
Oral-envolvido vs. Letrado-Informacional	Características Positivas (+) Verbos: aspecto progressivo, verbos no tempo não passado, verbos de ação Pronomes: Pronomes de 1ª. pessoa, Pronomes de 2ª. pessoa Características de posicionamento: verbos de desejo + oração com "para", verbos mentais, adjetivos atitudinais (sem controlar uma oração complementar); advérbios de posicionamento) (razão forma/palavra, verbo + oração relativa QU)	LÍRICO DISCURSO FALADO DISCUSSÃO INTERATIVA COMO-FAZER OPINIÃO	OPINIÃO: *CONSELHO* FALADO: *TV* LÍRICO: *MÚSICA* FALADO: *ENTREVISTA* NARRATIVA: *BLOG PESSOAL*	+ +
VERSUS				
		PERSUASÃO INFORMACIONAL	NARRATIVA: *FICÇÃO* LÍRICO: *POEMA* INFORMACIONAL: *BLOG DEINFORMAÇÃO* FALADO: *DISCURSO FORMAL*	0 0 0 0
	CARACTERÍSTICAS NEGATIVAS (-) Artigos definidos, sintagmas preposicionais, orações relativas passivas não finitas, (substantivos concretos)	NARRATIVA DESCRIÇÃO INFORMACIONAL	OPINIÃO: *NOTÍCIAS+BLOG DE OPINIÃO* OPINIÃO: *BLOG RELIGIOSO* INFORMACIONAL: *ARTIGOS DE PESQUISA; ARTIGOS DE ENCICLOPÉDIA* *NARRATIVA: ARTIGO HISTÓRICO*	- -

(continua)

TABELA 4. (Continuação)

Interpretação da Dimensão	Características Linguísticas Coocorrentes na dimensão	Resumo dos escores da dimensão para os registros gerais	Resumo dos escores da dimensão para os sub-registros selecionados	Nível de escore da Dimensão
DIMENSÃO 2				
Elaboração oral	CARACTERÍSTICAS POSITIVAS (+) Verbos: verbos existenciais, verbos mentais, verbos epistêmicos (sem controlar uma oração complementar) Oração verbal + complemento: verbos de probabilidade + orações relativas (com '*that*'), verbos de certeza + orações relativas (com '*that*') (verbo + oração com para/to (excluindo verbos de desejo), verbo + oração relativa - QU Apagamento de '*that*' (pronomes de 1ª. pessoa)	LÍRICO DISCUSSÃO INTERATIVA DISCURSO FALADO OPINIÃO	LÍRICO: *MÚSICAS* NARRATIVA: FICÇÃO; BLOG PESSOAL LÍRICO: *POEMA*	+ +
VERSUS				0
	CARACTERÍSTICAS NEGATIVAS (-) (verbos de ação; substantivos próprios' razão forma/palavra)	PERSUASÃO INFORMACIONAL COMO-FAZER DESCRIÇÃO INFORMACIONAL NARRATIVA	NARRATIVA: *REPORTAGEM DE NOTÍCIA, ARTIGO HISTÓRICO* COMO-FAZER: *RECEITA*	0 - - - - - - - -

(continua)

ISSN eletrônico 2446-6905

TABELA 4. (Continuação)

Interpretação da Dimensão	Características Linguísticas Coocorrentes na dimensão	Resumo dos escores da dimensão para os registros gerais	Resumo dos escores da dimensão para os sub-registros selecionados	Nível de escore da Dimensão
DIMENSÃO 3				
Narrativa oral vs. Informação escrita	CARACTERÍSTICAS POSITIVAS (+) Verbos: verbos no passado, verbos no aspecto perfeito, verbos de ação Advérbios: de tempo, de lugar, (o total de outros advérbios), sintagmas adverbiais (excluindo condicional) Pronomes: de 1ª. pessoa (3ª. pessoa "it") (características de sintagma nominal: artigos definidos, substantivos concretos)	LÍRICO DISCUSSÃO INTERATIVA DISCURSO FALADO NARRATIVA	NARRATIVA: *FICÇÃO* LÍRICO: *MÚSICA* FALADO: *TV* NARRATIVA: *BLOG PESSOAL* NARRATIVA: *REPORTAGEM ESPORTIVA* NARRATIVA: *BLOG DE VIAGEM* LÍRICO: *POEMA*	+ +
VERSUS				0
	CARACTERÍSTICAS NEGATIVAS (-) Palavras longas Substantivos: comuns, processuais Modificadores nominais: adjetivos atributivos; substantivos pré-modificadores (orações relativas finitas)	COMO-FAZER OPINIÃO PERSUASÃO INFORMACIONAL DESCRIÇÃO INFORMACIONAL	INFORMACIONAL: *DESCRIÇÃO DE PESSOA* NARRATIVA: *REPORTAGEM DE NOTÍCIA* FALADO: *DISCURSO FORMAL* INFORMACIONAL: *ARTIGO DE PESQUISA*	0 - - - - - - - - - - - - - - - -

(continua)

DOI: 10.12957/matraga.2021.59614

TABELA 4. (Continuação)

Interpretação da Dimensão	Características Linguísticas Coocorrentes na dimensão	Resumo dos escores da dimensão para os registros gerais	Resumo dos escores da dimensão para os sub-registros selecionados	Nível de escore da Dimensão
DIMENSÃO 4				
Comunicação reportada	CARACTERÍSTICAS POSITIVAS (+) Verbos de comunicação + orações relativas (com '*that*'), verbos de ligação sem controlar uma oração complementar Apagamento de *that* Verbos de probabilidade + oração relativa (com '*that*') (verbos no aspecto perfeito, no tempo passado, modais de predição, substantivos comuns)	NARRATIVA DISCURSO FALADO	NARRATIVA: *REPORTAGEM DE NOTÍCIAS* NARRATIVA: *FICÇÃO*	+ + + + + + + +
VERSUS				
	CARACTERÍSTICAS NEGATIVAS (-) Nominalizações (pronomes de 2ª pessoa)	DISCUSSÃO INTERATIVA OPINIÃO DESCRIÇÃO INFORMACIONAL COMO-FAZER PERSUASÃO INFORMACIONAL	NARRATIVA: *ARTIGO HISTÓRICO*	0 0 0 - - - - - - - - - - - -

(continua)

ISSN eletrônico 2446-6905

TABELA 4. (Continuação)

Interpretação da Dimensão	Características Linguísticas Coocorrentes na dimensão	Resumo dos escores da dimensão para os registros gerais	Resumo dos escores da dimensão para os sub-registros selecionados	Nível de escore da Dimensão
DIMENSÃO 5				
Irrealis vs. Narração Informacional	CARACTERÍSTICAS POSITIVAS (+) Verbos modais: modais de predição, necessidade, possibilidade Orações adverbiais condicionais Verbos: de ligação 'BE', verbos não no passado Pronomes de 2ª. pessoa, adjetivos epistêmicos (sem controlar uma oração complementar), pronomes demonstrativos, coordenação	LÍRICO COMO-FAZER DISCUSSÃO INTERATIVA DISCURSO FALADO OPINIÃO	LÍRICO: *MÚSICA* FALADO: *TV* OPINIÃO: *CONSELHO* INFORMAÇÃO: *PERGUNTAS FREQUENTES (FAQS)* OPINIÃO: *BLOG RELIGIOSO*	+ +
VERSUS				0
	CARACTERÍSTICAS NEGATIVAS (-) Verbos no tempo passado (verbos no aspecto progressivo; substantivos próprios, adjetivos atributivos; razão forma/palavra; palavras longas)	DESCRIÇÃO INFORMACIONAL NARRATIVA	FALADO: *DISCURSO FORMAL* OPINIÃO: *CRÍTICA* NARRATIVA: *REPORTAGEM DE NOTÍCIA; FICÇÃO* INFORMAÇÃO: *ARTIGO DE ENCICLOPÉDIA; ARTIGO DE PESQUISA* NARRATIVA: *BLOG DE VIAGEM* INFORMAÇÃO: *DESCREVER UMA PESSOA* NARRATIVA: *ARTIGO HISTÓRICO*	- -

(continua)

DOI: 10.12957/matraga.2021.59614

TABELA 4. (Continuação)

Interpretação da Dimensão	Características Linguísticas Coocorrentes na dimensão	Resumo dos escores da dimensão para os registros gerais	Resumo dos escores da dimensão para os sub-registros selecionados	Nível de escore da Dimensão
DIMENSÃO 6				
Discurso Procedural/ explanatório	CARACTERÍSTICAS POSITIVAS (+) Verbos causativos/facilitadores, verbos no aspecto progressivo, verbos + orações relativas com '*that*' (excluindo verbos de desejo) Substantivos processuais (verbos de ação, modais de possibilidade, palavras longas, advérbios de ligação, verbos de comunicação sem controlar uma oração complementar)	COMO-FAZER DESCRIÇÃO INFORMACIONAL	INFORMAÇÃO: *ARTIGO DE PESQUISA; PERGUNTAS FREQUENTES* OPINIÃO: *CONSELHO*	+ + + + + + + + + + + + +
VERSUS				0 0
	CARACTERÍSTICAS NEGATIVAS (-) Substantivos próprios (artigos indefinidos, contrações)	PERSUASÃO INFORMACIONAL NARRATIVA OPINIÃO DISCUSSÃO INTERATIVA DISCURSO FALADO LÍRICO	INFORMAÇÃO: *DESCREVER UMA PESSOA* OPINIÃO: *CRÍTICA*	- - - - - - - - - - - - - -

(continua)

ISSN eletrônico 2446-6905

TABELA 4. (Continuação)

Interpretação da Dimensão	Características Linguísticas Coocorrentes na dimensão	Resumo dos escores da dimensão para os registros gerais	Resumo dos escores da dimensão para os sub-registros selecionados	Nível de escore da Dimensão
DIMENSÃO 7				
Posicionamento Letrado	CARACTERÍSTICAS POSITIVAS (+) Características de sintagmas nominais de posicionamento: substantivos cognitivos; substantivo de posicionamento + sintagma preposicional, substantivos de posicionamento + oração complementar, outros substantivos de posicionamento (verbos epistêmicos sem controlar uma oração complementar, substantivos processuais, verbos de comunicação sem controlar uma oração complementar)	OPINIÃO DESCRIÇÃO INFORMACIONAL	INFORMAÇÃO: *ARTIGO DE PESQUISA* FALADO: *DISCURSO FORMAL*	+ + + + + + + + + + + + + + +
VERSUS				
		NARRATIVA DISCURSO FALADO PERSUASÃO INFORMACIONAL COMO-FAZER DISCUSSÃO INTERATIVA LÍRICO	 INFORMAÇÃO: *ARTIGO DE ENCICLOPÉDIA; DESCRIÇÃO DE COISA; PERGUNTAS FREQUENTES* FALADO: *TV*	0 0 - - - - - - -

(continua)

DOI: 10.12957/matraga.2021.59614

TABELA 4. (Continuação)

Interpretação da Dimensão	Características Linguísticas Coocorrentes na dimensão	Resumo dos escores da dimensão para os registros gerais	Resumo dos escores da dimensão para os sub-registros selecionados	Nível de escore da Dimensão
DIMENSÃO 8				
Descrição de humanos	CARACTERÍSTICAS POSITIVAS (+) Pronomes de 3ª. pessoa Orações relativas finitas Artigos indefinidos (verbos de comunicação + oração relativa com *'that'*, verbos de comunicação sem controlar uma oração complementar, verbos no tempo passado)	PERSUASÃO INFORMACIONAL NARRATIVA	NARRATIVA: *FICÇÃO* INFORMAÇÃO: *DESCREVER UMA PESSOA; ARTIGO DE ENCICLOPÉDIA* OPINIÃO: *BLOG RELIGIOSO* LÍRICO: *POEMA*	+ + + + + + + + + + + + + + + + + + +
VERSUS		DISCURSO FALADO OPINIÃO DESCRIÇÃO INFORMACIONAL DISCUSSÃO INTERATIVA LÍRICO COMO-FAZER	 LÍRICO: *MÚSICA* NARRATIVA: *BLOG DE VIAGEM* COMO-FAZER: *RECEITA*	0 0 0 - - - - - - - - -

(continua)

ISSN eletrônico 2446-6905

DOI: 10.12957/matraga.2021.59614

TABELA 4. (Continuação)

Interpretação da Dimensão	Características Linguísticas Coocorrentes na dimensão	Resumo dos escores da dimensão para os registros gerais	Resumo dos escores da dimensão para os sub-registros selecionados	Nível de escore da Dimensão
DIMENSÃO 9				
Descrição não-técnica	CARACTERÍSTICAS POSITIVAS (+) Substantivos concretos, substantivos comuns Artigos indefinidos (substantivos pré-modificadores)	COMO-FAZER DISCUSSÃO INTERATIVA	COMO-FAZER: *RECEITA* INFORMAÇÃO: *BLOG DE INFORMAÇÃO, PERGUNTAS FREQUENTES* OPINIÃO: *CONSELHO, CRÍTICA*	+ + + + + + + + + + + + + + + +
VERSUS		OPINIÃO DESCRIÇÃO INFORMACIONAL		0 0
	CARACTERÍSTICAS NEGATIVAS (-) Nominalizações	NARRATIVA DISCURSO FALADO PERSUASÃO INFORMACIONAL LÍRICO	INFORMAÇÃO: *ARTIGO DE PESQUISA* OPINIÃO: *BLOG DE OPINIÃO* OPINIÃO: *NOTÍCIAS + BLOG DE OPINIÃO*	- - - - - - - - - - -

Legenda para os níveis dos escores das dimensões:

> ± 1.5 representado por + + + + + + / – – – – – –

> ± 1.2 representado por + + + + + / – – – – –

> ± 0.9 representado por + + + + / – – – –

> ± 0.6 representado por + + + / – – –

> ± 0.3 representado por + + / – –

> ± 0.15 representado por + / –

Registros não-marcados (com escores próximos de 0.0) representados por 0.

