

Uma metodologia de perfilação gramatical sistêmica baseada em *corpus*

Towards a corpus-based methodology for grammatical systemic profiling

Giacomo Figueredo *

RESUMO: Embasado nos pressupostos da Linguística de *corpus* (LC), este artigo apresenta uma metodologia de investigação de funções gramaticais, a qual inclui as etapas de compilação, extração e análise de dados da gramática. Denominada perfilação gramatical, a metodologia apresentada neste artigo possibilita a identificação de padrões e subsequente descrição, principalmente da forma como a gramática é empregada na organização do texto – a dinâmica textual. A perfilação gramatical lança mão da teoria sistêmico-funcional devido ao fato de essa teoria abranger as relações entre a gramática, sua expressão e a organização do texto. Para a perfilação gramatical deste artigo, foi utilizado como Exemplo de Pesquisa um *corpus* composto por dez minibiografias produzidas em português brasileiro. Os passos da perfilação incluem um mapeamento das funções gramaticais, o estabelecimento da distância topológica entre as funções e, por fim, o movimento do emprego das funções no espaço gramatical – constituindo assim a perfilação. Como resultado, este artigo apresenta a forma pela qual a metodologia da perfilação gramatical pode contribuir para o desenvolvimento das pesquisas de *corpus*, em particular quando o foco é a gramática. Especificamente, a perfilação gramatical se mostra efetiva na compilação e análise de *corpora* para além da relação item lexical/token, compreendendo igualmente categorias teóricas/funções gramaticais. Demonstra ainda como é possível compilar e extrair dados de *corpora* inclusive para a forma como são empregados na dinâmica textual, o

ABSTRACT: This paper develops a methodology to investigate grammar functions based on corpus Linguistics. The methodology is grammar profiling and includes compiling, retrieving and analyzing data from the grammar. Grammar profiling is shown to be a useful methodology for handling grammar patterning, especially when dealing with such patterning in the development of text. Grammar profiling is grounded on systemic functional theory for it needs a theory comprehensive enough to deal with grammar metaredundancy in relation to both text organization and expression. The grammar profiling of this paper is applied to investigate a corpus of ten mini-biographies in Brazilian Portuguese as a Research Example. The steps taken during profiling identify grammar functions, establish their topological distances and places them within the grammar space. Results show that grammar profiling is a useful methodology for corpus research, especially when the need is for a non-lexical corpus – one made up of functions – is investigated. It is also useful for compiling corpora and retrieving data from text development, capturing not only corpus synopsis, but also text dynamics.

* Doutor em Linguística Aplicada. Universidade Federal de Ouro Preto.

que, de outra forma, não seria possível.

PALAVRAS-CHAVE: Perfilação gramatical. Metodologia baseada em *corpus*. Dinâmica de sistemas. Espaço gramatical.

KEYWORDS: Grammatical profiling. corpus-based methodology. Systems dynamics. Grammar space.

1. Introdução

Este artigo focaliza as relações entre a metodologia advinda da Linguística de *Corpus* (LC) e a análise dos sistemas gramaticais. Motivado pela investigação dos fenômenos linguísticos, em particular de sua organização paradigmática (ou sistêmica) de forma objetiva e replicável, este artigo pauta-se pelos pressupostos da Linguística de *Corpus* (LC) para apresentar uma metodologia de investigação de funções gramaticais. Toma-se aqui como base a constante preocupação da LC de investigar a língua para além da intuição do falante/analista (VIANA, 2011), bem como o estabelecimento de metodologias confiáveis de compilação, extração e análise de dados que busquem garantir resultados uniformizados. A metodologia apresentada neste artigo, denominada perfilação gramatical, possibilita a compilação de *corpora* de funções gramaticais que, posteriormente, podem ser analisadas para a busca de padrões e subsequente descrição da língua.

A necessidade da perfilação gramatical para os estudos afiliados à LC surge do fato de que a maior parte das funções gramaticais, diferentemente dos itens lexicais, não é realizada de forma aparente pela expressão (i.e., típica, aberta), mas sim pela relação entre itens linguísticos (i.e., criptotípica, encoberta) (WHORF, 1987). Diferentemente dos itens lexicais, que são identificados e recuperados com facilidade – uma vez que são realizados por palavras, codificadas como *tokens* – as funções gramaticais, por comparação, raramente são realizadas por palavras, especialmente quando estas cumprem o papel de organizar o texto (CONRAD, 2010). Por tal motivo, essas últimas requerem a anotação manual do analista e dificultam a tarefa do computador na extração de dados.

Na perfilação gramatical, a anotação do *corpus* para funções gramaticais torna-se um passo metodológico, a qual abre a possibilidade para a compilação de *corpora* de funções gramaticais (cf. PAGANO *et al.*, 2014). Dessa forma, a perfilação gramatical contribui para o reconhecimento de padrões que, de outra forma, não seriam reconhecidos, o que tornaria a análise assim suscetível à variação conforme o analista (BIBER *et al.*, 2004). Igualmente, a perfilação gramatical auxilia a busca automática pelo computador, convertendo as funções gramaticais em “*tokens* gramaticais”.

Como aporte teórico, a perfilação gramatical baseia-se na Linguística Sistêmico-Funcional (doravante LSF) (cf. HALLIDAY e MATTHIESSEN, 2004). Na qualidade de teoria abrangente, a LSF proporciona um arcabouço de descrição e análise do sistema linguístico como um todo, em seus diferentes estratos, de forma que a gramática pode ser examinada tanto em relação aos elementos que a expressam quanto ao texto que organiza. Além disso, a LSF privilegia a organização paradigmática da língua, dando ênfase à disposição sistêmica das funções gramaticais, as quais são o foco do presente artigo.

