

Artigos

Considerações sobre a organização do texto e da instanciação sob a perspectiva sistêmico-funcional

On text organization and instantiation from a systemic functional perspective

Giacomo Figueredo¹
Gabriel Gomes Botelho Freitas²
Laura Scaramussa Azevedo³
Lucas Alexandre Damasceno⁴

RESUMO

Este trabalho considera o aspecto probabilístico da organização do texto a partir da instanciação. Cada instância manifesta de texto se associa à tipologia textual, tanto pelo contexto — como se registra amplamente na literatura — quanto pela configuração linguística mais provável, que é o foco deste trabalho. Descrever como a organização de probabilidade e frequência, assim como os mecanismos pelos quais as frequências emergem, motiva uma configuração linguística e determina como um texto é instanciado torna-se passo necessário para incluir propriedades quantitativas na modelagem. O artigo parte de um corpus de procedimentos, Subcorpus 1 — experimentos científicos e receitas; e Subcorpus 2 — referente à atividade social de “assar bolo”. Em seguida, identifica o perfil metafuncional dos textos e estabelece regras de associação entre os sistemas para identificar sua chavicidade. Por fim, aplica a metodologia da dinâmica de sistemas. Nos resultados, descrevem-se os mecanismos de subpotencialização, bem como determinam as probabilidades para uma instância ao mesmo tempo em que são constrangidos pela variação do registro. O trabalho aponta a

1. Universidade Federal de Ouro Preto, UFOP, Programa de Pós-graduação em Letras, Mariana, Brasil. <https://orcid.org/0000-0002-0967-244X>. E-mail: giacomo.figueredo@ufop.edu.br
2. Universidade Federal de Ouro Preto, UFOP, Programa de Pós-graduação em Letras, Mariana, Brasil. <https://orcid.org/0000-0001-9501-0671>. E-mail: gomesgabriel1997@gmail.com
3. Universidade Federal do Espírito Santo, UFES, Programa de Pós-graduação em Linguística, Vitória, Brasil. <https://orcid.org/0000-0002-1501-3993>. E-mail: laurasazevedo@outlook.com
4. Universidade Federal de Ouro Preto, UFOP, Programa de Pós-graduação em Letras, Mariana, Brasil. <https://orcid.org/0000-0003-3017-0599>. E-mail: lucas.damasceno@aluno.ufop.edu.br

alternativa complementar à análise dos gêneros para se compreender como os textos se organizam e são estruturados.

Palavras-chave: *linguística sistêmico-funcional; instanciação; dinâmica de sistemas; registro.*

ABSTRACT

The paper departs from instantiation to explore text in relation to the quantitative properties of language — more specifically how probability plays a part in the matching of an instance and a particular socially-oriented use. Such matching not only links instantiation and stratification, but also helps explaining how different context configurations (patterns of register and genre) are accordingly related to steps in subpotentialization. The fact that language is an inherently probabilistic system may yield a quantitative account of this process if a more complete modelling of language is to be obtained. As part of our exploration, we compiled a corpus of procedures and applied dynamics modelling to tease out quantitative properties of instances associated with context. We described some properties of instantiation — reweighting, association, and dynamics —, which determine the probabilities of an instance configuration, as they also behave following register variation. The paper goes on to show quantitatively how the process of subpotentialization is obtained, leading up to a complementary definition of ‘text’ as a probable unit of instantiation.

Keywords: *systemic functional linguistics; instantiation; system dynamics; text modelling; register.*

1. Introdução

The only generalization that can be made is that the normal state of a language is one of constant change (Halliday e McIntosh, 1966, p. 169).

Este artigo tem como objeto de investigação o texto a partir de sua organização sistêmica (Halliday, 1978; Martin, 1992; Matthiessen, 2015), segundo a qual os fenômenos da língua são interpretados, primeiramente, como fenômenos motivados pelo uso, para os quais a produção de significado linguístico possui orientação social (van Dijk, 1997; Painter, 1999). A produção de significado, por sua vez, não é determinada univocamente [f(língua_x) = contexto_x], mas admite um grau de liberdade tal que a ela confere sua natureza probabilística. A partir disso, compreende-se o texto como unidade

fundamental do discurso, pois é aquela com propósito comunicativo e social claramente definido pela relação provável entre a língua e seu contexto de produção (de Beaugrande, 1997; Martin & Rose, 2008). Adicionalmente, o texto é também uma unidade fundamental de probabilidade, ou seja, de uma configuração de escolhas que probabilisticamente se associam na forma de uma instância. O artigo se concentra precisamente nessa visão do texto como instância, bem como nos mecanismos que possibilitam as escolhas linguísticas se configurarem de maneira probabilística.

Uma decorrência dessa visão sobre o texto é que, dada a complexidade das atividades humanas e das relações sociais, os contextos sociais para a produção linguística variam assim como a função social do texto. Os textos (i.e., instâncias reais da unidade ‘texto’) se dispõem em diferentes tipos, constituindo uma tipologia de base discursiva, delimitada pelo contexto. Segundo essa visão, os textos se associam aos gêneros. Esses últimos, por sua vez, dariam propósito social aos textos, conferindo-lhes funcionalidade e motivação para o uso.

De maneira complementar, no que lhe concerne, essa variação pode também ser observada como um subpotencial, ou um conjunto de recursos prováveis que determina um agrupamento limitado (embora não determinado de forma específica) de opções a serem selecionadas nos sistemas. Desde o ponto de vista da variação, tanto a análise quanto a descrição do texto poderia, em um primeiro momento, prescindir do papel que a funcionalidade e o uso desempenham na sua organização. Isso porque o texto seria visto, a princípio, apenas como uma configuração de recursos prováveis, arregimentados segundo mecanismos de distribuição de frequências. Naturalmente, é fundamental ressaltar que essa visão não vai de encontro àquela. De maneira diversa, complementa-a na medida em que oferece configurações linguísticas padronizadas para as configurações recorrentes de uso.

Ademais, essa visão sobre o texto indica as principais características da dimensão da estratificação relativamente ao conceito de ‘texto’ e sua relação com a instanciação. O texto, observado a partir da relação entre estratos, torna-se um elemento fundamental para a explicação de diversos aspectos da metarredundância (Lemke, 1985), tais como a variação quanto ao uso (Firth, 1957; Matthiessen, 1993); a realização entre sistemas semióticos denotativo (língua) e conotativo (registro/gênero) (Lukin et al., 2008); e a própria realização entre os estratos da língua (semântica discursiva, gramática e fonologia). Por conseguinte, é possível definir de forma objetiva o conceito de ‘texto’ relativamente aos estratos do contexto - ‘registro’ e

‘gênero’: uma unidade linguística (texto) que realiza um conjunto de variações dessa produção linguística (registro) orientada contextualmente com propósito social definido (gênero), (Plum & Cowling, 1987; Nesbitt & Plum, 1988; Zappavigna & Martin, 2018) que se reconhece como diferente de outros conjuntos de variações que, por sua vez, possuem propósitos sociais distintos (Rose, 2019).

De modo complementar, quando o texto é observado do ponto de vista da subpotencialização, deve-se considerar que a tipologia textual demanda organização dos sistemas de toda a língua. Portanto, dessa maneira, independem do estrato ao qual esses sistemas se associam. O que importa é a forma como co-ocorrem e como se predizem uns aos outros, na forma como emergem em uma instância (provável e recorrente). Com isso, inclui-se nessa visão a partir da instanciação o fato de que todos os textos são gerados pelos mesmos recursos linguísticos, mas em proporções diferentes (Firth, 1957; Ellis & Ure, 1969; Martin & Christie, 1997). Mais importante ainda, algumas escolhas possuem mais peso do que outras na diferenciação entre os tipos de texto (cf. Lemke, 1995; Rose, 2019).

Por esse motivo, o paradigma que este artigo toma é que não basta estabelecer a relação entre estratos para se obter uma definição completa de ‘texto’, pois uma dada configuração linguística provável pertence, ainda, à esfera do que é possível, ou um “texto em potencial” (Martin, 2008). O mecanismo pelo qual esse potencial sistêmico se manifesta efetivamente na forma de uma unidade de texto se deve aos processos da instanciação (Halliday, 1991; 1992), a qual pode ser definida como o princípio que responde pelo desdobramento, no tempo, do potencial sistêmico nas seleções reais manifestadas como texto. É importante ressaltar que, caso uma determinada instância (i.e., configuração provável), ou um “texto em potencial”, venha a ser sancionada pela cultura e se tornar parte do comportamento linguístico como manifestação de um uso associada a um gênero, esse é um processo posterior à produção da instância. Afinal, uma instância somente poderia ser reconhecida, associada a um gênero e adotada por uma comunidade de falantes após ser produzida. É, por conseguinte, esse fato que permitiu a este artigo, nesse primeiro momento, optar pela escolha de investigar a instanciação de maneira independente da estratificação; e o texto independente do gênero ao qual se associa.

A questão que merece relevo nesse modelo — a qual motivou a pesquisa que aqui é apresentada — é que, apesar de identificada e teorizada, a dimensão da instanciação ainda carece das descrições de mecanismos pelos

quais esse processo de subpotencialização acontece; principalmente em sua relação com os tipos de texto (cf. Martin, 2009). Segundo os estudos nesse tema (Plum & Cowling, 1987; Halliday & James, 1993; Neumann, 2013), as descrições que podem aumentar a compreensão sobre esses mecanismos são importantes justamente pelo fato de que os recursos do sistema linguístico são os mesmos para todas as variações de texto, e o que os torna distintos são (i) a distribuição diferenciada desses recursos aliada a (ii) o peso que cada recurso adquire em (iii) um determinado momento no desdobramento temporal do texto. É nesse ponto que uma perspectiva probabilística/quantitativa se torna condição de entrada para a descrição da instanciação, as quais no futuro poderão ser associadas aos estratos e motivação funcional e do uso para que existam.