4. Padrões multidimensionais de variação de registro na Internet

4.1. Dimensões oral-letrada

A análise dos escores da dimensão para os registros gerais (coluna 3 na Tabela 4) indica que as dimensões 1-3 são semelhantes em diferenciar-se entre os registros orais e os registros letrados na Internet. Os textos da categoria LÍRICO são os mais marcados em todas as três dimensões, enquanto os textos presentes nas categorias FALADOS e DISCUSSÃO INTERATIVA também têm escores positivos altos nas três dimensões. Os registros escritos têm escores negativos em todas as três dimensões.

Além dessas similaridades, existem algumas diferenças menos perceptíveis nos padrões de registro. O registro DESCRIÇÃO INFORMACIONAL tem escore negativo na Dimensão 1, refletindo ausência de características positivas nessa dimensão, simultaneamente, à presença de altas frequências para as características negativas da Dimensão 1. A categoria NARRATIVA tem escore negativo na Dimensão 1, enquanto as categorias COMO-FAZER e OPINIÃO têm escores positivos entre baixos e moderados na Dimensão 1. DESCRIÇÃO INFORMACIONAL e NARRATIVA também são os únicos registros com escores negativos na Dimensão 2, embora a magnitude desses escores seja pequena. Uma das principais diferenças entre as duas dimensões é que a Dimensão 1 define uma clara oposição entre dois estilos de discurso — oral-envolvido versus letrado-informacional — enquanto a Dimensão 2 identifica principalmente um único estilo de discurso: o estilo de elaboração encontrado na oralidade (o qual é bastante ausente nos registros DESCRIÇÃO INFORMACIONAL e NARRATIVA). Por fim, a Dimensão 3 combina duas influências funcionais que se contrastam: um contraste entre oral e letrado e um contraste entre narrativo e não narrativo. Aparentemente, todos os registros gerais orais em nosso *corpus* tendem a ter propósitos comunicativos narrativos, assim como os documentos narrativos escritos. Na outra extremidade, tanto DESCRIÇÃO INFORMACIONAL quanto PERSUASÃO INFORMACIONAL estão marcados pela ausência de características positivas da Dimensão 3, simultaneamente com altas frequências das características negativas nessa dimensão.

Apesar das semelhanças nos padrões de registro, essas três dimensões são claramente distintas em sua composição linguística. A Dimensão 1 é composta por verbos de ação dinâmica, frases com verbo no aspecto progressivo (gerúndio) e verbos no tempo presente, associados a pronomes e características de posicionamento. Já a Dimensão 2 é composta por classes de verbos estáticos e construções de orações complementares. A Dimensão 3 é composta por verbos no tempo passado e verbos no tempo perfeito (aspecto perfeito), ao mesmo tempo que por advérbios e pronomes (em oposição a palavras longas e características associadas a sintagmas nominais complexos).

Essas diferenças linguísticas refletem os diferentes fundamentos funcionais das dimensões. Assim, as características positivas na Dimensão 1 são características estereotipicamente orais, mas também refletem um alto grau de interatividade e envolvimento pessoal. Tais características são comuns não apenas ao DISCURSO FALADO e à DISCUSSÃO INTERATIVA escrita; também são comuns em documentos escritos do tipo COMO-FAZER, que refletem alto envolvimento pessoal.

A Tabela 4 mostra que essas características de envolvimento pessoal são raras no registro descrição informacional (que tem escore negativo na Dimensão 1). Em vez disso, encontramos uso frequente das características negativas da Dimensão 1 nesses registros: frases com artigos definidos, sintagmas preposicionais e orações relativas passivas não finitas. Essas características são usadas para transmitir informações, mas também refletem um tipo de "distanciamento" impessoal em oposição às funções "envolvidas" e dinâmicas das características positivas.

O padrão de variação da Dimensão 1 para os registros gerais define uma oposição fundamental entre os registros falados/orais/envolvidos e os registros escritos/letrados/informacionais. Por exemplo, os exemplos (1)-(3) ilustram o denso uso de características positivas da Dimensão 1 "orais" em uma música (1), uma transcrição de tv (2) (categoria que inclui o diálogo de filmes e de televisão), e um fórum de discussão (3), que, embora seja escrito é interativo. Em cada um dos exemplos, marcamos características linguísticas importantes referentes à dimensão em itálico, em negrito e sublinhado.

(1) Lírico: Música[8]

(**Negrito marca os verbos que não estão no tempo passado**; *itálico marca os verbos no aspecto progressivo finito (gerúndio) e orações não finitas no gerúndio*; sublinhado marca pronomes de 1a. e 2a. pessoas)

Neither of us ever thought, back on the first day that we met
We were both so convinced that the dream would never end
Nobody saw the signs man, it'**s** funny how love **is** blind man
'Cause it can **be** over just as fast as it **begins**
So now I be ***sitting*** here, I'**m** alone ***trying*** to **stuff** my fear
Checkin' out one of the pictures there up on the shelf
The story behind my song, **is** that you never **know** what you **got** 'til it**'s** gone
(fonte: http://lyrics.wikia.com/Scatman_John:Sorry_Seems_To_Be_The_Hardest_Word>)

(2) Discurso Falado: Transcrição de TV

(**Negrito marca verbos que não estão no passado;** *itálico marca verbos de posicionamento*; sublinhado marca pronomes de primeira e segunda pessoas)

V : Fortunately, I got to you before they did.
Evey Hammond : You got to me? You did this to me? You cut my hair? You tortured me? You tortured me! Why?
V : You *said* you *wanted* to **live** without fear. I ***wish*** there'd been an easier way, but there wasn't. I ***know*** you may never **forgive** me . . . but nor will you ***understand*** how hard it was for me to **do** what I did. Every day I saw in myself everything you **see** in me now. Every day I *wanted* to **end** it, but each time **you** *refused* to **give** in, I knew I couldn't. [. . .] See, at first

[8] Nota das tradutoras: as traduções de todos os exemplos usados no artigo estão no apêndice 2.

I *thought* it was hate, too. Hate was all I *knew*, it built *my* world, it imprisoned me, taught me how to **eat**, how to **drink**, how to **breathe**.
[. . .]
V: So if you**'ve** seen nothing, if the crimes of this government **remain** unknown to you then I would ***suggest*** you ***allow*** the fifth of November to **pass** unmarked. But if you **see** what I **see**, if you ***feel*** as I ***feel***, and if you would ***seek*** as I ***seek***, then I ***ask*** you to **stand** beside me one year from tonight, outside the gates of Parliament, and together we shall **give** them a fifth of November that shall never, ever be ***forgot***. (fonte: http://www.imdb.com/character/ch0002908/quotes)

(3) Discussão Interativa: Fórum

(**Negrito marca os verbos que não estão no tempo passado**; *itálico marca os verbos no aspecto progressivo finito (gerúndio) e orações não finitas no gerúndio*; sublinhado marca pronomes de 1a. e 2a. pessoas)

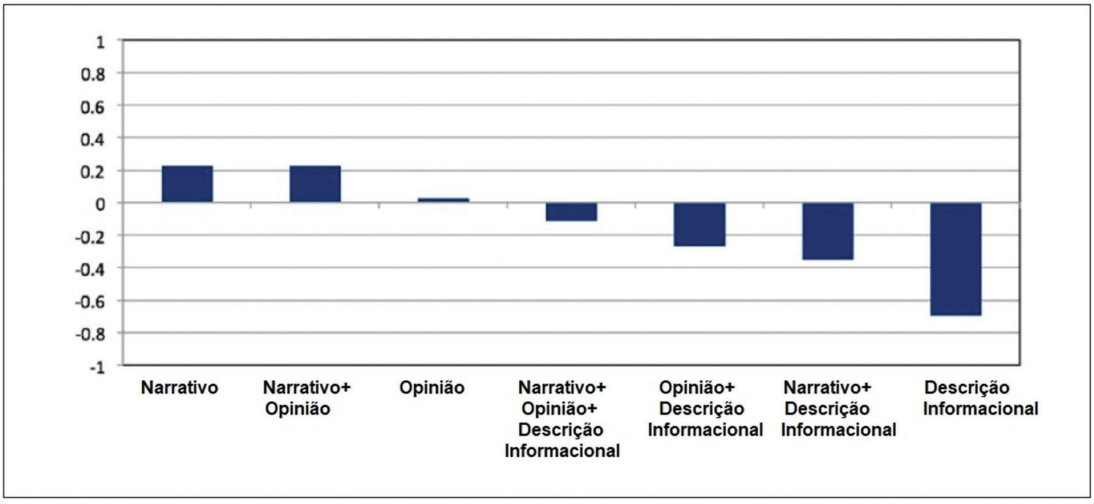
It also **depends** where your [sic] ***downloading*** from. sometimes its not VM's fault. Also maybe you ***getting*** traffic **managed**, have you considered this?

If you signed up for a 12 month contract with virgin media broadband and your out of the initial 30 day period I dont **think** you can simply "**cancel**". You can still **get** BT **installed** but you will be **required** to **keep** ***paying*** for VM.

If someone's helped you out **say** thanks by ***clicking*** on the Kudos Star. If someone's solved your problem, why not **mark** their message as na Accepted Solution

I was only ***trying*** to **make** you aware that if you signed up for BT and a 12 month contract that if you are still under a 12 month contract with virgin media it would not **be** that easy to **cancel**. But if you want to **act** like an immature child then **be** my guest.

FIGURA 1. Escores da Dimensão 3 para Registros Simples e Híbridos: Narrativa Oral versus Informação Escrita



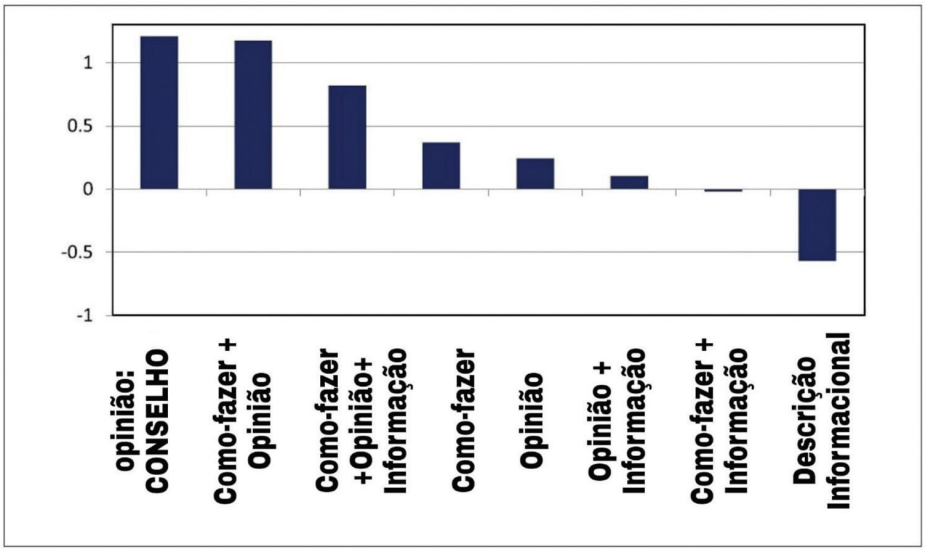
Em contraste, o exemplo (4) ilustra o uso de características negativas da Dimensão 1 "letrado" (e a ausência relativa de características positivas da Dimensão 1) em um ARTIGO DE PESQUISA.