Este artigo tem o seguinte objetivo: apresentar uma descrição quantitativa do texto, de forma a defini-lo como unidade de variação das escolhas nos sistemas linguísticos. Em seguida, objetiva apresentar os mecanismos pelos quais as escolhas podem se tornar frequentes e formar padrões previsíveis, de forma a oferecer modelos que, posteriormente, poderão ser empregados como os recursos linguísticos para que cumpra um propósito social. Igualmente, visa explicar como esta descrição quantitativa contribui para a visão do texto como unidade de probabilidade e para a explicação do motivo pelo qual se divide em tipos.

Como é de conhecimento amplo, lidar com representatividade em amostras linguísticas, particularmente para aquelas com grande variedade de tipos de texto, não é uma tarefa de fácil execução. Por consequência, fazer asserções sobre “a língua toda” é um trabalho ainda mais árduo, ou mesmo inexecutável. Assim, este artigo será limitado pelo escopo do corpus investigado. Com isso, os objetivos de descrever o texto de maneira quantitativa e de apresentar os mecanismos de padronização das escolhas estão limitados ao corpus. Espera-se que a pesquisa em tela possa motivar estudos adicionais, oferecendo algum direcionamento descritivo e de metodologia para a identificação e análise das configurações linguísticas e de instanciação em outros contextos.

Para que se cumpram os objetivos, o *corpus* investigado é constituído de textos de procedimentos, divididos em dois *subcorpora*. Um primeiro, de procedimentos em geral, nomeado SC1, inclui textos dos domínios de experimentos científicos e receitas culinárias. Já o segundo subcorpus, SC2, segue critérios para que seja significativamente mais homogêneo, e se refere especificamente a um tipo de procedimento, a preparação de bolos — isto é, o domínio experiencial somente do procedimento de *assar bolo*.

Para tanto, parte das hipóteses seguintes: (a) uma abordagem quantitativa é necessária para a compreensão da instanciação, pois assim pode oferecer conhecimento para a maneira como se vincula à estratificação; (b) a tipologia textual contextualizada (i.e., mudança linguística determinada pela variação do contexto) é um efeito da redistribuição e associação que determinados recursos linguísticos assumem em um dado contexto; e (c) esse processo acontece de maneira dinâmica, no desdobramento temporal - o que significa que *redistribuição e associação não são propriedades dos sistemas, mas da instanciação*.

Metodologicamente, o artigo realiza a análise do perfil sistêmico de um *corpus* com tipos de texto semelhantes, aliada à modelagem dinâmica de como se desenvolvem no tempo. Assim, possibilita compreender a relação entre a variação contextual, a distribuição e as associações que são designadas aos recursos em diferentes momentos dos textos. Em seguida, compara as análises de perfil e modelagem e as relaciona a tipos de texto semelhantes, como forma de averiguar as hipóteses.

2. A caracterização do texto em uma abordagem sistêmico-funcional

Esta seção procura caracterizar o texto dando maior destaque à visão da instanciação. Contudo, devido à relação evidente entre o texto e a estratificação, busca-se aqui, em alguma medida, estabelecer os paralelos entre o texto como um a unificação semântica que gera unidades de discurso, que se associa aos gêneros; e o texto como uma emergência das configurações de probabilidade.

Do ponto de vista da organização discursivo-funcional (Halliday & Hasan, 1976; Halliday, 1978; Martin, 1992), o texto é o resultado da produção de significado concomitantemente em todos os sistemas da língua, incluindo-se, portanto, os níveis da fonologia, gramática e semântica-discursiva. Soma-se a isso o fato de que um sistema não produz significado de maneira isolada. Diferentemente, a escolha em um sistema depende de seleções em outros sistemas (Lemke, 1991; Halliday, 2005). Logo, a produção de significado é, na verdade, a unificação semântica de todas as expressões fonológicas de todas as realizações estruturais de todas as opções sistêmicas de todos os paradigmas da língua. À unidade resultante dá-se o nome de 'texto'; e é justamente a unificação semântica que confere ao texto a condição de unidade básica da língua.

Estrutura e coerência

A estruturação é o processo pelo qual a “unificação semântica” acontece. Cada estrato possui sua própria realização pela estrutura funcional. Assim, a língua é composta por sistemas fonológicos realizados por estruturas funcionais fonológicas; sistemas gramaticais realizados por estruturas funcionais gramaticais, etc. Quando “unificadas” na unidade ‘texto’, essas estruturas funcionais formam uma macroestrutura (estrutura de funções composta por outras estruturas). Ver exemplo 1, extraído e adaptado do texto [PB_EDP_2_E_MN] e Quadro 2.

- (1)
 Por que o botão grande cai dentro da garrafa?
 Ora, o botão grande cai é provavelmente por causa da inércia inicial que ele tinha.

Quadro 1 — Análise de alguns sistemas da segunda oração de (1) na matriz estrato-ordem-função

	Ora	o botão grande	cai	é	provavelmente	por causa	d[e]	a inércia inicial que ele tinha
CONEXÃO	[consequência:causa:esperada]							
NUCLEARIDADE		[nuclear]	[central]	[periférica]				
TAXONOMIA		[repetição]						
NEGOCIAÇÃO	[K1]							
E. DISCURS.	[demandar/informação/respondente]							
ARRANJO	[explícito periódico]							
CONTEXTUALIZ.	[continuidade:elaboração]							
TRANSITIVIDADE		Ator	[material]: Processo					
CIRCUNST.							[consequência]: Causa	
TEMA	[interp.]	[default]						
MODO	[ind.declarat.]							
	Sujeito	Finito/ Pred.						
MODALIDADE					[probabilidade]: Adjunto			
POLARIDADE	[positiva]							
VALIDAÇÃO	[verificação: conclusão]: Posicionador							
PREDICAÇÃO				[focalização: proeminência: circunstância]				
INFORMAÇÃO							[nova]: Novo	
ENTE		[comum:obj.]					[comum: abstração]	
CLASSIFICADOR								[conjunto]
QUALIFICADOR								[qualificado: oracional]
EPÍTETO		[descritivo: sensível]						
DÊIXIS		[específico]						

A macro-estruturação não diz respeito ao “empilhamento” de estruturas, uma vez que não é um processo aleatório; não se refere, pois, a uma estrutura semântica sobre uma estrutura gramatical. De outra forma, o aspecto macro-estrutural da unificação semântica promove a integração dessas estruturas, que é feita pela covariação entre seleções nos diferentes sistemas, que realizam cada estrutura (Martin, 2015). Covariar, por sua vez, significa que a escolha em um sistema A (realizada por sua estrutura funcional A) tem uma chance maior que o acaso de ocorrer junto com outra escolha no sistema B (realizada por sua estrutura funcional B). Caso o sistema A faça uma escolha diferente, então existe uma chance maior que o acaso que o sistema B faça uma escolha diferente também; isto é, os sistemas A e B variam juntos (Lemke, 1995).

A covariação pode ser vertical, por exemplo: uma opção de NEGOCIAÇÃO pressupõe a escolha de uma opção de FUNÇÃO DISCURSIVA, que pressupõe escolha no MODO e assim por diante. Ao mesmo tempo, pode ser horizontal; uma opção de MODO pressupõe uma de TRANSITIVIDADE, e outra de TEMA. Desde a estratificação, o que permite aos sistemas covariarem é, em última análise, o propósito social do texto (Martin, 2015; Rose, 2019). Em outras palavras, o texto deve possuir coerência (Halliday & Hasan, 1976; Martin, 1992; Martin & Zappavigna, 2018). Conseqüentemente, a coerência é o efeito da organização macro-estrutural pautada pelo propósito.

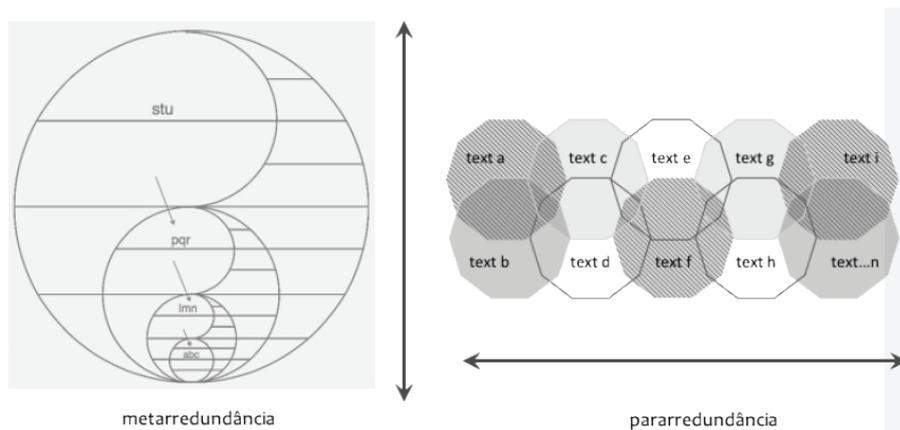
Do ponto de vista da instanciação, porém, persiste o problema de se saberem exatamente os mecanismos da unificação semântica em um texto em particular. Na estratificação, a unificação semântica trata abstratamente de como o texto é organizado, ou quando ‘texto’ pode ser definido como ‘macro-estrutura que realiza a covariação’. Não obstante, isso ainda não explica como uma macro-estrutura abstrata se manifesta na forma de um texto que existe (cf. Martin, 2008).

Nesse ponto, a visão quantitativa adotada neste artigo pode mais uma vez contribuir. Por esse motivo, é necessário, em primeiro lugar, entender o contexto não mais como um estrato, mas sim como um conjunto de sistemas que também fazem parte da covariação. Assim como a NEGOCIAÇÃO covaria com FUNÇÕES DISCURSIVAS e o MODO, outros sistemas (que denominamos contextuais), tais como os de STATUS e CONTATO e HONORIFICAÇÃO, também variam com aqueles. Da mesma maneira, cumpre também observar como a variação do contexto determina o princípio probabilístico (não-aleatório) das escolhas sistêmicas.