(4) DESCRIÇÃO INFORMACIONAL: ARTIGO DE PESQUISA
(**Negrito marca preposições**; *itálico marca artigos definidos*; sublinhado marca orações relativas passivas não-finitas)

Follow up experiments used macrophage isolated **from** *the* bone marrow (BM) **of** wildtype (WT) mice. *The* data **from** WB mushroom extracts is shown since *the* highest TNF secretion **from** RAW 264.7 cells was found **in** *the* WB stimulated cultures (Figure 1). The BMDM cells are untransformed mouse macrophage that produce a variety **of** cytokines. Three cytokines produced **by** macrophage were picked **for** analysis: TNF, interleukin (IL)-10 and IL-1? (Figure 2). BMDM cells **plus** DMSO (control) cultures did not produce TNF, IL-10 or IL-1 (data not shown). **Like** *the* results **from** *the* RAW 264.7 cell line, WB extracts alone stimulated TNF production.
(fonte: http://www.biomedcentral.com/1471-2172/10/12)

Os escores da Dimensão 1 para outros sub-registros específicos ajudam ainda mais na interpretação funcional. Por exemplo, a Tabela 4 mostra que existem diferenças importantes entre os sub-registros opinativos e persuasivos em relação a essas características linguísticas: os documentos pertencentes a CONSELHO são extremamente envolvidos, enquanto os BLOGS DE OPINIÃO + NOTÍCIAS (e EDITORIAIS) e BLOGS RELIGIOSOS são distantes e informacionais, em vez de envolvidos. Os exemplos (5) e (6) ilustram essas diferenças.

FIGURA 2. Escores da Dimensão 1 para Registros Simples e Híbridos: Oral Envolvido vs. Letrado Informacional



(5) Opinião: Conselho
(**Negrito marca verbos que não estão no passado**; *itálico marca orações no aspecto progressivo finitas e não finitas (gerúndio);* sublinhado marca pronomes de primeira e segunda pessoas)

You will stop ***pulling*** on others to **make** you special only when you accept the full responsibility of ***making*** yourself **feel** special. This **means** ***learning*** to give yourself all that you may be ***trying*** to **get** from others. [. . .] Instead of ***trying*** to **get** others to **give** you what you want, you can:

Attend to your feelings throughout the day and **explore** what you may be ***doing*** that is ***causing*** painful feelings, rather than ***making*** others responsible for your feelings.

Attend to your own needs rather than ***expecting*** others to meet your needs.

[. . .]

(0290769; http://www.streetdirectory.com/travel_guide/7814/self_improvement_and_motivation/the_need_to_feel_special.html)

(6) Opinião: Blog Religioso
(**Negrito marca preposições**; *itálico marca artigos*)

It is a snare **to** a man to utter a vow (**of** consecration) rashly, and **after** vows to inquire whether he can fulfill them. Both clauses are a protest **against** *the* besetting sin **of** rash and hasty vows. Compare *the* marginal reference.

Who devoureth that which is holy - It is a sin to take that which belongs **to** God, his worship, or his work, and devote it **to** one's own use.

And **after** vows to make inquiry - That is, if a man be inwardly making a rash vow, *the* fitness or unfitness, *the* necessity, expediency, and propriety **of** *the* thing should be first carefully considered. (0202156; http://bible.cc/proverbs/20-25.htm)

Da mesma forma, há uma ampla gama de variação entre os sub-registros dentro da categoria geral narrativa em relação aos escores da Dimensão 1. Por exemplo, blogs pessoais são altamente envolvidos nessa dimensão, como no exemplo (7); a ficção é intermediária; enquanto artigos históricos são extremamente informacionais em suas características da Dimensão 1, como é observado em (8).

(7) Narrativa: Blog Pessoal
(**Negrito marca verbos de posicionamento**; *itálico marca orações no aspecto progressivo finitas e não finitas (gerúndio);* sublinhado marca pronomes de primeira e segunda pessoas)

Like last year (2011 event recaps here and here), I've come home with my head *spinning* and full of ideas that I can't wait to implement over the coming months.

The event was brilliant. On the networking front (although I **prefer** "*connecting*" or, let's face it, *chatting* up a storm) I had a ball *catching* up with the women I met last year, as well as *meeting* some of my beautiful readers in person — and let's not **forget** the B-School babes. There were fabulous ladies at every turn.

Other highlights of the weekend were having a good ol' chat to Sarah Wilson at the post-event drinks on Friday night (yes, she is beyond amazing).

[. . .]

What a great wrap up Rach! Can't **believe** we didn't get a chance to chat. I **promised** myself we would on Saturday but by Saturday my dear sinuses were *tormenting* me so I wasn't in the right shape to **talk** to anyone.

I'm very much ***looking*** forward to *having* a bit of extra time over my holidays where I can spend some time on my blog . . . I **miss** it so much but life is just *getting* in the way at the moment. (Fonte: http://inspacesbetween.com/blogging-business/problogger-event--2012-the-full-wrap-up-pt-1/)

(8) Narrativa: Artigo Histórico

(**Negrito marca preposições**; *itálico marca artigos definidos;* sublinhado marca orações relativas não-finitas)

In 1851, **during** *the* time that there was a gold rush **in** California, a gold rush began **in** Australia. *The* gold **in** California was mainly **in** the form **of** very fine grains, called gold dust.

However, **in** Australia, it was not unusual **for** gold nuggets, some very large, to be found.

The Largest Australian Nuggets

In October 1872 Holtermann's Nugget was found. **At** that time it was *the* world's largest specimen **of** reef gold. It weighed 286 kg and measured 150cm **by** 66cm. Also famous are: *The* Hand **of** Faith (27.2 kg), *the* Welcome Stranger (73.4 kg) and *the* Welcome (69.9 kg) nuggets. (fonte: http://www.kidcyber.com.au/topics/gold.htm)

Mesmo os sub-registros falados e os sub-registros escritos-informacionais variam, de certa forma, ao longo dessa dimensão: dentro do registro geral discurso falado, as transcrições de tv são extremamente envolvidas (ver 2 acima), enquanto discursos formais têm uma pontuação intermediária da Dimensão 1. Ademais, dentro do registro geral descrição informacional, blogs informacionais são intermediários ao longo da Dimensão 1, enquanto sub-registros como artigos de pesquisa e artigos de enciclopédia apresentam os maiores escores negativos da Dimensão 1 (ver 4 acima).

A dimensão 2 pode ser interpretada similarmente, como apresentando oposição geral entre oral e letrado, mas a base linguística desse contraste refere-se mais à elaboração estrutural do que a envolvimento pessoal. Ou seja, as características linguísticas coocorrentes na Dimensão 2 incluem verbos estáticos (verbos existenciais e verbos mentais) e três tipos de orações dependentes orações complementares: orações relativas (com *that*), e orações relativas QU). O

discurso LÍRICO – especialmente as MÚSICAS – é extremamente marcado para o uso dessas características elaboradas, porém, elas aparecem comumente em todos os sub-registros falados de nosso *corpus* (ver exemplos 1 e 2 acima, repetidos aqui como 9 e 10, com construções de oração complementar, destacada em negrito; os exemplos imbricados estão destacados com sublinhando ou duplo sublinhado).

(9) LÍRICO: MÚSICA
Neither of us ever thought, back on the first day that we met
We were both so convinced **that the dream would never end**
Nobody saw the signs man, it's funny **how love is blind** man
'Cause it can be over just as fast as it begins
So now I be sitting here, I'm alone trying **to stuff my fear**
Checkin' out one of the pictures there up on the shelf
The story behind my song, **is that you never know what you got 'til it's gone**
(Fonte: http://lyrics.wikia.com/Scatman_John:Sorry_Seems_To_ Be_The_Hardest_Word)

(10) DISCURSO FALADO: TRANSCRIÇÃO DE TV
V : Fortunately, I got to you before they did.
Evey Hammond : You got to me? You did this to me? You cut my hair? You tortured me? You tortured me! Why?
V : You said **you wanted to live without fear**. I wish **there'd been an easier way**, but there wasn't. I know **you may never forgive me** . . . but nor will you understand **how hard it was for me to do what I did**. Every day I saw in myself everything you see in me now. Every day I wanted **to end it**, but each time you refused **to give in**, I knew **I couldn't**. [. . .] See, at first I thought **it was hate, too**. Hate was all I knew, it built my world, it imprisoned me, taught me **how to eat, how to drink, how to breathe**.
[. . .]
V: So if you've seen nothing, if the crimes of this government remain unknown to you then I would suggest **you allow the fifth of November to pass unmarked**. But if you see **what I see**, if you feel as I feel, and if you would seek as I seek, then I ask you **to stand beside me one year from tonight**, outside the gates of Parliament, and together we shall give them a fifth of November that shall never, ever be forgot. (Fonte: http://www. imdb.com/character/ch0002908/quotes)

Essas características também são bastante comuns em registros escritos orais, como DISCUSSÃO INTERATIVA (FÓRUNS DE DISCUSSÃO OU FÓRUNS DE PERGUNTA-RESPOSTA - ver 3 acima), assim como FICÇÃO (ver 11 abaixo) e BLOG PESSOAL (ver 7 acima).

Finalmente, a Dimensão 3 compreende características linguísticas estereotipicamente narrativas, como verbos no tempo passado, verbos no aspecto perfeito, verbos de ação, advérbios de lugar e tempo, e a maioria dos tipos de orações adverbiais (exceto orações condicionais). O estilo narrativo capturado pela Dimensão 3 é principalmente o estilo em primeira pessoa

(sendo percebido pelo alto valor para pronomes de primeira pessoa), embora os pronomes de terceira pessoa também coocorram com essas características. Os registros gerais mais marcados na web com relação à Dimensão 3 são textos líricos (especialmente músicas - ver 1 acima) e discussão interativa (ver 3 acima). No entanto, ao se considerar os sub-registros específicos (ver Tabela 4), percebe-se que a ficção é, na verdade, o registro que mais depende dessas características, como em (11).

(11) Narrativa: Ficção
(*Itálico marca verbos de ação*; **negrito marca verbos no tempo passado (incluindo passado perfeito)**; sublinhado marcar pronomes de primeira e segunda pessoa)

My family have been prominent, well-to-do people in this Middle Western city for three generations. The Carraways are something of a clan, and we have a tradition that we're descended from the Dukes of Buccleuch, but the actual founder of my line **was** my grandfather's brother, who ***came*** here in fifty-one, ***sent*** a substitute to the Civil War, and **started** the wholesale hardware business that my father carries on to-day.
[. . .]
The practical thing **was** to *find* rooms in the city, but it **was** a warm season, and I **had** just **left** a country of wide lawns and friendly trees, so when a young man at the office **suggested** that we take a house together in a commuting town, it **sounded** like a great idea. He ***found*** the house, a weather-beaten cardboard bungalow at eighty a month, but at the last minute the firm **ordered** him to Washington, and I ***went*** out to the country alone. (Fonte: http://ebooks.adelaide.edu.au/f/fitzgerald/f_scott/gatsby/complete.html)

blogs pessoais também são extremamente marcados para o uso dessas características (ver 7 acima). Curiosamente, porém, artigos históricos têm apenas escore narrativo intermediário na Dimensão 3 (ver 8 acima), enquanto relatos de notícias são verdadeiramente informacionais e não narrativos, em seu escore na Dimensão 3, como ilustrado em (12).

(12) Narrativa: Reportagem de Notícia
(*Itálico marca substantivos*; **negrito marca substantivos pré-modificadores**; sublinhado marca adjetivos atributivos)

There are now *reports* that at least 12 *people* have been killed in *today*'s *crackdown*. *Reuters* reports that the *number* killed in *Hama* has risen to six, and *Avaaz* claims a further *six* were killed in *Homs*. Some of the latest ***YouTube*** videos to emerge from *Syria* are too gruesome to even link to.

The European *Union* has agreed to impose sanctions on 14 Syrian *officials* for their *part* in a violent ***government*** *crackdown* against *protesters*, but ***President*** *Bashar al-Assad* was not among those targeted.

> After a *meeting* of **EU** *ambassadors*, the 27 ***country*** *bloc* said it would impose ***travel*** *restrictions* and ***asset*** *freezes* on the 14 *individuals*, with the *measures* to be formally approved early next *week* if no ***member*** *state* objects.
>
> While *Assad* is not on the *list*, there is the *possibility* that he will be added in *time*, the *official* said. (Fonte: http://www.guardian.co.uk/ news/blog/2011/may/06/syria-libya-middle-east-unrest-live)

Surpreendentemente, TRANSCRIÇÕES DE FALAS DE TV também são extremamente narrativas, em relação às características da Dimensão 3 (ver 2 acima), enquanto DISCURSOS ORAIS FORMAIS são realmente marcados pela ausência dessas características narrativas.

Linguisticamente, o polo negativo da Dimensão 3 é composto por palavras mais longas (tamanho das palavras é uma variável contínua, cuja medida é feita simplesmente pelo número de caracteres) e várias características relacionadas a sintagmas nominais complexos: substantivos comuns, substantivos processuais e três tipos de modificadores nominais (adjetivos atributivos, substantivos pré-modificadores e orações relativas finitas). Essas características são especialmente presentes em registros informacionais, incluindo PERSUASÃO INFORMACIONAL E DESCRIÇÃO INFORMACIONAL. (Ao mesmo tempo, a Dimensão 3 mostra que esses registros informacionais tendem a ser linguisticamente não narrativos.) Entretanto, há variação entre os sub-registros informacionais. Por exemplo, DESCRIÇÕES DE UMA PESSOA E ARTIGOS DE ENCICLOPÉDIA são intermediários ao longo da Dimensão 3, empregando características linguísticas tanto narrativas quanto informacionais. Em contrapartida, ARTIGOS DE PESQUISA são extremamente marcados na Dimensão 3, pela presença densa da característica negativa 'sintagma nominal', simultaneamente à rara presença da característica positiva 'narrativa' (ver 4 acima).

As dimensões 1-3 distinguem-se entre os registros e os sub-registros de formas semelhantes. Os textos do registro LÍRICO (particularmente MÚSICAS) são especialmente dignos de destaque pela ocorrência frequente de características positivas de todas as três dimensões. DISCUSSÕES INTERATIVAS E DISCURSO FALADO (especialmente TRANSCRIÇÕES DE TV) também são altamente marcados para a presença frequente das características linguísticas que definem todas as três dimensões. Na outra extremidade, documentos informacionais escritos têm escores negativos em todas as três dimensões.

Sob uma perspectiva estatística, as dimensões são definidas em termos de padrões de coocorrência linguística. Ou seja, cada fator é extraído como um conjunto de características linguísticas que tendem a coocorrem frequentemente nos textos - sem considerar a categoria de registro desses textos. As similaridades nos padrões de variação do registro em todas as dimensões refletem o fato de que a coocorrência linguística é funcional. Particularmente, a oposição entre discurso oral e letrado foi identificada em estudos anteriores de AMD como o único parâmetro mais importante de variação de registro (ver BIBER, 2014). No entanto, esse parâmetro funcional tem múltiplas correspondências gramaticais/estruturais, as quais foram separadas em múltiplas dimensões de variação na presente análise. Mais especificamente, as características da Dimensão 1 são principalmente aquelas associadas a envolvimento pessoal e interatividade (e.g., aspecto progressivo, pronomes de primeira e segunda pessoa, verbos de posicionamen-

to, adjetivos de posicionamento, advérbios de posicionamento), enquanto as características da Dimensão 2 são aquelas associadas à estrutura sintática típica do discurso oral (especialmente construções do tipo verbo + oração complementar). A Dimensão 3 é definida de forma semelhante, como uma oposição entre verbos (e advérbios) e características de sintagmas nominais, somados ao elemento mais específico: verbos no tempo passado / aspecto perfeito / ação.

Em geral, os registros orais estereotípicos apresentam tanto as características de envolvimento pessoal associadas à Dimensão 1, quanto as características de oração finita complementar associadas à Dimensão 2. No entanto, há exceções para essa generalização. Por exemplo, os documentos OPINIÃO-CONSELHO são extremamente marcados para o uso frequente de características positivas da Dimensão 1 (associadas ao envolvimento pessoal), mas eles apresentam uso moderado das estruturas de oração complementar associadas à Dimensão 2. Por outro lado, essas estruturas de orações orais dependentes são frequentemente empregadas em NARRATIVA-FICÇÃO (resultando em escore positivo alto na Dimensão 2). Todavia, FICÇÃO não é um registro especialmente marcado pelas características linguísticas de envolvimento pessoal que definem a Dimensão 1. Ademais, a Dimensão 3 mostra padrões um tanto diferentes de variação de registro. Por exemplo, FICÇÃO é o registro mais marcado em relação às características orais positivas que definem essa dimensão. No entanto, outros registros mais informacionais, como REPORTAGEM ESPORTIVA e BLOG DE VIAGEM, também são dignos de destaque no que diz respeito à presença frequente de características positivas da Dimensão 3 (embora não sejam especialmente marcados para o uso frequente das características positivas associadas às Dimensões 1 e 2).

Em resumo, todas as três primeiras dimensões do presente estudo podem ser consideradas como oposições entre oral e letrado. Contudo, elas são definidas por diferentes conjuntos de características linguísticas coocorrentes. Os padrões de variação de registro ao longo dessas três dimensões são semelhantes, diferenciando-se entre registros orais estereotípicos como DISCURSO FALADO, DISCUSSÃO INTERATIVA e MÚSICA, em oposição a registros estereotípicos escritos como ARTIGOS DE PESQUISA e ARTIGOS DE ENCICLOPÉDIA. Ao mesmo tempo, porém, cada dimensão está associada a seu próprio padrão único de variação de registro, refletindo o conjunto específico de características linguísticas coocorrentes.

4.2. Dimensões narrativas

As dimensões 3, 4 e 5 apresentam certa relação com propósitos comunicativos narrativos. A Dimensão 3 - discutida acima - contrasta os estilos narrativos estereotípicos encontrados em ficção e outros registros orais com os estilos informacionais encontrados em registros como artigos de pesquisa acadêmica.

A Dimensão 4, interpretada como "Comunicação Relatada", também contrasta narrativa com registros informacionais, mas há duas diferenças significativas em relação à Dimensão 3: registros orais (DISCURSO FALADO, DISCUSSÃO INTERATIVA e discurso LÍRICO) não são marcados em relação à Dimensão 4; e as relações entre os sub-registros narrativos específicos são notadamente diferentes da Dimensão 3: RELATÓRIOS DE NOTÍCIAS são especialmente marcados em relação às características da Dimensão 4, enquanto ARTIGOS HISTÓRICOS e FICÇÃO não são marcados.

Linguisticamente, a Dimensão 4 tem uma base completamente diferente da Dimensão 3. Como discutido na Seção 4.1, a Dimensão 3 consiste em características narrativas estereotípicas, incluindo verbos no passado e verbos no aspecto perfeito, advérbios de tempo e lugar e pronomes. Em contrapartida, a Dimensão 4 é composta por verbos de comunicação (controlando uma oração complementar e ocorrendo em outros contextos), que coocorrem com verbos de probabilidade que controlam essas orações relativas (com *that*). Essa dimensão é interpretada como refletindo a comunicação relatada, uma função que é especialmente dominante em REPORTAGEM DE NOTÍCIAS (ver 12 acima). Por outro lado, é mais provável que FICÇÃO inclua discurso direto (ver 11 acima), enquanto registros narrativos informacionais, como ARTIGOS HISTÓRICOS, incluem relativamente poucos relatos de comunicação no passado (ver exemplo 8 acima).

A função primária da Dimensão 5 é identificar o discurso Irrealis (ver Seção 4.3). Nesse caso, o discurso narrativo está no polo oposto (negativo). Linguisticamente, o polo negativo da Dimensão 5 é definido por poucas características coocorrentes: verbos no tempo passado, sintagmas preposicionais e outras características linguísticas com valores menores nos fatores. Como resultado, a característica linguística mais importante de registros negativos da Dimensão 5 é a ausência de características irrealis, em vez da presença de características narrativas. A discussão sobre as características irrealis será retomada em 4.3.1.

4.3. Dimensões com funções de discurso mais específicas

4.3.1. Discurso Irrealis. A Dimensão 5 é interpretada como sinalizadora do discurso irrealis em oposição ao discurso informacional/narrativa. As funções primárias associadas ao discurso irrealis são a comparação/contraste de várias condições, possibilidades, obrigações e eventualidades. Linguisticamente, essas funções são determinadas pelo uso frequente de verbos modais (possibilidade, obrigação, predição), orações condicionais, pronomes de segunda pessoa, adjetivos epistêmicos e verbo de ligação 'BE'. Vários dos exemplos acima ilustram esse estilo irrealis, incluindo amostras de registros LÍRICO (especialmente MÚSICA, como em 1) e DISCUSSÃO INTERATIVA (como em 3). Os documentos do sub-registro COMO-FAZER/INSTRUCIONAL também apresenta essa constelação de características linguísticas, como ilustrado em (13).

(13) COMO-FAZER/INSTRUCIONAL
(*itálico marca verbos modais*; **negrito marca orações condicionais**; sublinhado marca pronomes de segunda pessoa (incluindo determinantes possessivos)

To register for ANZ Internet Banking you need a Customer Registration Number (CRN) and Telecode.

ANZ Phone Banking customers
If you have already registered for ANZ Phone Banking use this CRN and Telecode to register for ANZ Internet Banking.

What's a CRN and how *can* I get one?

This *will* be either a nine digit number provided to you by an ANZ Customer Service Consultant or your 15 or 16 digit card number.

Enter your password. It *should* be 8-16 characters long and a combination of numbers and letters. You *will* need at least one number and one letter in your password. Your password *should* have no spaces or symbols.

How do I register for ANZ Internet Banking?

To register for ANZ Internet Banking you need a CRN and a Telecode. **If you have not already been supplied with a valid CRN** [. . .], please call the ANZ Internet Banking team [. . .]

If you have already registered for ANZ Phone Banking use the same CRN and Telecode to register for ANZ Internet Banking.

If you do not have a valid CRN or have forgotten your CRN, please call the ANZ Internet Banking team [. . .] (Fonte: http://www.anz.com/Internet-banking/help/getting-started/register/)

Os registros gerais DISCURSO FALADO, OPINIÃO E DESCRIÇÃO INFORMACIONAL normalmente não apresentam esse estilo irrealis. No entanto, sub-registros específicos dentro dessas categorias gerais costumam empregar discurso irrealis, incluindo TRANSCRIÇÕES DE TV (um sub-registro de DISCURSO FALADO; ver 2 acima) e CONSELHO (um sub-registro de OPINIÃO; ver 5). Da mesma forma, as FAQS (PERGUNTAS FREQUENTES) são um exemplo de sub-registro dentro de DESCRIÇÃO INFORMACIONAL, que frequentemente apresenta características irrealis, como exemplificado em (14).

(14) DESCRIÇÃO INFORMACIONAL: PERGUNTAS FREQUENTES (FAQ)
(*itálico marca verbos modais*; **negrito marca orações condicionais**; sublinhado marca pronomes de segunda pessoa (incluindo determinantes possessivos)

Q: I am having trouble losing the extra pounds from the turkey I ate over the holidays and I entered a lightweight category. *Will* I be able to race?

A: **If you are over the 135lb max for women or 165lb for men**, you *will* not be allowed to race in that category. NO EXCEPTIONS. However, you *will* be able to race in an open weight category for your age group and the registration team *will* give priority to accommodate you in a heavier category closest to your age and eligibility. There is no additional cost or penalty for changing weight categories. **If you are unsure**, please consult your coach for advice on your weight. Ideally, you should be at your category weight two weeks prior to February 5, 2012.

Q: *Can* I pick which ergometer I get to race on?

A: No, lane assignments are done randomly by computer. It is very important to sit in the correct lane matched by your name on the screen. Lane officials *will* be able to help you find your lane **if you are unsure** and each erg is labelled.

Q: **If I have to scratch my race** *can* I get my entry fee refunded?