Instanciação e análise quantitativa

Se os estratos se associam por realização, as instâncias se associam por semelhança no comportamento. Assim, as instâncias se agrupam conforme o grau de generalização de suas configurações. Igualmente, a propriedade que permite a realização é a metarredundância (Lemke, 1989), ou a codificação da produção de significado em níveis cada vez maiores de abstração: {gênero realizado por {registro realizado por [semântica realizada por gramática (realizada por fonologia)]}}. O comportamento, por sua vez, é possível pela propriedade da pararredundância, ou o limite de uma região de probabilidade que estabelece o comportamento semelhante: {{{[(este texto) é receita de bolo] é receita} é procedimento} é informativo} (Figura 1).

Figura 1 — Metarredundância e pararredundância



Fonte: elaboração própria.

Devem-se ressaltar duas características importantes da pararredundância. A primeira é que nem todos os sistemas mudam completamente suas escolhas quando observamos dois tipos de texto. Muitos sistemas se conservam com escolhas e probabilidades semelhantes. Por exemplo, a probabilidade da escolha pelo modo interrogativo em uma receita de bolo (procedimento) e no resumo de artigo acadêmico (argumentativo) é basicamente a mesma (cf. Oliveira, 2022). A segunda é que a variação não é sempre linear. Por exemplo, os tipos de texto que utilizam mais imperativo não necessariamente se agrupam em oposição aos que utilizam indicativo, pois outras escolhas, em outros sistemas, sempre interferem na configuração dos tipos de texto.

Dessas duas características, é possível levantar a hipótese de que, dentre todas as escolhas feitas na instanciação de um texto, algumas são mais decisivas do que outras, por distribuírem os recursos e associá-los de forma a caracterizar o texto dentro de um tipo mais evidente ou reconhecível. Esse reconhecimento pode, de forma mais técnica, ser identificado por meio da probabilidade para além de um só sistema, mas de sua covariação (cf. Halliday, 1991; Lemke, 1995; Neumann, 2012; Martin, 2015). Para esse tipo de sistema e opções, utilizam-se os termos “sistema-chave” e “opção-chave”, de forma análoga às “palavras-chave” na metodologia da Linguística de *Corpus* (cf. Saioro, 2021). No caso dos sistemas, a chavicidade sempre é associada a outros sistemas por meio da covariação — daí o termo chavicidade associativa.

3. Metodologia

Metodologia de compilação do corpus

A compilação do *corpus* de procedimentos envolveu dois *subcorpora*. Um primeiro, para procedimentos em geral, nomeado SC1, incluiu textos dos domínios de experimentos científicos e receitas culinárias. Já o segundo subcorpus, SC2, foi significativamente mais homogêneo, referente à preparação de bolos — isto é, o domínio experiencial somente da atividade social de *assar bolo*.

A compilação para ambos os *subcorpora* foi realizada a partir de coleta de textos diversificada entre o livro de receitas *Dona Benta: comer bem, 76ª edição*, e o *website* de compartilhamento de receitas *TudoGostoso* e *websites* de experimentos científicos. Todos os textos coletados foram escritos originalmente em português brasileiro (Tabela 1). Foi considerada somente a etapa de execução do procedimento na compilação do corpus; listas de ingredientes e materiais, ou sugestões e dicas, não foram coletadas.

Tabela 1 — *Corpus* da pesquisa

	Procedimentos SC1	Receitas de bolo SC2	total
textos	30	20	50
tokens	1695	1031	2726
orações	291	145	436

Metodologia de análise do corpus

O primeiro passo metodológico para a análise foi a criação de uma tabela de anotação, em um software de planilhas eletrônicas (neste caso, *Excel* e *Google Sheets*). O modelo construído contou com 15 colunas, que representam opções dos sistemas analisados, inseridas por meio do recurso “validação de dados”, padronizadas para a análise de todos os textos que compuseram o *corpus*. Para cada oração, foram anotadas as escolhas em diferentes sistemas, como por exemplo, TEMA, PROCESSO, MODO, MODALIDADE, POLARIDADE, sistemas do grupo nominal, sistemas discursivos (TAXE, LÓGICO-SEMÂNTICO, MENSAGEM, CONEXÃO, IDEIAÇÃO), entre outros. Finalizada a anotação, foram concatenados os dados. Segue um exemplo de concatenação, para as 4 primeiras orações do texto PB_EDP_31_E_MN.

ORAÇÃO 1:

procedimento_principal_textcontinuativo_electtemainterp_temaprocess_imperativojussivo_selectmodal_positiva__selectvalid__selectvocat_materialtransforma_circmodo__selectcirc__selectcirc_

ORAÇÃO 2:

procedimento_principal_textconjuntivo_electtemainterp_temaprocess_imperativojussivo_selectmodal_positiva__selectvalid__selectvocat_materialtransforma__selectcirc__selectcirc__selectcirc_

ORAÇÃO 3:

procedimento_principal_selecttematext_electtemainterp_temaprocess_imperativojussivo_selectmodal_positiva__selectvalid__selectvocat_materialtransforma__selectcirc__selectcirc__selectcirc_

ORAÇÃO 4:

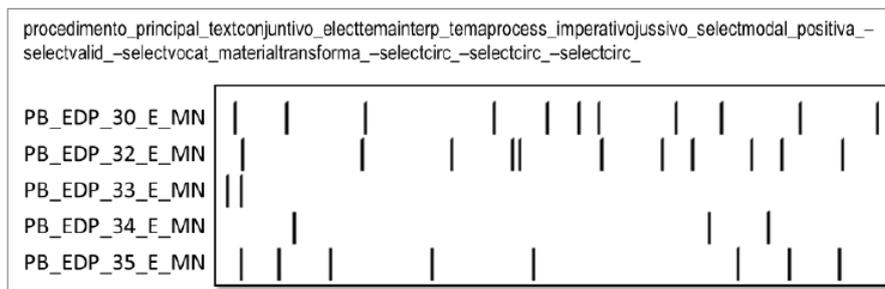
procedimento_subordinada_textconjuntivo_electtemainterp_temaP2_imperativojussivo_selectmodal_positiva__selectvalid__selectvocat_materialtransforma_circextens<o__selectcirc__selectcirc_

A partir da concatenação, gera-se um dado sobre a escolha nos sistemas. E, mais além, de como as escolhas em todos os sistemas se relacionam no “tempo” de 1 oração. Destaca-se que foram analisados sistemas gramaticais, mas igualmente discursivos e do contexto.

Em seguida, a concatenação foi transferida para o concordanciador *AntConc*, onde os textos foram analisados usando a função *Plot*, por meio da qual foi possível observar a distribuição das opções nos textos, além da quantidade de vezes que foram selecionadas. Assim, foi possível observar se essas co-ocorrências observadas em uma oração aconteceram apenas uma vez, ou se foram recorrentes ao longo do corpus, formando um padrão.

Ainda, por meio do *Plot*, foi possível observar em que “tempo” do texto esses padrões ocorrem (Figura 2).

Figura 2 — *Plot* de um padrão de co-ocorrência ao longo de 5 textos



Fonte: elaboração própria.

Como mostra a Figura 2, a concatenação em questão forma um padrão, que é recorrente e acontece em vários momentos dos textos. Por meio da concatenação e do *Plot*, foi possível analisar todas as escolhas no corpus completo para os sistemas analisados. O próprio AntConc consegue elaborar uma lista de todos esses padrões e mostrar quantas vezes apareceram, isto é, a sua frequência (é basicamente o mesmo mecanismo para contar tokens, pois uma concatenação funciona como um “token sistêmico”).

Para além desse ponto, não foi utilizada nenhuma outra ferramenta computacional. Os dados foram todos analisados manualmente. As ferramentas computacionais auxiliaram para registrar a anotação dos sistemas, concatenação e *Plot*, contagem da quantidade de vezes que cada padrão aparece. Os dados gerados se relacionam à análise linguística apenas no que se refere às escolhas nos sistemas. Assim, nesse momento, não estão relacionadas o significado dos textos, o seu uso ou lugar na cultura, nem ao seu comportamento como prática discursiva em uma comunidade de falantes. Desde o ponto de vista escolhido para o artigo, aquele da instânciação e da língua como um sistema probabilístico, os dados revelam apenas como um volume grande de escolhas em vários sistemas vão se repetindo em determinados momentos dos textos e, assim, vão formando padrões de escolhas recorrentes.

Metodologia de generalização e subpotencialização

Com os dados reunidos, seguiu-se à análise cruzada dos dados, visando avaliar a viabilidade da chavicidade associativa nos textos que compuseram o *corpus*. Para tal, utilizaram-se regras de associação (Hipp et al., 2000), cujo método possibilita descobrir relações entre variáveis identificadas em um *corpus*.

Não foi necessário aplicar testes estatísticos, uma vez que o *corpus* foi analisado em sua totalidade. Não se tratou de investigar uma amostra que infere sobre a população, pois investigou-se toda a população. A frequência das correlações encontradas diz respeito, de fato, às correlações que existem no *corpus*. Assim, os termos ‘frequência’ (quantidade de eventos reais ocorridos em um determinado período de tempo) e ‘probabilidade’ (formulação teórica para a chance de um evento ocorrer — i.e., a chance de a frequência se confirmar) levam ao mesmo resultado quando aqui aplicamos as regras de associação, uma vez que no *corpus* totalmente analisado a probabilidade sempre confirma a frequência.

As regras de associação são utilizadas para descobrir como ou por que determinados itens covariam — a sua probabilidade de associação. Mais especificamente, lançamos mão de três regras: **suporte**, **confiança** e **atração**.

Isso foi necessário pela seguinte razão. Os dados gerados pela concatenação e *Plot*, bem como a recorrência de padrões observada mostram quais são as co-ocorrências mais frequentes. Contudo, não explicam o motivo pelo qual essas são as mais frequentes. Além disso, não revelam como a escolha de um padrão — por exemplo, na oração 7 — pode interferir na formação de outros padrões — por exemplo nas orações 8, 9 e 10. As regras de associação são uma forma de depreender essas relações entre padrões.

Antes mesmo de utilizar as regras, é possível pensar em um exemplo. Como se sabe, para textos de procedimento, o modo [imperativo] é uma escolha muito recorrente. Uma vez que o imperativo não co-seleciona modalidade (ou seja, não existe a opção [imperativo: modalidade]), como é possível responder a seguinte pergunta: a escolha mais importante para o procedimento é selecionar [imperativo], ou não selecionar [modalidade]? Qual dessas duas escolhas é uma marca definidora do texto de procedimento?

Por meio da análise do discurso, da funcionalidade, dos usos e do contexto, é possível chegar à conclusão que a resposta é que o modo [imperativo]

é mais importante. Porém quando se consideram apenas as frequências e padrões, a resposta vem justamente das associações que essas duas escolhas fazem nos textos e no corpus como um todo. Nesse caso, as regras de associação mostram que o [imperativo] se associa fortemente a outros sistemas, como PROCESSO, CONJUNÇÃO, TEMA TEXTUAL, CIRCUNSTÂNCIA, ao passo que a não seleção da [modalidade] não tem nenhuma associação forte no corpus (ela é, por assim dizer, “aleatória”).

A seguir, apresentam-se as regras de associação que foram utilizadas na pesquisa em tela. Ressalta-se que, metodologicamente, os cálculos e contagens foram feitos de forma manual, para cada padrão do corpus.

O suporte de um item ‘x’ é a porcentagem das ocorrências em um *corpus* C que contém ‘x’. Sendo assim, o suporte da regra $x \rightarrow y$ em um *corpus* C é o suporte de $x \cup y$ em C, calculado através da equação apresentada em (1), na qual ‘ $x \cup y$ ’ indica o número total de orações em que x e y foram encontrados lado a lado para o número total de ocorrências de orações:

$$(1) \quad \text{Suporte} = \frac{\text{Freq}(x \cup y)}{\text{Total orações}}$$

Por exemplo, a covariação das opções [material transformativo] e [tema processo], no texto [PB_EDP_41_E_MN], que possui um total de 7 orações (equação 2). Encontrou-se que 6 das 7 orações apresentaram essa covariação, demonstrando assim um suporte de 85%.

$$(2) \quad \text{Suporte} = \frac{\text{Freq}(\text{mat. transf} + \text{tema proc.})}{\text{Total orações}} = \frac{6}{7} = 0,85$$

A confiança da regra $x \rightarrow y$ em um *corpus* C é a proporção do número de ocorrências em C que contém a interseção entre x e y, ou $x \cap y = \{a \mid a \in x, e a \in y\}$, com o número de ocorrências que contém ‘x’ em C (equação 3). Sendo assim, a confiança é a medida da frequência de vezes em que o consequente é verdadeiro quando o antecedente também é verdadeiro:

$$(3) \quad \text{Confiança} = \frac{\text{Freq}(x \cap y)}{\text{Freq}(x)}$$

Por exemplo, a covariação das opções [material transformativo] e [imperativo jussivo] no texto [PB_EDP_43_E_MN], que contém 9 orações,

demonstrou uma confiança de 100%. Ou seja, 100% das vezes que [material transformativo] foi selecionado, [imperativo jussivo] também o foi.

$$(4) \quad \text{Confiança} = \frac{\text{Freq}(\text{mat.transf} \cap \text{imp.juss})}{\text{Freq}(\text{mat.transf})} = \frac{9}{9} = 1$$

Por fim, a atração indica o valor de covariação do suporte de duas categorias, cujo objetivo é observar sua correlação (equação 5). Portanto, se itens 'x' e 'y' são independentes, ou seja, não há correlação entre eles, o valor da atração é de 1. A atração acima de 1 indica uma relação para a qual a correlação é positiva. Quando a atração é abaixo de 1, a relação indicada é de repulsão (i.e., correlação negativa).

$$(5) \quad \text{Atração} = \frac{\text{Suporte}(x + y)}{\text{Suporte}(x) \times \text{Suporte}(y)}$$

Por exemplo, a covariação das opções [tema textual conjuntivo] e [circunstância de local] no texto [PB_EDP_44_E_MN], demonstrou uma atração de 0,61. Ou seja, há uma relação negativa entre as categorias cruzadas (i.e., ocorrem menos frequentemente que a sorte).

$$(6) \quad \text{Atração} = \frac{\text{Suporte}(x + y)}{\text{Suporte}(x) \times \text{Suporte}(y)} = \frac{0,09}{0,145} = 0,62$$

Uma vez identificadas as regras de associação entre as categorias analisadas, tornou-se possível determinar diferentes aspectos da chavicidade dos sistemas. Tendo o objetivo de determinar o nível de associação considerado para a chavicidade, para a presente pesquisa, convencionou-se o primeiro grau de desvio padrão da distribuição normal como parâmetro, que é de 2/3 ou 67%. Dessa forma, foram consideradas as associações por suporte, confiança ou atração com valor maior ou igual a 0,67.

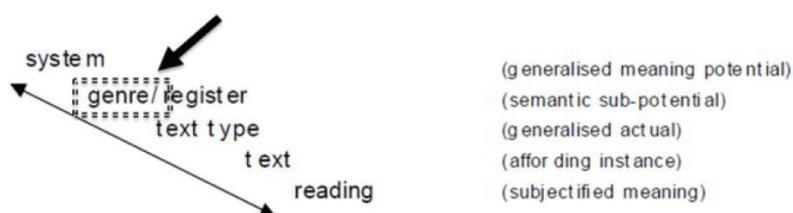
Dessa maneira, a chavicidade levou em conta não somente a probabilidade de uma opção ocorrer em um texto, mas também a maneira como está associada (por suporte, confiança e atração maiores que 67%) à probabilidade de ocorrência de outras opções.

Metodologia de modelagem

Como foi mencionado anteriormente, este artigo teve como objetivo analisar um corpus do ponto de vista da instanciação. Com isso, questões da estratificação (como análise do discurso, funcionalidade, contexto, gêneros) foram consideradas apenas marginalmente. Assim, os resultados que aqui serão apresentados, são atinentes somente a este corpus e, por conseguinte, não podem ser generalizados para toda a língua. Servem, contudo, como um conhecimento exploratório e uma indicação de como os sistemas podem estar associados e formar padrões também para a língua, extrapolando o corpus.

Tendo isso em mente, iremos aqui utilizar o termo ‘gênero’ para se referir aos padrões recorrentes que, ao se associarem fortemente, estabelecem uma estrutura de probabilidade de escolhas que foi característica do *corpus*. Dessa maneira, após o *corpus* analisado, para efeitos da pesquisa em tela, é possível considerar que a análise do *corpus* é, efetivamente, a análise do gênero. Isso porque o *corpus* se configura como a população inteira, indo além de uma amostra representativa, é a própria língua. Com isso, o que chamamos de ‘gênero procedimento’, ou ‘receita de bolo’, é, na verdade, o modelo global dos padrões de comportamento linguístico encontrado no *corpus*. A escolha pelo termo ‘gênero’ nesse caso decorre da terminologia empregada por Martin (2008, p. 33) quando se refere ao potencial semântico da instanciação (Figura 3).

Figura 3 — gênero como categoria da instanciação em Martin (2008, p. 33)



Fonte: Martin (2008, p. 33)

Os modelos para o gênero procedimento, a partir dos recursos linguísticos identificados na análise, foram estabelecidos segundo os critérios de frequência e relevância desses recursos. A frequência — aqui compreendida como uma ocorrência de redistribuição indica quais recursos foram utilizados e sua proporção. Já a relevância indica como os recursos interagem, tanto por covariação, quanto por chavicidade associativa.

Da mesma maneira que no caso da associação, também aqui determinamos o primeiro nível de desvio padrão, 67%, para indicar a frequência e relevância esperadas. Por conseguinte, qualquer recurso com probabilidade de ocorrência maior que 67%, ou a soma dos recursos mais frequentes até 67%, foi tomado para compor o modelo.

Em seguida, o ‘tempo’ foi incluído no modelo. Incluir o tempo como variável significa entender como cada porção de texto vai sendo produzida e acrescentada à estrutura textual, segundo uma sequência determinada. Para a presente pesquisa, a oração foi escolhida como a porção de texto. Com isso, o tempo passa a ser medido em orações. Um dos propósitos de se introduzir o tempo como variável do modelo é tornar possível a observação das probabilidades mais frequentes e relevantes de forma simultânea, para todos os recursos que são escolhidos durante o passar do tempo de uma oração. O outro motivo para a introdução do tempo é a possibilidade de se preverem escolhas futuras, baseadas em escolhas presentes.

Uma vez que as orações foram dispostas no tempo, calculou-se, subsequentemente, a diferença de *valeur* entre elas na ordem cronológica. Assim, foi possível observar a quantidade de significado de cada oração medida em *valeur*, a sequência temporal dos recursos dispostos entre as orações e também a variação de *valeur* no tempo. Nesse sentido, a variação de *valeur* passa a ser compreendida como um “movimento do texto” que se desloca no espaço.