A: It depends when you scratch. **If you do not scratch before February 1**, 2012 by 5:00 p.m. your race fees are non-refundable for any reason. **If you reported your withdrawal and scratch to the Entries Co-ordinator and received a reply email that it was received before Feb 1 deadline** you are eligible for a refund. However, you *will* have to wait. Refunds *will* only be processed starting 14 days after the event and mailed out to you.
(Fonte: http://www.cdnindoorrowing.org/faqs.html)

4.3.2. Discurso Processual/Explicativo. A constelação de características linguísticas coocorrentes associadas à Dimensão 6 funciona para dizer aos leitores o que fazer e como fazê-lo e, portanto, constituem "discurso processual/explicativo". Essas características incluem verbos causativos/facilitadores (e.g., causa, resulta em) e substantivos processuais (e.g., procedimento, processo), simultaneamente com verbos no aspecto progressivo (gerúndio) e verbos de ação. Existem vários sub-registros na web - incluindo documentos do tipo COMO-FAZER ou INSTRUCIONAIS, PERGUNTAS FREQUENTES e BLOGS DE CONSELHO, que frequentemente apresentam essas características para descrever procedimentos ou oferecer outras explicações. Nos exemplos de (15) a (17), os verbos causativos/facilitadores estão marcados em negrito, os substantivos processuais estão destacados em itálico e verbos de ação estão sublinhados.

(15) Exemplos de documentos de COMO-FAZER/INSTRUCIONAL

That extra step can **cause** the *process* to drag on three times as long as a normal home. (Fonte: http://money.cnn.com/2009/01/27/real_estate/hort_sale.moneymag/index.htm)

Basically, since this *procedure* **resulted in** lost visitor paths, I switched to automatic tagging. (Fonte: http://www.ga-experts.com/blog/2006/11/ how-to-get-detailed-ppc-keyword-data-from-google-analytics/)

To **help** you discover ancestors who left these shores (or arrived on them), it's good to have a look through Immigration & Emigration records. (Fonte: http://landing.ancestry.co.uk/intl/uk/gettingstarted.aspx)

(16) PERGUNTAS FREQUENTES (FAQ)

Among other unforeseen problems, indiscriminate *use* of joint ownership can **cause** an increase in estate taxes over the joint lives of married persons, **force** double probates in the event of simultaneous *deaths*, create unfairness as to who pays for funeral expenses and *claims* against the decedent, raise undesired exposure during life to the debts of co-owners, and **cause** a shortage of funds for payment of estate taxes which can **cause** litigation with the taxing authorities.
(Fonte: http://www.floridabar.org/tfb/tfbconsum.nsf/48e76203493b82ad-852567090070c9b9/a0091ab18d4875d085256b2f006c5b75?OpenDocument)

(17) Conselho

The run might elevate heart rate and get a nice boost of serotonin, but a longer brisk walk will actually **facilitate** fat burning and weight loss.
(Fonte: http://lajollamom.com/2011/01/drink-warm-lemon-waterin-the-morning/)

Curiosamente, essas características (especialmente substantivos processuais) também são comuns em artigos de pesquisa, que incluem descrições dos procedimentos utilizados para a pesquisa empírica, como também os resultados desses procedimentos, conforme demonstrado em (18).

(18) Descrição Informacional: Artigos de Pesquisa

The ENCODE project aims at characterising the entire human hereditary information in more detail in order to identify functions for the large, non-protein-coding part of the human genome and to place it in context with the *regulation* of gene *activity*. One prerequisite was the *development* of novel *methods* for large-scale experimental *approaches* as well as for data handling and analysis. Using biochemical and bioinformatics *approaches*, it was possible to identify "candidates" of DNA elements that co-determine when and where in the human body a gene is active. (Fonte: http://www.alphagalileo.org/ViewItem.aspx?ItemId=123846&CultureCode=en)

4.3.3. Posicionamento Letrado. Artigos de pesquisa é o único sub-registro especialmente marcado para características da Dimensão 7, que é interpretada como "posicionamento letrado". Essas características são compostas quase totalmente por substantivos de posicionamento, que ocorrem em uma gama de diferentes ambientes gramaticais (orações complementares de controle e sintagmas preposicionais, assim como em outros contextos). Esses substantivos de posicionamento podem expressar vários significados epistêmicos, como certeza (e.g., fato, conhecimento), probabilidade, possibilidade, dúvida ou uma afirmação ou reivindicação de uma pessoa. Os substantivos de posicionamento também podem expressar significados relacionados a planos ou propostas futuras. Os substantivos de posicionamento geralmente são bastante raros na maioria dos registros da web. No entanto, são comparativamente comuns em artigos de pesquisa, como em (19).

(19) Exemplos de Descrição Informacional: Artigo de Pesquisa

(**Negrito marca substantivos de posicionamento**; *itálico marca substantivos de posicionamento + oração com 'TO'*; sublinhado marca substantivos de posicionamento + oração relativa (*that*), incluindo orações relativas predicativas com *that*)

The **need** *to consider different genetic materials* is also highlighted by the **fact** that varieties of many crops [. . .] show great production variation. One **possibility** is **that females may find the questions more sensitive than males**.

A recurrent **claim** is that the criminal justice system does not place value on the perspectives of victims (Fonte: http://rspb.royalsocietypublishing.org/content/274/1608/303.full)

4.3.4. Descrição de Seres Humanos. A Dimensão 8 é uma dimensão altamente especializada, interpretada como relacionada à descrição de seres humanos. Existem poucas características linguísticas agrupadas nesta dimensão, incluindo substantivos que se referem a humanos (e.g., menino, menina, cara, 'mina', pai, esposa, adulto, acadêmico, conselheiro, advogado, supervisor), pronomes de terceira pessoa e orações relativas finitas (que frequentemente servem para identificar a referência de uma pessoa). Essas características são comuns em DESCRIÇÃO DE PESSOA, como em (20).

(20) DESCRIÇÃO INFORMACIONAL: DESCRIÇÃO DE PESSOA
(**negrito marca substantivos humanos**; *itálico destaca artigos indefinidos*)

Priscilla Beaulieu Presley (born Priscilla Ann Wagner on May 24, 1945, in Brooklyn, New York) is *an* American **model**, **author** and **actress** and the only **wife** of Elvis Presley. Her biological **father**, James Wagner, was *a* pilot who was killed in *a* plane crash when Priscilla was just *an* **infant**. (Fonte: http://priscilla.elvispresley.com.au/)

No entanto, essas características linguísticas são ainda mais comuns em narrativa ficcional. O Exemplo 11 acima ilustra essas características em uma passagem de um romance.

4.3.5. Descrição Não Técnica. Finalmente, a Dimensão 9 apresenta a oposição entre dois estilos nominais: escores positivos na Dimensão 9 refletem densidade de uso de substantivos concretos e outros substantivos comuns; escores negativos na Dimensão 9 refletem densidade de uso de nominalizações. O único registro geral que é especialmente marcado em relação a essas características é COMO-FAZER, e o sub-registro RECEITAS apresenta-se significativamente marcado com o uso de substantivos concretos e outros substantivos comuns. De fato, listas de sintagmas nominais foram frequentemente encontradas nesses textos, apresentando pouca necessidade de orações completas. O Exemplo 21 ilustra esse estilo de discurso.

(21) COMO-FAZER: RECEITA
(**negrito marca substantivos concretos**; *itálico marca artigos indefinidos*; sublinhado marca outros substantivos comuns)

- 1/2 cup **sugar**
- 1.75 cups plain **flour**
- 1 tbsp baking **powder**
- 2 **eggs**, slightly whisked already
- 1/2 cup **milk**
- *A* little under half *a* **stick** of **butter** (110g) — melted

- lots of **vanilla** essence (3-5 spoons)
- lots of **blueberries** (fresh/frozen . . . doesn't matter)
- **Oven**: 200°C

> Spray a 12-hole **muffin pan** with cooking **spray**. Mix dry ingredients (**sugar**, **flour**, baking **powder**) together in *a* **bowl**. Whisk wet ingredients (**eggs**, **milk**, **butter**, essence) together for about *a* minute or two, then add to dry ingredients. [. . .] (Fonte: http://crissybakes.wordpress.com/2012/03/09/blue-muffins-for-a-long-weekend/)

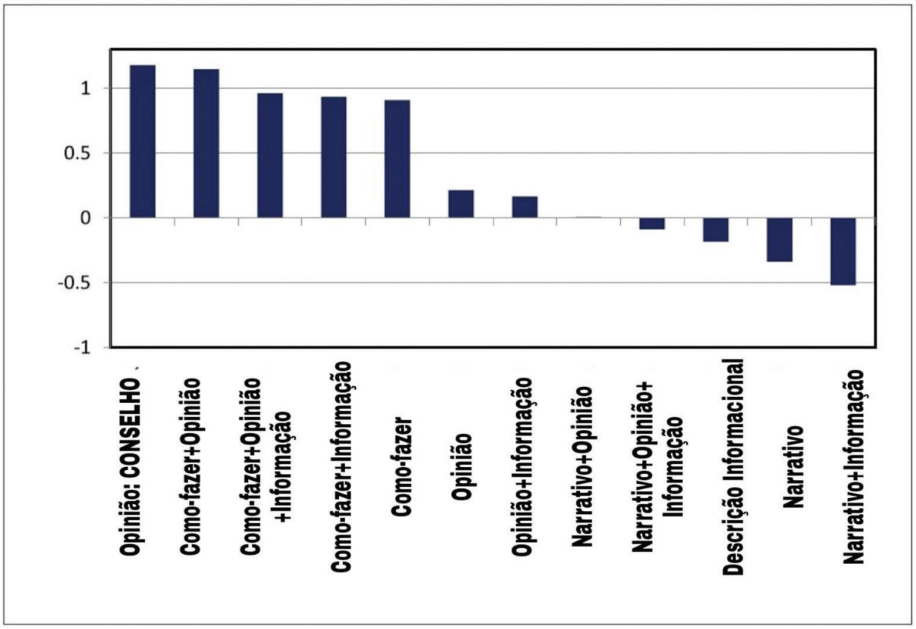
4.4. A análise linguística de registros híbridos

Os registros híbridos foram introduzidos na Seção 2.2. Esses documentos foram codificados com uma divisão 2-2 ou 2-1-1. Por exemplo, dois avaliadores podem ter codificado uma determinada página como NARRATIVA, enquanto outros dois avaliadores classificaram a mesma página como DESCRIÇÃO INFORMACIONAL. Uma possível interpretação dessas divisões é que elas simplesmente mostram falta de concordância entre os avaliadores, refletindo falta de confiança no quadro de registros. No entanto, conforme observado em 2.2, a divisão verdadeira desses pares sugere uma interpretação diferente, pois apenas algumas combinações teoricamente possíveis, de fato, ocorrem comumente (mais notavelmente representando combinações de registros NARRATIVAS, OPINIÕES e INFORMACIONAIS). Em Biber, Egbert e Davies (2015), discutimos vários documentos desse tipo, mostrando como eles representam registros híbridos que combinam os propósitos comunicativos e outras características situacionais de dois ou mais registros gerais.

A AMD fornece evidências que sustentam a natureza híbrida desses documentos. Em geral, os principais registros híbridos têm escores de dimensão que são intermediários entre as categorias "pai" de registro geral. Por exemplo, a Figura 1 mostra os escores da Dimensão 3 ("Narrativa Oral versus Informações Escritas") para três registros gerais (OPINIÃO, NARRATIVA, DESCRIÇÃO INFORMACIONAL), além dos quatro registros híbridos que representam combinações dessas categorias gerais. Linguisticamente, essa dimensão é interpretada com a oposição entre características narrativas (por exemplo, verbos no tempo passado, verbos no aspecto perfeito, advérbios de tempo e lugar) e características informacionais (por exemplo, palavras longas, substantivos e modificadores nominais). Os registros híbridos OPINIÃO + DESCRIÇÃO INFORMACIONAL e NARRATIVA + DESCRIÇÃO INFORMACIONAL são intermediários em seus escores na Dimensão 3, demonstrando que esses documentos empregam características linguísticas positivas e negativas. Por outro lado, NARRATIVA + OPINIÃO tem essencialmente a mesma pontuação na Dimensão 3 que NARRATIVA simples, indicando que os propósitos comunicativos narrativos têm uma influência muito mais forte no uso de características da Dimensão 3 do que propósitos de opinião.

As dimensões 1 e 5 mostram padrões semelhantes para os híbridos que combinam NARRATIVA / OPINIÃO / DESCRIÇÃO INFORMACIONAL. As Figuras 2 e 3 mostram que esses registros híbridos são geralmente intermediários entre os registros "pais" simples. No entanto, há algumas exceções interessantes. Por exemplo, na Dimensão 1, os híbridos com OPINIÃO + COMO-FAZER

FIGURA 3. Escores da Dimensão 5 para Registros Simples e Híbridos: Irrealis vs. Narração Informacional



têm escores positivos extremamente altos ("oral-envolvido"), em comparação a escores mais moderados para os registros gerais de Como-Fazer e Opinião.

A Figura 3 mostra padrões semelhantes para a Dimensão 5. Nesse caso, o registro simples Como-Fazer é extremamente marcado pelas características positivas ("irrealis") da Dimensão 5, e os registros híbridos com Como-Fazer são ainda mais marcados. Na outra extremidade, o registro híbrido Narrativa + Descrição Informacional tem escore negativo maior do que qualquer um dos dois registros "pais".

Nas seções acima, observamos como os sub-registros específicos dentro de uma categoria geral são muitas vezes mais definidos linguisticamente do que seus registros gerais de nível superior. O sub-registro Conselho, dentro da categoria geral Opinião, é um exemplo desse tipo: enquanto Opinião geral tem escores intermediários nas Dimensões 1 e 5, o sub-registro Conselho está entre as categorias de texto mais marcadas, com escores positivos extremamente elevados em ambas as dimensões (ver Tabela 4). As Figuras 2 e 3 mostram que o sub-registro Conselho é quase idêntico ao Como-Fazer + Opinião + híbridos de Descrição Informacional em seus escores nas Dimensões 1 e 5. Esse achado linguístico nos fez reconhecer uma redundância em nosso esquema de codificação: em termos de seus propósitos comunicativos, documentos que dão conselhos implicam uma integração híbrida de Como-Fazer, Opinião e Descrição Informacional. Assim, em pesquisas futuras, essas categorias de registros serão combinadas.

Também planejamos realizar análises mais detalhadas de outros registros híbridos em nossas futuras pesquisas. No entanto, mesmo as descrições dadas aqui sendo preliminares, elas demonstram que essas são categorias distintas de registros, apresentando características linguísticas prontamente interpretáveis. Na maioria dos casos, os registros híbridos têm escores

de dimensão que são intermediários entre os registros "pais", indicando uma mistura de características linguísticas que reflete a mistura de propósitos comunicativos identificados pelos codificadores. Em outros casos, porém, os registros híbridos apresentam caracterizações linguísticas mais extremas do que qualquer um de seus registros "pais", indicando que os codificadores identificaram um tipo específico de sub-registro que vai além de uma simples integração de propósitos comunicativos gerais. Como-Fazer + híbridos de Opinião são exemplos desse tipo de sub-registro, que se mostraram altamente semelhantes aos documentos de Conselho.

5. Conclusão

Os resultados da presente análise MD corroboram os universais de variação de registro propostos, que surgiram de estudos anteriores de MD (ver discussões em BIBER, 1995, p. 7 e p. 10; Biber, 2014). Esses universais incluem:

- Dimensões linguísticas refletindo a oposição fundamental entre o discurso oracional/oral e o discurso frasal/letrado.
- Dimensões linguísticas refletindo diferentes aspectos de discurso narrativo versus não narrativo.
- Dimensões linguísticas refletindo a expressão de posicionamento.

Essas três considerações funcionais estão fortemente refletidas na presente AMD de registros da web. Ao mesmo tempo, há diferenças notáveis entre nossa AMD de registros da web e AMD anteriores com outros domínios do discurso. A presente análise identificou mais dimensões de variação do que análises anteriores. Essas dimensões são representadas por um número relativamente grande de características linguísticas coocorrentes, as quais são prontamente interpretáveis em termos funcionais, em virtude de funções compartilhadas por essas características simultaneamente com a distribuição de registros ao longo da dimensão. No entanto, a análise global resultou em distinções mais detalhadas do que em estudos anteriores. Essas distinções refletem a natureza abrangente do *corpus* (bem como o grande número de características léxico-gramaticais incluídas na análise). Ademais, o *corpus* está baseado em uma grande amostra aleatória da totalidade da web pesquisável, incorporando um conjunto consideravelmente mais diversificado de registros do que em estudos MD anteriores. Os registros incluem exemplos de lírico, transcrições de TV, entrevistas faladas, conselho, blogs pessoais, como-fazer, perguntas frequentes, ficção, reportagens de notícias, blogs de opinião pessoal, artigos de enciclopédia, artigos de pesquisa, etc.

Embora nosso *corpus* represente uma amostra, quase aleatória, abrangente de documentos da web pública pesquisável, é importante lembrar os tipos de textos que não estão incluídos. Obviamente, nosso estudo não inclui registros privados da Internet (por exemplo, mensagens de e-mail, *tweets* e postagens de outras mídias sociais). Em contrapartida, porém, nosso *corpus* também não inclui documentos da *deep web* (a Internet oculta). São os milhões de documentos

informacionais protegidos por senha e armazenados na Internet, incluindo artigos de pesquisa publicados em revistas acadêmicas referenciadas e documentos técnicos encontrados em sites corporativos ou institucionais. Nossos resultados mostraram que documentos informacionais compreendem < 15% dos documentos encontrados na web pública pesquisável. No entanto, se fosse possível coletar amostras de documentos da *deep web* privada, essa proporção seria muito maior, incluindo um número muito maior de textos altamente técnicos.

O domínio do discurso da web pesquisável difere ainda mais do domínio do discurso analisado em estudos MD mais antigos, por sua ausência de discurso falado verdadeiro. Nosso *corpus* da web inclui Transcrições De TV, de Entrevistas e de Discursos Formais. O *corpus* também inclui textos Líricos, incluindo letras de músicas. Todavia, esses textos têm quantidade limitada e não há exemplos de conversa cara a cara. Apesar dessas diferenças, três das dimensões da análise descrita aqui referem-se a uma oposição oral fundamental, que incorpora as mesmas características lexico-gramaticais orais que análises anteriores (e.g., verbos, advérbios, pronomes, orações adverbiais finitas, orações complementares finitas).[9]

Uma das perguntas levantadas no início deste trabalho foi se o Web-as-Corpus pode ser usado como *corpus* substituto a fim de representar o inglês como um todo. Nossos achados sugerem veementemente que a linguagem na web é distinta de outros domínios da linguagem, de muitas formas. Enquanto certos registros da web são semelhantes ou idênticos a seus pares tradicionais e não da web (por exemplo, artigos de enciclopédia, editoriais, reportagens de notícias, músicas, receitas), outros registros não existem fora da web (por exemplo, blogs de viagem, blogs de opinião pessoal, fóruns de discussão e fóruns de pergunta e resposta). Além disso, há muitos registros em inglês que não estão disponíveis na web pública pesquisável (como livros didáticos, romances de ficção contemporâneos, artigos de revistas acadêmicas referendadas, relatórios técnicos corporativos - bem como conversa e muitos outros registros genuinamente falados). Portanto, defendemos que um *corpus* de documentos da web representa a web - não o inglês em geral. Todavia, essa conclusão, tomada com cautela, não deve ser interpretada como um comentário negativo sobre o valor da web como recurso para a pesquisa linguística. Dada a compreensão da natureza e da composição da web, juntamente com uma compreensão de pesquisas na web (i.e., como interpretar as informações retornadas por uma busca na web),

[9] Em pesquisas futuras, planejamos realizar experimentos para prever automaticamente a categoria de registro de documentos da web. Com base nos resultados da presente AMD, combinaremos algumas categorias em nosso *corpus* presente que não sã bem distintas linguisticamente. Por exemplo, expandiremos a categoria de Persuasão Informacional simples para também incluir Descrição Informacional + híbridos de Persuasão Informacional. Criaremos uma nova categoria Persuasão Informacional opinada, que combina quatro híbridos: Opinião + Persuasão Informacional, Descrição Informacional + Opinião, Opinião + Descrição Informacional + Persuasão Informacional e Opinião + Persuasão Informacional + Narrativa. (Também certamente combinaremos a categoria Persuasão Informacional com a categoria Persuasão Informacional Opinada. A categoria Narrativa + Descrição Informacional será expandida para incluir Narrativa + Descrição Informacional + a híbrida Persuasão Informacional. E, finalmente, o registro geral Como-Fazer será expandido para incluir o híbrido Como-Fazer + Descrição Informacional, enquanto híbridos de Como-Fazer que incorporam Opinião serão amalgamados ao sub-registro Conselho. O objetivo final dessas reclassificações é criar um *corpus* com categorias que sejam claramente distintas tanto situacionalmente quanto linguisticamente. Em seguida, realizaremos uma série de experimentos para desenvolver e testar algoritmos para a predição automática de categorias de registro da web. Na etapa final desse processo, aplicaremos esses algoritmos ao *Corpus* GloWbE de 1,9 bilhões de palavras, criando um *corpus* extremamente grande de documentos da web que terá sido automaticamente codificado para categorias de registros.

o Web-as-Corpus fornece um recurso único para a pesquisa linguística. No presente trabalho, contribuímos para esse conhecimento basilar, utilizando a AMD para descrever os padrões de variação linguística entre os registros e os sub-registros na web.

REFERÊNCIAS

BARONI, Marco; BERNARDINI, Silvia. BootCaT: Bootstrapping corpora and terms from the web. In: LINO, Maria Teresa; XAVIER, Maria Francisca; FERREIRA, Fátima; COSTA, Rute; SILVA, Raquel (Eds.), **Proceedings of LREC 2004**, p. 1313-1316. Lisbon: ELDA, 2004

BARONI, Marco; BERNARDINI, Silvia; FERRARESI, Adriano; ZANCHETTA, Eros. The WaCky wide web: A collection of very large linguistically processed web-crawled corpora. **Language Resources and Evaluation**, v. 43, n. 3, p. 209-226, 2009.

BERBER-SARDINHA, Tony. 25 years later: Comparing Internet and pre-Internet registers. In: BERBER-SARDINHA, Tony; VEIRANO-PINTO, Marcia (Eds.), **Multi-dimensional analysis, 25 years on: A tribute to Douglas Biber**, p. 81-105. Philadelphia: John Benjamins, 2014.

BERBER-SARDINHA, Tony; VEIRANO-PINTO, Marcia (Eds.). **Multi-dimensional analysis, 25 years on: A tribute to Douglas Biber**. Philadelphia: John Benjamins, 2014.

BIBER, Douglas. Investigating macroscopic textual variation through multi-feature/multidimensional analyses. **Linguistics,** v. 23, n. 2, p. 337-360, 1985.

BIBER, Douglas. Spoken and written textual dimensions in English: Resolving the contradictory findings. **Language,** v. 62, n. 2, p. 384-414, 1986.

BIBER, Douglas. **Variation across speech and writing**. Cambridge: Cambridge University Press, 1988.

BIBER, Douglas. **Dimensions of register variation:** A cross-linguistic perspective. Cambridge: Cambridge University Press, 1995.

BIBER, Douglas. University language: A corpus-based study of spoken and written registers. Amsterdam: John Benjamins, 2006.

BIBER, Douglas. Using multi-dimensional analysis to explore cross-linguistic universals of register variation. **Languages in Contrast**, v. 14, n. 1, p. 7-34, 2014.

BIBER, Douglas; CONRAD, Susan. **Register, genre, and style.** Cambridge: Cambridge University Press, 2009.

BIBER, Douglas; EGBERT, Jesse; DAVIES, Mark. Exploring the composition of the searchable web: A corpus-based taxonomy of web registers. **Corpora,** v. 10, n. 1, p. 11-45, 2015.

BIBER, Douglas; KURJIAN, Jerry. Towards a taxonomy of web registers and text types: A multi-dimensional analysis. In: HUNDT, Marianne; NESSELHAUF, Nadja; BIEWER, Carolin (Eds.), **Corpus linguistics and the web**. Amsterdam: Rodopi, p. 109-132, 2007.

CONRAD, Susan; BIBER, Douglas. (eds.). **Multi-dimensional studies of register variation in English**. London: Longman, 2001.

CROWSTON, Kevin; KWASNIK, Barbara; RUBLESKE, Joseph. Problems in the use-centered development of a taxonomy of web genres. In: MEHLER, Alexander; SHAROFF, Serge; SANTINI, Marina (Eds.). **Genres on the web:** Computational models and empirical studies. New York: Springer, p. 69-86, 2010.

EGBERT, Jesse; BIBER, Douglas; DAVIES, Mark. Developing a bottom-up, user-based method of web register classification. **Journal of the Association for Information Science and Technology**, v. 66, n. 9, p. 1817-1831, 2015.

FLETCHER, William H. Corpus analysis of the World Wide Web. In: CHAPELLE, Carol. (Ed.), **Encyclopedia of applied linguistics**. Oxford: Wiley-Blackwell, p. 1339-1347, 2012.

FRIGINAL, Eric (Ed.). Twenty-five years of Biber's Multi-dimensional analysis [Special Issue]. **Corpora**, v. 8, n. 2, 2013.