4. Resultados

Perfil sistêmico dos subcorpora

Os textos de ambos os *subcorpora* — procedimentos em geral (SC1) e receitas de bolo (SC2) — pertencem ao tipo que denominamos ‘procedimento’. Cumpre destacar que, como é de conhecimento amplo, disseminado em diversas publicações (por exemplo, Martin e Rose, 2008), existe uma

expectativa para o padrão do gênero procedimento, do ponto de vista da estratificação. O que se espera com a análise dos textos de procedimento do ponto de vista da instanciamento é que, ao final, possam corresponder ao potencial sistêmico que instanciará o gênero.

Com isso, a expectativa confirmada era de que a configuração linguística contribuísse para que as etapas do gênero fossem cumpridas; mais especificamente a etapa que instrui sobre como uma atividade deve ser realizada. Neste caso, ‘etapas do gênero’, tal como interpretadas do ponto de vista da instanciamento, significa que há uma configuração de co-seleções entre sistemas que se destaca do ponto de vista de sua frequência relativamente a outras configurações. Em outras palavras, aquilo que se denomina ‘etapa do gênero procedimento’ por meio da análise do discurso, da funcionalidade e do contexto na estratificação, aqui esse rótulo é conferido a um padrão de probabilidade (ver Tabela 2).

Para ambos os *subcorpora*, a configuração ideacional mais frequente empregou orações materiais, com circunstâncias de localização. Do ponto de vista interpessoal, houve maior frequência de orações imperativas, com pouca modalidade ou avaliação. Finalmente, do ponto de vista textual, o uso de [tema textual] como forma de expandir atividades dentro de uma mesma sequência, quer por continuativos ou conjuntivos, teve frequência destacada (Tabela 2). Por exemplo:

- (7) SC1 [PB_EDP_8_E_MN]
Coloque cubos de gelo em um copo longo e decore
 Material Circ. Local Tema Textual Material
- (8) SC2 [PB_EDP_36_E_MN]
Por último Junte o conhaque, as passas e as frutas cristalizadas
 Tema Textual Material

Tabela 2 — Perfil sistêmico dos *subcorpora*

		freq. relativa SC1	freq. relativa SC2
oração	principal	69,2%	76,8%
	dependente	10,6%	6,8%
	hipotática	11,5%	6,8%
	reduzida	5,5%	9,6%
	não-oração	3,2%	—
		total 100%	total 100%
tema textual	textual		
	continuativo	37,9%	34,5%
	textual conjuntivo	62,1%	65,5%
		total 100%	total 100%
tema ideacional	processo		
	p1	58,4%	100%
	p2	33,0%	—
	c2	3,8%	—
	c2	4,5%	—
		total 100%	total 100%
modo	indicativo		
	declarativo	31,7%	0,8%
	imperativo jussivo	52,2%	99,2%
	imperativo	0,9%	—
	hortativo	—	—
	menor	15,2%	—
polaridade	positiva	95,2%	100%
	negativa	4,8%	—
		total 100%	total 100%
modalidade	probabilidade	60%	100%
	frequência	40%	—
		total 100%	total 100%

Na seção anterior, foi colocada a questão sobre quais são os mecanismos que possibilitam que um padrão se torne recorrente de modo que sua frequência caracterize um tipo de texto. Como consequência, colocou-se também a questão de quais são as escolhas mais importantes em sistemas específicos que possibilitam definir um tipo de texto (no caso em tela, o [imperativo], por exemplo; mas pode-se pensar em qualquer outro tipo de texto). A solução para essa questão, é importante reforçar, pode ser dada a

partir da estratificação, por meio da análise do discurso, da funcionalidade, dos usos e do contexto. De maneira complementar, a solução pode também ser alcançada quando se observam esses mecanismos pela sua recorrência nos textos, pela forma que essa recorrência se torna de distinta de outras recorrências de outros tipos de texto, bem como pelo momento no texto em que elas aparecem.

Destaca-se que esse é precisamente um dos objetivos deste artigo: apresentar os mecanismos pelos quais as escolhas podem se tornar frequentes e formar padrões previsíveis, de forma a oferecer modelos (no caso, dos textos que compõem os corpora analisados).

A Tabela 2 permite que isso seja precisamente observado no caso dos subcorpora investigados. A Tabela 2 mostra que, do ponto de vista da frequência relativa, houve uma distribuição reduzida no uso de recursos de SC2 relativamente a SC1. Da mesma forma, todos os recursos empregados em SC2 estão, necessariamente, presentes em SC1. Com isso, em SC1, as opções mais frequentes são aquelas que em SC2 se tornam as opções únicas ou possuem sua frequência aumentada. Esse é o primeiro mecanismo da instanciação em funcionamento: *os recursos do sistema linguístico são redistribuídos em padrões distintos*, e assim selecionados para realizar as combinações possíveis para cada texto.

Além disso, esse fato indica que, de SC1 para a língua a mesma hipótese pode ser levantada, para a qual esse mesmo processo se repetiria. Se a hipótese testada em Halliday e James (1993) for extrapolada para os dados desta pesquisa, a língua pode ser também comparada (Tabela 3). Halliday e James (1993) confirmaram a hipótese para a qual os sistemas linguísticos se comportam de duas formas: equiprováveis, quando a probabilidade de escolha de uma opção é igual a das outras opções (50%-50% para sistema de 2 opções; 33,33% para sistema de 3 opções e assim por diante); ou enviesados, quando a probabilidade de escolha de uma opção é muito maior que a das outras (90%-10%).

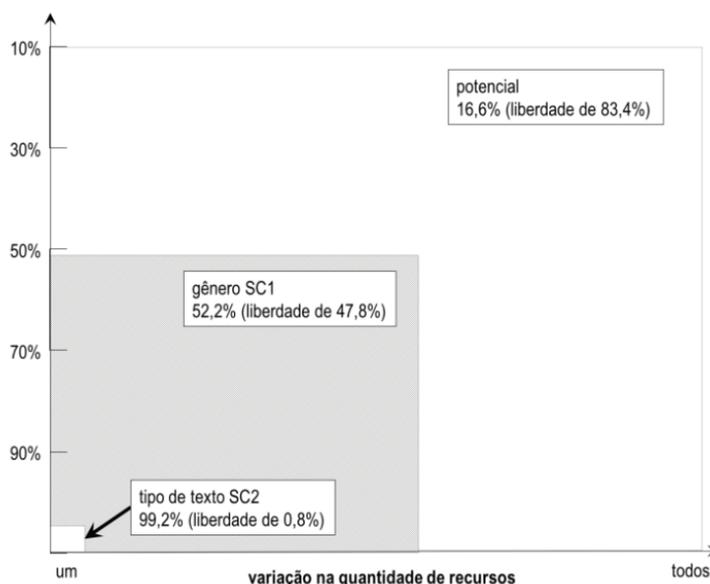
Tabela 3 — Comparação entre SC1 e 2 e hipótese equi-/viés para a língua

		freq. relativa SC1	freq. relativa SC2	freq. equi./ enviesada p/língua
tema textual	textual			
	continuativo	37,2%	34,5%	33%
	textual conjuntivo	62,1%	65,5%	33%
tema	processo			
ideacional		58,4%	100%	12,5%
	p1	33,0%	—	12,5%
	p2	3,8%	—	12,5%
	c2	4,5%	—	12,5%
modo	indicativo			
	declarativo	31,7%	0,8%	16,6%
	imperativo jussivo	52,2%	99,2%	16,6%
	imperativo	0,9%	—	16,6%
	hortativo	3,8%	—	—
	menor	15,2%	—	16,6%
modalidade	modalização	0%	0%	90%
	modulação	0%	0%	10%

A Tabela 3 revela o movimento de “afunilamento” na escolha de recursos quando os textos são mais específicos. Soma-se a isso o fato de que a subpotencialização acontece tanto entre recursos de um mesmo sistema — como no caso do MODO —, quanto entre sistemas — como no caso da MODALIDADE.

Destaca-se, ainda, que o “afunilamento” não é absoluto, mas indica uma probabilidade, que varia entre textos. A frequência relativa para, por exemplo, [imperativo jussivo] em um texto da língua selecionado aleatoriamente é de 16,6%; em SC1 é de 52,2%; e em SC2 é de 99,2% (Figura 4).

Figura 4 — Área da instanciação com exemplo de [imperativo jussivo] no *corpus*



Fonte: elaboração própria.

Um outro mecanismo importante da subpotencialização é a associação entre os sistemas e sua covariação; ou seja, a chavicidade que os sistemas vão assumindo à medida que o potencial “vai afunilando” para um determinado tipo de texto ou instância.

Durante a etapa de análise, foi observada a covariação entre todos os sistemas. Como é esperado acontecer na chavicidade, nem todas as covariações ocorreram, uma vez que nem todos os sistemas foram empregados (por exemplo, MODALIDADE ou VALIDAÇÃO). Com isso, o cruzamento de dados para o cálculo das regras de associação foi relevante justamente entre os sistemas que de fato são determinantes para os corpora em questão — a saber, aqueles cuja subpotencialização fez aumentar a frequência relativa quando associados. Após a combinação, foram encontrados 7 pares de sistemas que satisfazem essa condição (Tabela 4). As opções escolhidas para o cruzamento dos pares 1, 2 e 3 foram motivadas por serem as opções com maior frequência nos textos analisados, o que pôde ser identificado com a análise dos *Plots*.

De forma semelhante, para os pares 4, 5 e 6, foram selecionadas opções bastante proeminentes em todos os textos analisados (imperativo jussivo,

tema processo e polaridade positiva) junto a opções raramente ou nunca selecionadas (modalidade obrigação, vocativo chamado, validação persuasão) esperando-se que não houvesse associação entre elas. O par 7 foi escolhido pois as duas opções que o formam aparecem distribuídas semelhantemente nos textos.