GATTO, M. **Web as corpus:** Theory and practice. New York: Bloomsbury Academic, 2014.

GRIEVE, Jack; BIBER, Douglas; FRIGINAL, Eric; NEKRASOVA, Tatiana. Variation among blog text types: A multi-dimensional analysis. MEHLER, Alexander; SHAROFF, Serge; SANTINI, Marina (Eds.). **Genres on the web:** Corpus studies and computational models. New York: Springer, p. 303-322, 2011.

HARDY, Jack; FRIGINAL, Eric. Linguistic variation among Filipino and American blogs and online opinion columns. **World Englishes,** v. 31, v. 1, p. 1-19, 2012.

HERRING, Susan C.; PAOLILLO, John C. Gender and genre variation in weblogs. **Journal of Sociolinguistics,** v. 10, n. 4, p. 439-459, 2006.

KILGARRIFF, Adam; GREFENSTETTE, Gregory. Introduction to the special issue on the web as corpus. **Computational Linguistics,** v. 29, n. 3, p. 333-347, 2003.

REHM, Georg; SANTINI, Marina; MEHLER, Alexander; BRASLAVSKI, Pavel; GLEIM, Rudiger; STUBBE, Andrea; SYMONENKO, Svetlana; TAVOSANIS, Mirko; VIDULIN, Vedrana. Towards a reference corpus of web genres for the evaluation of genre identification systems. In: **Proceedings of LREC 2008**. Marrakech, Morocco, p. 351-358, 2008.

ROSSO, Mark A.; HAAS, Stephanie W. Identification of web genres by user warrant. MEHLER, Alexander; SHAROFF, Serge; SANTINI, Marina (Eds.). **Genres on the web:** Computational models and empirical studies, New York: Springer, p. 47-68, 2010.

SANTINI, Marina. Characterizing genres of web pages: Genre hybridism and individualization. In: SPRAGUE, Ralph H. (Ed.), **Proceedings of the 40th Hawaii International Conference on System Sciences** (HICSS-40). Waikoloa, Hawaii. p. 1-10, 2007.

SANTINI, Marina. Zero, single, or multi? Genre of web pages through the users' perspective. **Information Processing and Management.** v. 44, n. 2, p. 702-737, 2008.

SANTINI, Marina; SHAROFF, Serge. Web genre benchmark under construction. **Journal for Language Technology and Computational Linguistics.** v. 25, n. 1, p. 125-141, 2009.

SHAROFF, Serge. Creating general-purpose corpora using automated search engine queries. In: BARONI, Marco; Silvia, BERNARDINI (eds.), **WaCky! Working papers on the web as corpus**. Gedit, Bologna: University of Bologna, p. 63-98, 2005.

SHAROFF, Serge. Open-source corpora: Using the net to fish for linguistic data. **International Journal of Corpus Linguistics**. v. 11, n. 4, p. 435-462, 2006.

SHAROFF, Serge; WU, Zhili; MARKERT, Katja. The web library of Babel: Evaluating genre collections. In: CALZOLARI, Nicoletta; CHOUKRI, Khalid; MAEGAARD, Bente; MARIANI, Joseph; ODIJK, Jan; PIPERIDIS, Stelios; ROSNER, Mike; TAPIAS, Daniel (Eds.), **Proceedings of LREC 2010**, Valetta, Malta. p. 3063-3070, 2010.

TITAK, Ashley; ROBERSON, Audrey. Dimensions of web registers: An exploratory multidimensional comparison. **Corpora.** v. 8, n. 2, p. 239-271, 2013.

VIDULIN, Vedrana; LUŠTREK, Mitja; GAMS, Matjaz. Multi-label approaches to web genre identification. **Journal for Language Technology and Computational Linguistics**. v. 24, n. 1, p. 97-114, 2009.

DECLARAÇÃO DE CONFLITO DE INTERESSE

O(s) autor(s) declararam não existirem conflitos de interesse em relação à pesquisa, autoria e/ou publicação deste artigo.

Financiamento – O(s) autor(es) divulgaram o recebimento do seguinte apoio financeiro para a pesquisa, autoria e/ou publicação deste artigo: Esse material foi desenvolvido com o apoio da Fundação Nacional de Ciências sob o nº 1147581.

APÊNDICE 1

CARACTERÍSTICAS LINGUÍSTICAS

1. Verbo de ligação "be"
2. Verbos no aspecto progressivo (gerúndio)
3. Verbos no tempo não-passado
4. Verbos no tempo passado
5. Verbos no aspecto perfeito (particípio)
6. Verbos de ação
7. Verbos existenciais
8. Verbos mentais
9. Verbos epistêmicos (sem controlar uma oração complementar)
10. Verbos de comunicação (*not controlling a complement clause*)
11. Verbos causativos/facilitadores
12. Verbos de desejo + oração com "para" ('to')
13. Verbo + oração relativa QU (WH)
14. Verbo de probabilidade + oração relativa (com '*that*')
15. Verbos de certeza + oração relativa (com '*that*')
16. Verbo + oração com "para" ('to') (excluindo verbos de desejo)
17. Verb + oração relativa QU (WH)
18. apagamento de 'that'
19. Verbos de comunicação + oração relativa (com '*that*')
20. Coordenação
21. Pronomes de 1ª. pessoa
22. Pronomes de 2ª. pessoa
23. Pronomes de 3ª. pessoa
24. Pronomes demonstrativos
25. Pronome neutro ('*it*')
26. Advérbios de posicionamento
27. Advérbios de tempo
28. Advérbios de lugar
29. Total dos outros advérbios
30. Orações adverbiais (excluindo condicional)
31 Orações adverbiais condicionais
32. Advérbios de ligação
33. Artigos indefinidos
34. Artigos definidos
35. Sintagmas preposicionais
36. Substantivos próprios

37. Substantivos concretos
38. Substantivos comuns
39. Substantivos processuais
40. Substantivos humanos
41. Substantivos pré-modificadores
42. Nominalizações
43. Substantivos cognitivos
44. Substantivo de posicionamento + Sintagma preposicional
45. Substantivo de posicionamento + oração complementar
46. Outros substantivos de posicionamento
47. Orações relativas finitas
48. Orações relativas passivas não finitas
49. Adjetivos atributivos
50. Adjetivos atitudinais (sem controlar uma oração complementar)
51. Adjetivos epistêmicos (sem controlar uma oração complementar)
52. Razão forma/palavra
53. Palavras longas
54. Contrações
55. Modais de possibilidade
56. Modais de predição
57. Modais de necessidade

APÊNDICE 2

TRADUÇÃO DOS EXEMPLOS

(1)

Nenhum de nós jamais pensou, lá no primeiro dia que nos conhecemos
Nós dois estávamos tão convencidos de que o sonho nunca **acabaria**
Ninguém viu os sinais, cara, **É** engraçado como o amor **é** cego, cara
Pois pode acabar (**estar** acabado) tão rápido quanto **começa**
Então agora eu **estou** ***sentando*** aqui, Eu **estou** sozinho ***tentando*** **encher** meu medo
Olhando uma das fotos em cima da estante
A estória por trás da minha música **é** que você nunca **sabe** o que você **tem** até que perder

(2)

V : Felizmente, Eu cheguei até você antes deles.

Evey Hammond : Você chegou até mim? Você fez isso comigo? Você cortou meu cabelo? Você me torturou? Você me torturou! Por quê?

V : Você *disse* que você *queria* **viver** sem medo. Eu ***gostaria*** que tivesse existido um jeito mais fácil, mas não houve. Eu ***sei*** que você pode nunca me **perdoar**. . . mas você também não vai ***entender*** o quanto foi difícil para mim **fazer** o que eu fiz. Todos os dias eu via em mim mesma tudo o que você **vê** em mim agora. Todos os dias eu *queria* **terminar** com isso, mas cada vez que você *recusou* a **ceder**, eu sabia que eu não poderia. [. . .] Veja, no início eu *pensei* que fosse ódio também. Ódio era tudo o que eu *conhecia*, que construiu o *meu* mundo, me aprisionou, me ensinou a **comer**, a **beber**, a **respirar**.

[. . .]

V: Então, se você não **tem** visto nada, se os crimes deste governo **permanecem** desconhecidos para você, então eu ***sugiro*** que você ***permita*** que cinco de novembro não **passe** despercebido. Mas, se você **vê** o que eu **vejo**, se você ***sente*** o que eu ***sinto***, e se você ***procura*** como eu ***procuro***, então, eu lhe ***peço*** que você **fique** ao meu lado por um ano, a partir desta noite, fora dos portões do Parlamento, e juntos nós poderemos **dar** a eles um cinco de novembro que jamais deverá ser ***esquecido***.

(3)

Também **depende** de onde você está ***baixando***. às vezes não é culpa da VM. Também, talvez você esteja ***tendo*** **gerenciamento** de tráfego, você já considerou isso?

Se você assinou um contrato de 12 meses com a operadora de banda larga virgin media e você está fora do período inicial de 30 dias, eu não **acho** que você pode simplesmente "**cancelar**". Você ainda pode **conseguir** que BT seja **instalado** mas eles **requerem** que você **continue** ***pagando*** pela VM.

Se alguém lhe ajudou **diga** obrigado ***clicando*** na Estrela do Kudos. Se algu´´m resolveu o seu problema, por que não **marcar** a mensagem dele com Solução Aceita

Eu estava apenas ***tentando*** lhe **deixar** ciente de que se você assinou BT e tem um contrato de 12 meses que você anida está sob o contrato de 12 meses com a virgin media não vai **ser** tão fácil **cancelar**. Mas se você quer **agir** como uma criança imatura então **fique** à vontade.

(4)

Experimentos posteriores usaram macrófagos isolados ***da*** medula óssea (MO) **de** camundongos do tipo selvagem (TS). *Os* dados **dos** extratos de cogumelos tipo WB estão apresentados desde que *a* maior secreção de TNF (Fator de Necrose Tumoral) **das** células RAW264.7 foram encontradas ***nas*** culturas de WB (Figura 1). *As* células BMDM são macrófagos não alterados de ratos que produzem uma série **de** citocinas. Três citocinas produzidas **por** macrófagos foram selecionadas **para** análise: TNF, interleucina (IL)-10 e IL-1? (Figure 2). *As* células BMDM **mais** as culturas DMSO (controle) não produziram TNF, IL-10 nem IL-1 (dados não apresentados). **Assim como** *os* resultados ***da*** linha celular RAW 264.7, os extratos de WB isoladamente estimularam a produção de TNF.

(5)

Você vai parar de ***forçar*** os outros a lhe **considerarem** especial apenas quando você aceitar a plena responsabilidade de ***fazer*** a si mesma **sentir** especial. Isso **significa** ***aprender*** a se dar tudo o que você pode estar ***tentando*** **conseguir** dos outros. [. . .] Ao invés de **tentar conseguir** que os outros lhe **forneçam** o que você quer, você pode:

Cuidar dos seus sentimentos ao longo do dia e **explorar** o que você pode estar ***fazendo*** que está ***causando*** sentimentos dolorosos, ao invés de ***colocar*** a responsabilidade nos outros pelos seus sentimentos.

Cuidar de suas próprias necessidades ao invés de ***esperar*** que os outros atendam suas necessidades.

[. . .]

(Fonte: http://www.streetdirectory.com/travel_guide/7814/self_improvement_and_motivation/the_need_to_feel_special.html)

(6) OPINIÃO: BLOG RELIGIOSO

(**Negrito marca preposições**; *itálico marca artigos*)

É uma armadilha **para** um homem pronunciar um voto (**de** consagração) impetuosamente, e **após** os votos questionar se ele consegue cumpri-los. As duas cláusulas são um protesto **contra** *o* pecado persistente **de** votos impetuosos e precipitados. Compare com *a* referência na margem.

Quem devora o que é sagrado - É pecado tirar aquilo que pertence **a** Deus, sua adoração ou seu trabalho, e devotá-lo **ao** próprio uso.

E **após** os votos questionar - Isso é, se um homem inadvertidamente consagrar um voto impetuoso, *o* adequado ou inadequado, *a* necessidade, expediente, e propriedade ***das*** coisas devem ser primeiro consideradas cuidadosamente. (Fonte: http://bible.cc/proverbs/20-25.htm)

(7)

Como no ano passado (eventos de 2011 lembrados aqui e acolá), eu cheguei em casa com minha cabeça *girando* e cheia de ideias que eu mal posso esperar para implementar nos próximos meses.