Tabela 4 — Chavicidade associativa comparada entre SC1 e SC2

covariação	regra	SCI	SC2
material transf. + imperativo jussivo	suporte	49,50%	89,85%
	conf. mat.	57,06%	93,70%
	conf. imp.	63,33%	96,15%
	atração	0,61	1,00
material transf. + tema processo	suporte	58,86%	92,65%
	conf. mat.	80,78%	93,40%
	conf. tema	72,65%	100,00%
	atração	0,73	1,00
tema processo + imperativo jussivo	suporte	50,36%	88,55%
	conf. mat.	65,25%	93,05%
	conf. imp.	90,35%	96,65%
	atração	0,64	0,99
modalid. obrigação + imp. jussivo	suporte	0,00%	0,00%
	conf. mat.	0,00%	0,00%
	conf. imp.	0,00%	0,00%
	atração	0,00	0,00
vocativo chamado + tema processo	suporte	0,00%	0,00%
	conf. mat.	0,00%	0,00%
	conf. imp.	0,00%	0,00%
	atração	0,00	0,00
validação persuasão + polaridade positiva	suporte	0,00%	0,00%
	conf. valid.	0,00%	0,00%
	conf. pol.	0,00%	0,00%
	atração	0,00	0,00
tema textual conj. + circ. local	suporte	4,40%	9,98%
	conf. tema	12,50%	27,90%
	conf. circ.	23,50%	23,65%
	atração	0,23	0,35
covariações observadas	7 covariações		

Conforme a Tabela 4, o grau de associação entre os sistemas começa com um valor relativamente alto (maior que 49,5%) em todas as medidas e aumenta na comparação entre SC1 e SC2 (na maioria dos casos, em torno de 90%). Esse dado permite observar de uma maneira diferente o que, metaforicamente, tem-se nomeado de “afunilamento”. O que se observa é

o segundo mecanismo da instanciação: *o aumento no grau de associação entre sistemas-chave*.

Tendo em mãos o perfil metafuncional aliado à chavicidade dos sistemas, o passo seguinte foi observar como esses se dispõem ao longo do texto, na perspectiva dinâmica. Para que o desenvolvimento dos textos fosse observado ao longo do tempo, o primeiro passo foi construir um modelo dos *subcorpora* analisados.

O modelo dinâmico se fez a partir da produção de um texto sintético — i.e., o resultado da análise do *corpus* acrescido das probabilidades maiores relativas ao tipo de texto em questão. O texto sintético é a generalização das probabilidades maiores, associadas pela chavicidade. Serão apresentados os modelos de SC1 e SC2.

O modelo de SC1 possui 12 orações, para as quais se acrescentam as probabilidades de ocorrência conforme os dados na Tabela 5.

Tabela 5 — Modelo de SC1

Oração →	→	1	2	3	4	5	6	7	8	9	10	11	12
TAXE	principal	100	100	66	62	74	100	100	67	100	67	100	100
	dependente	0	0	16	16	0	0	0	12	0	9	0	0
	não-finita	0	0	0	0	13	0	0	15	0	18	0	0
TEMA TEXTUAL	não-seleciona	46	67	36	30	43	43	53	60	50	46	33	43
	continuativo	51	18	16	20	16	13	13	13	26	13	23	24
	conjuntivo	0	15	46	50	40	43	33	26	23	40	43	33
TEMA IDEACIONAL	processo	100	100	67	60	66	66	64	66	67	66	67	66
	p1	0	0	26	23	33	34	30	23	30	34	30	26
MODO	imp.jussivo	100	100	60	53	56	50	56	53	60	46	53	43
	indicativo	0	0	16	0	13	30	26	27	26	27	40	30
	dependente	0	0	23	33	30	20	16	20	13	27	0	26
TRANSITIVIDADE	material												
	transf.	100	100	100	100	100	67	66	66	67	60	67	67
	relacional	0	0	0	0	0	23	30	16	0	26	23	30
	existencial	0	0	0	0	0	0	0	13	6	10	0	0
CIRCUNST. 1	não-seleciona	60	60	60	65	65	50	60	60	55	60	60	60
	local	20	25	20	20	15	10	20	20	25	20	20	20
	modo	20	15	15	15	10	25	20	20	30	20	20	20

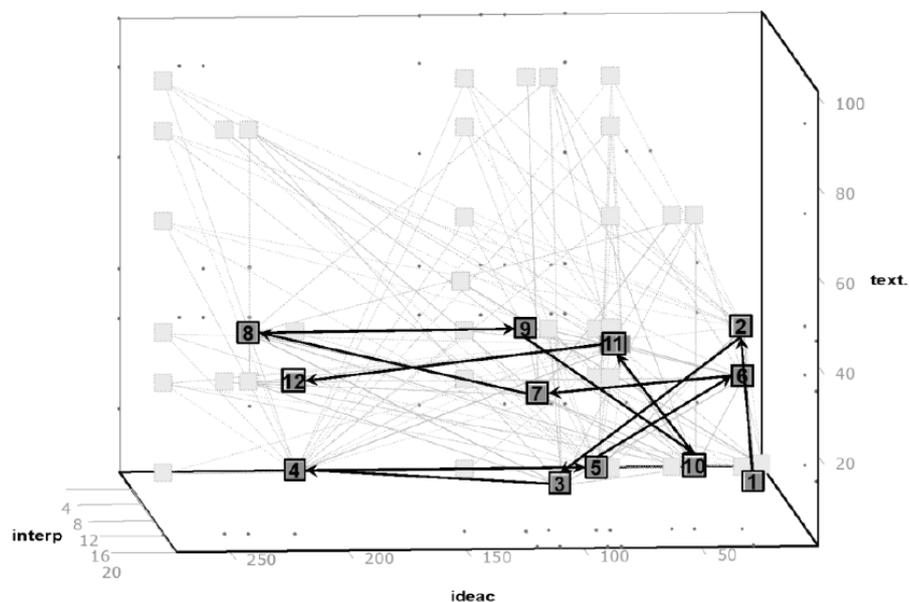
A Tabela 5 mostra os sistemas mais relevantes para SC1 — i.e., para os quais houve ocorrências — e os números nas células indicam a probabilidade de uma opção ocorrer. Por exemplo, para a oração 2, no sistema de TEMA TEXTUAL, existe a probabilidade de ocorrer [não-seleciona] 67%, [continuativo] 18% e [conjuntivo] 15%. Para o modelo, as probabilidades maiores

que 2/3 ou 67% foram consideradas como sendo o padrão para o *corpus*, e anotadas como 100%. Por exemplo, para a oração 1, no sistema de TAXE, a opção [principal] 100%. No caso das opções com probabilidade menor que 2/3 ou 67%, foram consideradas também as opções possíveis subsequentes.

A Tabela 5 mostra uma tendência geral de as orações seguirem a expectativa para as configurações linguísticas que, ao final, realizam o gênero procedimento. Igualmente, mostram também que pode haver variação nas escolhas dos sistemas em algumas das orações. Dinamicamente, o modelo de SC1 mostra o deslocamento no tempo do texto mais provável, e também as variações possíveis. Por exemplo, para a oração 8, no sistema de TRANSITIVIDADE, é mais provável ocorrer a opção [material transformativo] 66%; porém, também admite-se a variação incluindo [relacional] 16% e [existencial] 13%.

A Figura 5 apresenta o deslocamento no espaço gramatical para o modelo de SC1. Em cor mais forte, está uma opção de desenvolvimento de texto; as linhas e pontos semitransparentes indicam outras possibilidades para como o texto poderia se desenvolver.

Figura 5 — Deslocamento no espaço gramatical do modelo de SC1



Fonte: elaboração própria.

As questões decorrentes desses dados encerram a tentativa de entender: essa variação é motivada por algum mecanismo da instanciação? E, além disso, qual é o papel que variação e probabilidade desempenham no sistema linguístico, uma vez que a ele são inerentes?

Ao se abrirem mais opções para escolha, rapidamente obtêm-se um grande número de possibilidades para compor os textos. O cálculo da combinação simples entre os *valeurs* dos eixos x, y e z (cf. Figura 4) inclui as possibilidades de variação para cada uma das 12 orações, combinadas para x, y e z, o que dá: 10.405.955.952 possibilidades para o eixo x, 512 possibilidades para y e 544.195.584 para z. Isso indica que é potencialmente possível construir cerca de 2×10^{21} textos. O que essa enormidade de variações revela pode ser compreendido da seguinte maneira: o que percebemos da língua como sendo um ‘sistema probabilístico’, na verdade, são pequenas variações nas escolhas de alguma opção, em um ou outro sistema, que virtualmente nunca se repetem exatamente da mesma forma. A seguir, apresentam-se exemplos de pequenas variações em [PB_EDP_3_E_MS] a partir dos sistemas encontrados no *corpus*.

- (9a) Coloque a margarina e o chocolate
- (9b) E coloque a margarina e o chocolate
- (9c) Ainda, coloque a margarina e o chocolate
- (9d) Pode colocar a margarina e o chocolate
- (9e) Vai colocando a margarina e o chocolate
- (9f) Pode ir colocando a margarina e o chocolate
- (9g) Acrescente a margarina e o chocolate
- (9h) Acrescente a margarina e acrescente o chocolate

Contudo, desse número enorme, nem todas as opções de fato acontecem, pois uma determinada escolha não é completamente independente, mas influenciada por outras escolhas que a precedem. Além disso, embora essas pequenas variações sempre ocorram, quando se trata de chavidade associada, a tendência é de haver menor variação, e mais repetição. Ou seja, a variação é maior para sistemas menos relevantes no que diz respeito ao desenvolvimento do gênero; e é menor para os sistemas cruciais para o gênero (cf. Tabela 4).

Dessa forma, o deslocamento no espaço ao longo do tempo indica alguns padrões de movimento que coincidem com a disposição do gênero, assim como com o desenvolvimento do fluxo textual para os procedimentos. Por exemplo, para o TEMA TEXTUAL, nunca a oração 1 apresenta [conjuntivo].