O evento foi brilhante. Na frente de *networking* (embora eu **prefira** "*conexão*" ou, na real, *conversando* com uma tempestade) eu me diverti me *atualizando* com as mulheres que eu conheci ano passado, da mesma forma que *conhecendo* alguns das minhas lindas leitoras pessoalmente — e não vamos nos **esquecer** das gatinhas da faculdade de Business Havia mulheres fabulosas em todos os lugares.

Outro destaque do fim de semana foi ter um bom bate papo Sarah Wilson durante os drinks após o evento na sexta a noite (sim, ela é mais do que.

[. . .]

Que ótimo final, Rach! Nem consigo **acreditar** que nós não tivemos a chance de conversar. Eu me **prometi** que nós iríamos no sábado, mas, no sábado, minha sinusite estava me *atormentando,* então eu não estava bem para **conversar** com ninguém.

Eu estou ***esperando*** com muita ansiedade para *ter* um pouco de tempo extra nas minhas férias para eu poder dedicar um tempo ao meu blog . . . Eu **sinto** muita falta, mas a vida está simplesmente *atrapalhando no* momento. (Fonte: http://inspacesbetween.com/blogging-business/problogger-event-2012-the-full-wrap-up-pt-1/)

(8)

Em 1851, **durante** *o* período em que houve uma corrida por ouro **na** Califórnia, uma corrida por ouro começou **na** Austrália. *O* ouro **na** Califórnia era principalmente **na** forma **de** grãos muito finos, chamados poeira dourada.

Entretanto, **na** Austrália, não era incomum **de** serem encontradas pepitas de ouro, algumas muito grandes.

As Maiores Pepitas Australianas

Em outubro de 1872 a Pepita de Holtermann foi encontrada. **Naquela** época, ela era *a* maior espécie do mundo de veia **de** ouro. Pesava 286 kg e media 150 cm **por** 66 cm. Também são pedras famosas: *A* Mão **da** Fé (27.2 kg), *a* Benvindo Estranho (73.4 kg) e *a* Benvindo (69.9 kg). (Fonte: http://www.kidcyber.com.au/topics/gold.htm)

(9) É o mesmo texto da transcrição (1)

(10) É o mesmo texto da transcrição (2)

(11)

Minha família descendia de pessoas importantes e abastadas que tinham vivido numa cidade do centro-oeste dos Estados Unidos durante três gerações. Os Carraways compõem uma espécie de clã e, temos uma tradição familiar de que nós descendemos dos duques de Buccleuch, mas o verdadeiro fundador da minha linhagem **foi** o irmão do meu avô, que ***veio*** para cá no ano de

1851, ***enviou*** um substituto para combater no exército em seu lugar durante a Guerra da Secessão e **fundou** a empresa atacadista de ferragens a que meu pai administra até hoje

[. . .]

A coisa mais prática a fazer **era** *encontrar* acomodações na cidade, mas na época **fazia** a muito calor e eu **acabara de deixar** uma região de extensos gramados e árvores amigas; ; desse modo, quando um rapaz do escritório **sugeriu** nós alugássemos uma casa juntos em uma das cidades-dormitório próximas a Nova York, a ideia me **pareceu** bastante interessante. Ele até mesmo ***encontrou*** a casa, um sobradinho de madeira com divisórias de compensado e bastante maltratado pelo tempo, ao preço de oitenta dólares por mês.; mas no último minuto a firma **ordenou** que ele se transferisse para a filial de Washington, o que me levou a assumir o aluguel completo e ***parti*** sozinho para o campo. (Fonte: http://ebooks.adelaide.edu.au/f/fitzgerald/f_scott/gatsby/complete.html)

(fonte em português: https://esadmacommunication.files.wordpress.com/2017/02/f-scott--fitzgerald-o-grande-gatsby.pdf)

(12)

Foi agora *reportado que no mínimo* 12 *pessoas* foram mortas na *ação punitiva de hoje*. A *Reuters* informa que o *número* de mortos em *Hama* aumentou para *seis*, e a *Avaaz* afirma que mais *seis* foram mortos em *Homs*. Alguns dos últimos vídeos vindos da *Síria* disponíveis no ***YouTube*** são tão horripilantes mesmo vendo só o link.

A *União* Europeia concordou em aplicar penas a 14 *oficiais* ***sírios*** pela participação deles numa *ação punitiva* ***governamental*** violenta contra os *manifestantes*, mas o ***presidente*** *Bashar al-Assad* não estava entre os visados.

Após a *reunião* dos *embaixadores* da **UE**, o *bloco* de 27 **países** disse que imputaria *restrições* de ***viagem*** e *congelamento* de ***bens*** sobre os 14 *indivíduos*, as *medidas* ainda serão formalmente aprovadas no início da próxima *semana* se nenhum dos *estados* ***membros*** se opuser.

Embora *Assad* não esteja na *lista*, há a *possibilidade* de que ele seja adicionado a tempo, disse um *representante*. (Fonte: http://www.guardian.co.uk/ news/blog/2011/may/06/syria-libya--middle-east-unrest-live)

(13)

Para se registrar para o Internet Banking do ANZ você precisa de um Número de Registro de Cliente (NRC)e um Código de telefone.

Clientes de Fone Banking do ANZ

Se você já se registrou para o Fone Banking da ANZ use o mesmo NRC e código de telefone para se registrar para o Internet Banking do ANZ.

O que é um NRC e como eu *posso* conseguir um?

Será ou uma sequência numérica de nove dígitos que você receberá do Consultor de Serviço ao Consumidor do ANZ ou o número de 15 ou 16 dígitos do seu cartão.

Digite sua senha. Ela *deve* ter de 8 a 16 caracteres e ser uma combinação de números e letras. Você *precisará* ter pelo menos um número e uma letra em sua senha. Sua senha não *deve* ter espaços nem símbolos.

Como me registro para o Internet Banking do ANZ?

Para se registrar no Internet Banking do ANZ, você precisa de um NRC e um Código de telefone. **Se você ainda não recebeu um NRC válido** [. . .], por favor telefone para a equipe de Internet Banking do ANZ [. . .]

Se você já se registrou para o Fone Banking do ANZ use o mesmo NRC e Código de telefone para se registrar para o Internet Banking do ANZ.

Se você não tem um NRC válido ou esqueceu o seu, por favor telephone para a equipe de Internet Banking do ANZ [. . .] (Fonte: http://www.anz.com/Internet-banking/help/getting-started/register/)

(14)

Q: Estou tendo dificuldade em perder os quilos extras que eu ganhei do peru que eu comi nas festas e eu me inscrevi numa categoria peso leve. *Poderei* participar da corrida?

A: **Se você estiver com mais de 61 kg para mulheres ou 75 kg para homens, no máximo**, não *será* permitido que você corra nessa categoria. SEM EXCEÇÕES. Entretanto, *será* permitido que você corra numa categoria sem marcação de peso, por sua idade e a equipe de inscrição *dará* prioridade para acomodar você numa categoria mais pesada e mais próxima a sua idade e aptidão. Não há custo extra ou multa por mudar de categoria de peso. **Se você não tem certeza**, por favor consulte seu técnico para indicações sobre seu peso. O ideal é que você esteja em sua categoria de peso duas semanas antes de 5 de fevereiro de 2012.

Q: *Posso* escolher o ergômetro que eu vou usar para correr?

A: Não, a atribuição de pistas é feita aleatoriamente por. É muito importante estar na pista correta, que contém seu nome na tela. Os responsáveis pelas pistas *serão* capazes de ajudar-lhe a encontrar a sua pista **se você não tiver certeza** e cada ergômetro está etiquetado.

Q: **Se eu tiver que cancelar minha corrida** *posso* ter reembolso da taxa de inscrição?

A: Depende de quando você cancelar. **Se você não cancelar antes de 1º. de fevereiro** de 2012 até às 17h, suas taxas não fazem jus a reembolso por nenhuma razão. **Se você informou sua retirada e cancelamento para o coordenador de Inscrições e recebeu um e-mail antes 1º. de fevereiro,** você faz jus a reembolso. Entretanto, você *terá* que esperar. Os reembolsos apenas *serão* processados 14 dias após o evento e será enviado a você por e-mail. (Fonet: http://www.cdnindoorrowing.org/faqs.html)

(15)

Esse passo extra pode **provocar** que o *processo* se arraste três vezes mais que numa casa normal.

(Fonte: http://money.cnn.com/2009/01/27/real_estate/hort_sale.moneymag/index.htm)

Basicamente, desde que esse *procedimento* **resultou** em perda de caminho do usuário, eu troquei para etiquetagem automática. (Fonte: http://www.ga-experts.com/blog/2006/11/ how--to-get-detailed-ppc-keyword-data-from-google-analytics/)

Para **ajudar** você a descobrir ancestrais que deixaram estas terras (ou chegaram nelas), é bom dar uma olhada nos registros de Imigração & Emigração.
(Fonte: http://landing.ancestry.co.uk/intl/uk/gettingstarted.aspx)

(16)
Entre outros problemas não previstos, *uso* indiscriminado de posse conjunta pode **causar** um aumento nos impostos prediais sobre vidas conjuntas de pessoas casadas, **forçar** inventário duplo em caso de *mortes* simultâneas, criar injustiças sobre quem paga as despesas funerais e *alegações* contra o descendente, criar, em vida, exposição indesejada aos débitos dos coproprietários e **provocar** diminuição de fundos para o pagamento dos impostos prediais, o que pode **causar** litígio por parte das autoridades. (Fonte: http://www.floridabar.org/tfb/tfbconsum.nsf/48e76203493b82ad852567090070c9b9/a0091ab18d4875d085256b2f006c5b75?OpenDocument)

(17) CONSELHO
A corrida pode elevar a frequência dos batimentos cardíacos e promover um aumento positivo de serotonina, mas uma caminha ávida mais longa da irá, de fato, **facilitar** a queima de calorias e a perda de peso. (Fonte: http://lajollamom.com/2011/01/drink-warm-lemon-waterin--the-morning/)

(18)
O objetivo do projeto ENCODE é caracterizar a totalidade de informações da hereditariedade humana mais detalhadamente, a fim de identificar as funções da parte maior e não codificadora de proteína do genoma humano e colocá-lo no contexto da *regulamentação* da *atividade* do gene. Um pré-requisito foi o *desenvolvimento* de novos *métodos* para *abordagens* experimentais em larga escala, assim como para o tratamento e análise dos dados. Usando *abordagens* bioquímicas e bioinformáticas, foi possível identificar "candidatos" de elementos de DNA que co-determinam quando e onde um gene está ativo no corpo humano. (Fonte: http://www.alphagalileo.org/ViewItem.aspx?ItemId=123846&CultureCode=en)

(19)
A **necessidade** *de considerar materiais genético diferentes* também é salientada pelo **fato** de que variedades de muitas plantações [. . .] apresentam grande variedade de produção. Uma **possibilidade** é **que as fêmeas podem ser mais sensíveis às perguntas que os machos**.

Uma **alegação** recorrente é que o sistema de justiça criminal não valoriza a perspectiva das vítimas (Fonte: http://rspb.royalsocietypublishing.org/content/274/1608/303.full)

(20)

Priscilla Beaulieu Presley (nascida Priscilla Ann Wagner, em 24 de maio de 1945, no Brooklyn, em New York) é *uma* **modelo**, **escritora** and **atriz** americana e única **esposa** de Elvis Presley. Seu **pai** biológico, James Wagner, foi *um* piloto que morreu em *um* acidente de avião quando Priscilla era apenas *uma* **criança**. (Fonte: http://priscilla.elvispresley.com.au/)

(21)

- 1/2 xícara de **açúcar**
- 1.75 xícara de **farinha** comum
- 1 colher de sopa de fermento em **pó**
- 2 **ovos**, levemente batidos
- 1/2 xícara **leite**
- *Um* pouco menos que meio **tablete** de **manteiga** (110 g) — derretido
- muita essência de **baunilha** (3-5 colheres)
- muitos **mirtilos** (frescos/congelados . . . não importa)
- **Forno**: 200 °C

Unte uma **assadeira** para **muffin** com 12 forminhas com **spray** de untar. Misture os ingredientes secos (**açúcar**, **farinha**, fermento em **pó**) em *uma* **tigela**. Bata os ingredientes 'molhados' (**ovos, leite, manteiga**, essência) for cerca de *um* minuto ou dois, depois adicione os ingredientes secos. [. . .] (Fonte: http://crissybakes.wordpress.com/2012/03/09/blue-muffins-for-a-long-weekend/)