Da mesma forma, as orações iniciais sempre escolhem para o MODO a opção [imperativo], assim como o TEMA IDEACIONAL, escolhe [tema processo]. Já para a TRANSITIVIDADE, a escolha por [relacional] acontece na segunda metade do texto. Esse é o terceiro mecanismo da instanciação: *a dinâmica determina o tipo de variação possível no desenvolvimento do texto, que é inversamente proporcional à chavicidade.*

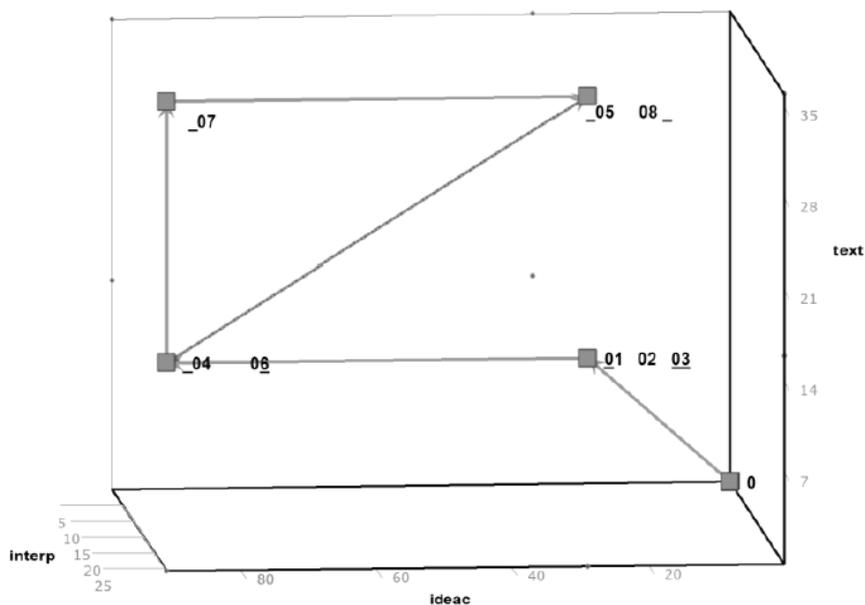
Passando ao modelo de SC2, esse foi composto por 8 orações, sendo que todas as opções-chave estavam associadas. Na oração principal, por exemplo, obteve-se tema ideacional processo, imperativo jussivo e circunstância de localização. A esses, acrescentam-se as probabilidades de ocorrência conforme os dados na Tabela 6.

Tabela 6 — Modelo de SC2

Oração →	→	1	2	3	4	5	6	7	8
TAXE	principal	100	100	100	100	100	100	100	100
TEMA TEXTUAL	não-seleciona	100	100	100	100		100		
	conjuntivo					100		100	100
TEMA IDEACIONAL	processo	100	100	100	100	100	100	100	100
MODO	imp.jussivo	100	100	100	100	100	100	100	100
TRANSITIVIDADE	material transf.	100	100	100	100	100	100	100	100
CIRCUNST. 1	não-seleciona	100	100	100		100			100
	local				100		100	100	

Todas as opções escolhidas em primeiro lugar obtiveram probabilidade maior que 67%. No caso de SC2, por consequência, todos os textos são iguais ao modelo; ou seja, não há variação entre as escolhas nos textos individualmente em relação à sua generalização. Por conta disso, poucos foram os recursos empregados em SC2, o que pode ser observado pelo movimento relativamente pequeno que seu desenvolvimento apresenta ao se deslocar pelo espaço gramatical (Figura 6).

Figura 6 — Deslocamento no espaço gramatical do modelo de SC2



Fonte: elaboração própria.

Mecanismos de instanciação e suas implicações para a estratificação

Quando observados dentro da arquitetura geral do sistema, os mecanismos exclusivos à instanciação — *redistribuição, associação e dinâmica* — se relacionam a todos os estratos. Observando essa relação “de cima”, o registro dispõe um número de configurações possíveis para a realização do gênero de maneira probabilística. Por exemplo, a diferença de *status* na sintonia para um mesmo tipo de texto, realizada por um comando no discurso para o gênero procedimento, pode ser realizada na gramática por:

- (10) SC1 [PB_EDP_6_E_MN]
Coloque a mistura na caneca
imperativo (52,2%)
- (11) SC1 [PB_EDP_1_E_MN]
O açúcar deve ser colocado a gosto
declarativo modulado (31,7%)

(12) SC1 [PB_EDP_1_E_MN]
Pode ir experimentando para saber o ponto
declarativo modalizado (4,7%)

(13) SC1 [PB_EDP_4_E_MN]
Separar quatro bolinhas de massa
menor (5,5%)

Em segundo lugar, a chavicidade pode ter mais ou menos força mesmo dentro de um único tipo de texto devido à essa variação relativamente ao contexto. Por exemplo, a chavicidade do MODO é sensível à variação na diferença de *status* da sintonia e, com isso, as opções [imperativo], [declarativo] ou [dependente] assumem maior ou menor força, como se vê no seguinte excerto:

(14) SC1 [PB_EDP_1_E_MN]
imperativo *Abra a panela*
imperativo *coloque o leite de vaca*
imperativo *e leve novamente ao fogo brando.*
imperativo *Coloque o restante dos ingredientes, menos o creme de leite.*
declarativo *O açúcar deve ser colocado a gosto.*
decl. modal *Pode ir experimentando*
dependente *para saber o ponto.*

Como mostra o exemplo, à medida que a diferença de *status* vai diminuindo em termos de expertise, a chavicidade da opção [imperativo] vai ficando “mais fraca”. No caso da dinâmica, tanto a redistribuição quanto a associação são dependentes do momento do texto, tanto por questões do desenvolvimento do campo quanto da negociação da sintonia.

Por fim, no que diz respeito à dinâmica, os modelos mostram que o desenvolvimento dos textos no tempo também se relaciona à configuração do registro e a forma como esse é realizado pela língua. Quanto menos recursos são chave, menor são o campo, a sintonia e o modo. Essa diferença fica clara quando se comparam os modelos dinâmicos de SC1 e SC2 (cf. Figuras 4 e 5). Se em SC1 existe uma variação a cada momento do texto, que leva a uma quantidade de possibilidades grandes nos momentos seguintes, em SC2 não existe variação, e o texto segue sempre o mesmo caminho em seu deslocamento pelo espaço gramatical.

Diante desses resultados, talvez a contribuição mais importante que o entendimento dos mecanismos de instanciação proporciona para a sua relação com a estratificação está em entender que todos os estratos instanciam (Martin, 2009), mas agora do ponto de vista da instanciação.

A partir da estratificação, dizer que cada estrato instancia significa compreender que os estratos não podem, por si sós, levar a um texto específico, ainda que se percorram os sistemas até suas opções mais delicadas, ou a realização até a expressão. Isso porque, o que cabe aos estratos é abstrair uma organização de “padrões de padrões” — o discurso é um padrão de padrões gramaticais; o registro é um padrão de padrões discursivos, etc. Por isso, uma instância qualquer inclui todos os padrões desde a expressão até o contexto (qualquer texto pode ser analisado desde a fonologia até o gênero).

A partir da instanciação, de outro modo, o que existem são associações mais prováveis entre sistemas, que se distribuem de formas distintas. Essas, por sua vez, acabam por gerar os estratos, que então podem ser entendidos como tipos de associação entre sistemas. Em outras palavras, o discurso é um padrão de padrões gramaticais porque um conjunto de sistemas (*a posteriori* nomeados de ‘discursivos’) mantém uma relação de chavidade associativa com um conjunto de sistemas (‘gramaticais’, *a posteriori*). Fenômenos como ‘*coupling*’ e ‘*commitment*’, identificados em Martin (2008), são exemplares desse processo.

5. Conclusão

Tendo como objetivo explicar como uma descrição quantitativa contribui para a visão do texto como unidade de probabilidade, visando ainda o definir como unidade de variação no emprego de recursos linguísticos, o artigo examinou a configuração linguística de dois *subcorpora* para observar princípios do funcionamento da instanciação, de maneira a revelar possíveis mecanismos de seu funcionamento.

Por meio das metodologias de associação e modelagem dinâmica, foram encontrados três mecanismos da instanciação. A redistribuição é o mecanismo que distribui os recursos de maneira enviesada nos diferentes registros (subpotenciais). Ela se deve ao fato de que os sistemas não são empregados de forma equiprovável em nenhuma situação real de uso da língua. Em seguida, a associação é o mecanismo que relaciona de forma

probabilística a escolha em diferentes sistemas, criando assim padrões de covariação. Uma vez que esses padrões também são redistribuídos, dão origem à chavicidade. Os sistemas-chave são fundamentais para a separação do potencial em registros; e esses em tipos de texto. Por fim, a dinâmica é o mecanismo que determina como a redistribuição irá ser associada ao longo do desenvolvimento do texto, no tempo.

Somando-se a isso, os resultados apontam que a probabilidade é uma propriedade da instanciação. Ademais, a natureza probabilística não é o resultado de uma “função aleatória intencional” do sistema linguístico. De outra forma, ela é o resultado dos trilhões de configurações possíveis para um texto (no universo de 2×10^{21} possibilidades de oração, a chance de dois textos serem exatamente iguais é nula).

Diante desses resultados, é possível indicar, ainda que parcialmente, alguns dos mecanismos que contribuem para a unificação semântica de um texto em particular, do ponto de vista da instanciação. Se na estratificação o ‘texto’ se define abstratamente como a macro-estrutura que realiza a covariação; na instanciação, ‘texto’ é *‘uma configuração de unidades linguísticas, distribuídas de forma probabilística, que possui sistemas-chave determinados por associação e se desenvolve no tempo’*.

Consequentemente, o conceito de instanciação pode ser reinterpretado como ‘a expressão da probabilidade dinâmica do texto’. Da mesma forma, reinterpreta-se a relação entre estratificação e instanciação. Se o olhar partir da estratificação indo para a instanciação, o que os resultados deste artigo mostram é como o processo de subpotencialização segue em paralelo ao aumento da delicadeza nos sistemas de todos os estratos, ou seja, quanto menor o potencial empregado em um tipo → subtipo → texto, maior o nível de delicadeza nos sistemas, chegando ao extremo de um texto sempre empregar as opções mais delicadas, tendo o número de opções reduzida em relação aos subtipos e tipos. Com isso, metáforas como “afunilar” podem ser substituídas pelos mecanismos de instanciação.

Já desde a instanciação, o que se tem é uma profusão de instâncias que a rigor variam sem obedecer qualquer norma de probabilidade pré-estabelecida. Afinal, o papel da instanciação é distribuir e redistribuir os recursos, criar associações e se desdobrar no tempo; mas nunca é determinar quais redistribuições ou associações deverão se tornar modelos de covariação (o papel da estratificação). Redistribuições e associações prototípicas só podem ser consideradas como tal sob a perspectiva complementar da estratificação.

Por fim, destaca-se que este é o ponto em que o artigo encontra suas limitações. Em primeiro lugar, os resultados apresentados são restritos ao corpus da pesquisa. Isso significa que esses mecanismos podem operar de maneira diferente para outros tipos de texto, ou mesmo não serem observados. Assim, pesquisas complementares são necessárias para que seja possível extrapolar os resultados da língua limitada a SC1 para a língua como um todo. Outras limitações incluem explicar o motivo pelo qual é a estratificação que determina quais modelos são possíveis e, assim, tendem a ser replicados e consolidados (será o acaso? Ou a relação entre língua e sociedade? Ou a ideologia? etc.). Assim, tendo como subsídio os resultados aqui apresentados, pesquisas no futuro poderão explorar com maior detalhe as implicações dos mecanismos de instanciação para a economia geral do sistema linguístico.

Agradecimentos

Agradecemos o apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico — CNPq no financiamento a esta pesquisa. (Processo CNPq: 313627/2020-7).

Conflito de interesses

Declaramos não ter qualquer conflito de interesse, em potencial, neste estudo.

Contribuições dos autores

Nós, Giacomo Figueiredo, Gabriel Freitas, Laura Scaramussa e Lucas Damasceno declaramos, para os devidos fins, que não temos qualquer conflito de interesse, em potencial, neste estudo. Todos nós participamos da conceptualização do estudo, metodologia, desenho do estudo, análise dos dados, coleta e geração de dados, escrita do texto do artigo. Todos os autores aprovam a versão final do manuscrito e são responsáveis por todos os aspectos, incluindo a garantia de sua veracidade e integridade.

Disponibilidade de dados

Todo o conjunto de dados que dá suporte aos resultados deste estudo foi publicado no próprio artigo.

Referências

- Beaugrande, R. de. (1997). Society, education, linguistics and language: inclusion and exclusion in theory and practice. *Linguistics and Education*, 9(2), 99-158. [https://doi.org/10.1016/S0898-5898\(97\)90011-5](https://doi.org/10.1016/S0898-5898(97)90011-5)
- Ellis, J., & Ure, J. (1969). Language varieties: register. In A.R. Meetham, & R. Hudson (Eds.), *Encyclopaedia of Linguistics, Information, and Control* (pp. 251-259). Pergamon Press.
- Firth, J.R. (1957). Modes of Meaning. In Palmer, F.R. (Ed.), *Papers in Linguistics*. Oxford University Press.
- Halliday, M.A.K. (1978). Ideas about language. *Journal of the Sydney University Arts Association*, Sydney, 11, 20-38.
- Halliday, M.A.K. (1991). Towards probabilistic interpretations. In Elija Ventola (Ed.), *Functional and Systemic Linguistics: approaches and uses* (pp. 39-61). De Gruyter Mouton. <https://doi.org/10.1515/9783110883527>
- Halliday, M.A.K. (1992). How do you mean?. In M. Davies, & L. Ravelli (Eds.), *Advances in Systemic Linguistics: Recent Theory and Practice* (pp. 20-25). Pinter Publishers.
- Halliday, M.A.K., & James, Z.L. (1993). A quantitative study of polarity and primary tense in the English finite clause. In J. Sinclair, M. Hoey, & G. Fox (Eds.), *Techniques of description: Spoken and written discourse* (pp. 32-66). Routledge.
- Halliday, M.A.K. (2005). On matter and meaning: the two realms of human experience. *Linguistics & the Human Sciences*, 1(1), 59-82. <https://doi.org/10.1558/lhs.2005.1.1.59>
- Halliday, M.A.K., & McIntosh, A. (1966). *Patterns of language: papers in general, descriptive and applied linguistics*. Longmans, Green and Co. Ltd.
- Halliday, M.A.K., & Hasan, R. (1976). *Cohesion in English*. Longman.
- Hipp, J., Güntzer, U., & Nakhaeizadeh, G. (2000). Algorithms for association rule mining: a general survey and comparison. *ACM sigkdd explorations newsletter*, 2(1), 58-64. <https://doi.org/10.1145/360402.360421>
- Lemke, J. (1985). Ideology, intertextuality, and the notion of register. In J. Benson, & W. Greaves (Eds.), *Systemic Perspectives on Discourse* (pp. 275-294). Ablex Publishing.
- Lemke, J. (1989). Making text talk. *Theory into practice*, 28(2), 136-141. <https://doi.org/10.1080/00405848909543392>
- Lemke, J. (1991). Text Production and Dynamic Text Semantics. In Elija Ventola (Ed.), *Functional and Systemic Linguistics: Approaches and Uses* (pp. 23-38). Mouton de Gruyter. <https://doi.org/10.1515/9783110883527>
- Lemke, J. (1995). Intertextuality and text semantics. In P. Fries, & M. Gregory (Eds.), *Discourse in society: Systemic functional perspectives* (pp. 85-114). Ablex Publishing.

- Lukin, A., Moore, A., Herke, M., Wegner, R., & Wu, C. (2008). Halliday's model of register revisited and explored. *Linguistics & the Human Sciences*, 4(2), 187-243. <http://dx.doi.org/10.1558/lhs.v4i2.187>
- Martin, J.R. (1992). *English Text: System and structure*. John Benjamins Publishing Company. <https://doi.org/10.1075/z.59>
- Martin, J.R. (2008). Tenderness: Realisation and instantiation in a Botswanan Town. *Odense Working Papers in Language and Communication*, 29, 30-58.
- Martin, J.R., & Christie, F. (1997). *Genre and Institutions: Social Processes in the Workplace and School*. Cassell.
- Martin, J.R., & Rose, D. (2008). *Genre Relations: Mapping Culture*. Equinox.
- Martin, J.R. (2009). Realisation, instantiation and individuation: some thoughts on identity in youth justice conferencing. *D.E.L.T.A., Especial*, 549-583. <https://doi.org/10.1590/S0102-44502009000300002>
- Martin, J.R. (2015). Meaning beyond the clause: Co-textual relations. *Linguistics and the human sciences*, 1(2-3), 203-235. <https://doi.org/10.1558/lhs.34711>
- Matthiessen, C.M.I.M. (1993). Register in the round: diversity in a unified theory of register analysis. In Mohsen Ghadessy (Ed.), *Register analysis: theory and practice* (pp. 221-292). Pinter.
- Matthiessen, C.M.I.M. (2015). Register in the round: registerial cartography. *Functional Linguistics*, 2(1), 1-48. <https://doi.org/10.1186/s40554-015-0015-8>
- Nacional, Editora (2011). *Dona Benta: Comer bem*. Editora Nacional.
- Nesbitt, C., & Plum, G. (1988). Probabilities in a systemic-functional grammar: The clause complex in English. *New developments in systemic linguistics*, 2, 6-38.
- Neumann, S. (2012). Applying register analysis to varieties of English. In *Anglistentag 2011 Freiburg Proceedings* (pp. 75-94). Microsoft Word - Sektion I_WVT.docx (researchgate.net) (accessed 21 April, 2022).
- Neumann, S. (2013). *Contrastive Register Variation: A Quantitative Approach to the Comparison of English and German*. De Gruyter Mouton.
- Oliveira, F. S. (2022). *O discurso do autocuidado em saúde: uma descrição de gêneros na covariação experto-leigo* [Tese de doutorado]. Universidade Federal de Minas Gerais.
- Painter, C. (1999). *Learning through Language in Early Childhood*. Cassell.
- Plum, G., & Cowling, A. (1987). Social constraints on grammatical variables: Tense choice in English. In R. Steele & T. Threadgold (Eds.) *Language Topics: Essays in honour of Michael Halliday* (pp. 281-306). John Benjamins. <https://doi.org/10.1075/z.lt2>
- Rose, D. (2019). *Selecting & Analysing Texts*. Reading to Learn.
- Saioro, R. S. B. B. (2021). *A construção do discurso científico: os gêneros do discurso científico no português brasileiro* [Dissertação de mestrado]. Universidade Federal de Ouro Preto.

Tudo Gostoso Internet Ltda. (2020). Tudo Gostoso. <https://www.tudogostoso.com.br/>. Accessed 23 July 2020.

Van Dijk, T. A. (1997). The Study of Discourse. In T. A. van Dijk (Ed.), *Discourse as Structure and Process* (pp. 1-34). SAGE Publications. <https://doi.org/10.4135/9781446221884>

Zappavigna, M., & Martin, J.R. (2018). *Discourse and Diversionary Justice: An Analysis of Youth Justice Conferencing*. Palgrave Macmillan. <https://doi.org/10.1007/978-3-319-63763-1>

Recebido em: 14.12.2022

Aprovado em: 18.03.2024