Para a perfilação gramatical deste artigo, como Exemplo de Pesquisa, será utilizado um *corpus* composto por dez minibiografias de matemáticos importantes, extraídas de livros didáticos de matemática publicados em português brasileiro (PB). A metodologia de análise inclui um mapeamento das funções gramaticais, o estabelecimento da distância topológica entre as funções e, por fim, o movimento do emprego das funções no espaço gramatical relativamente à dinâmica textual (LEMKE, 1991).

Com isso, este artigo espera demonstrar como a metodologia da perfilação gramatical pode contribuir para o desenvolvimento das pesquisas de *corpus*, em particular quando o foco é a gramática. Especificamente, a perfilação gramatical permite a compilação e análise de *corpora* para além da relação item lexical/*token*, compreendendo igualmente categorias teóricas/funções gramaticais.

2. Análise linguística e análise de dados linguísticos pelo computador

As metodologias advindas da LC possibilitam a compilação de amostras representativas da língua – i.e., *corpora* – bem como a identificação de padrões dos itens linguísticos (cf. VIANA, 2011). As metodologias da LC se relacionam ao uso do computador devido ao fato de os *corpora* serem armazenados em formato eletrônico e os procedimentos de investigação envolverem o uso de *softwares* (TOGNINI-BONELLI, 2001).

Nesse processo, contudo, cabe distinguir a análise linguística, por um lado, da análise de dados linguísticos pelo computador, por outro. Isso se faz necessário para evitar o emprego de *tokens*, concordâncias e frequência de itens computacionais como explicação para os fenômenos linguísticos (PAGANO *et al.*, 2014).

O computador consegue acessar o *corpus* apenas indiretamente, por meio de uma versão dos itens da linguagem natural codificada em uma forma que possa ser lida por *softwares*. Por exemplo, a palavra (unidade gramatical constituída por morfemas) é

“traduzida” como um *token* (conjunto de caracteres que o *software* consegue ler). Por conseguinte, os itens da língua natural são etiquetados; e as etiquetas, juntamente com os padrões que formam, é que são analisadas. Uma vez que o computador processa dados linguísticos indiretamente, *tokens*, linhas de concordância e frequências de ocorrência, por si só, não explicam fenômenos linguísticos (cf. KE, 2012). Nesse ponto, a LC proporciona os meios pelos quais as etiquetas do computador podem ser “retraduzidas” para a linguagem natural.

Do ponto de vista dos estudos de *corpora*, Tognini-Bonelli (2001) detalha a distinção entre as pesquisas baseadas em *corpus* (*corpus-based*) e aquelas direcionadas pelo *corpus* (*corpus-driven*). Os estudos baseados em *corpus* servem como forma de validar/refutar princípios da teoria, ao passo que os estudos direcionados pelo *corpus* generalizam os resultados, contribuindo para a construção gradual da teoria. A busca pelos padrões linguísticos que não são típicos ou abertos, na forma de itens isolados ou estruturas, faz com que a teoria reflita mais de perto os dados (cf. HALLIDAY e JAMES, 1993).

A trajetória desde uma abordagem baseada em *corpus* para outra direcionada pelo *corpus* – na qual os resultados extraídos pelo *software* são reinterpretados em termos linguísticos – abre a possibilidade da construção de uma “microteoria” do *corpus* que permite sua análise. Nesse processo, os dados extraídos pelo *software* – linhas de concordância, listas de palavras, colocações, etc. – são reinterpretadas em termos linguísticos. Os *types* e *tokens* revelam padrões da classe gramatical ‘palavra’; a colocação revela padrões da classe semântica ‘colocação lexical’; as linhas de concordância revelam padrões da gramática criptotípica; e assim por diante.

Do ponto de vista da LSF, Halliday (1992) faz a distinção entre o estrato da organização e o estrato da expressão no sistema linguístico. Os padrões extraídos pelo *software* são relativos ao estrato da expressão da linguagem natural – em particular na forma grafológica. Com isso, o estrato da expressão é, por sua vez, a manifestação da organização linguística, a qual pode ser modelada desde os padrões sistêmicos mais gerais – a gramática – até itens específicos – o léxico. A investigação do *corpus*, por meio da perfilação gramatical revela a organização linguística codificada nos dados extraídos pelo *software*.

Com isso, (i) o intercâmbio entre teoria e dados (estudos baseados em *corpus*/estudos direcionados pelo *corpus*) e (ii) a mudança do estrato da expressão (conjunto grafológico arbitrário) para o estrato da organização (a gramática e o léxico) se tornam os passos adotados

para se extraírem dados linguísticos do *corpus* por meio do computador. Quando o interesse da análise se concentra na gramática, inclui-se um novo passo, (iii) passar das ocorrências de itens específicos, o léxico, para os padrões mais gerais de organização.

Em certo sentido, a investigação é mais “fácil” (HALLIDAY, 1992, p. 64) quando (i) parte da abordagem baseada em *corpus* para a direcionada pelo *corpus*; (ii) concentra-se no estrato da expressão; e (iii) limita-se ao léxico. Por outro lado, quando a investigação tem como objeto a gramática, outros aspectos devem ser levados em conta.

Os itens lexicais constituem a região mais delicada (i.e., detalhada, específica) dos sistemas gramaticais, e por isso abarcam os itens mais distintos. Como tal, podem ser, em geral, considerados itens únicos. Um item lexical é, portanto, um item específico da gramática, em geral realizado por um elemento da ordem da palavra (ou, com menor frequência, da ordem do grupo). Consequentemente, sua expressão também é única. Esta é, via de regra, a definição do item lexical: a sobreposição de um significado em um item gramatical em uma expressão (HALLIDAY, 2002).

Na investigação do *corpus*, o computador consegue identificar com maior facilidade os itens lexicais justamente por causa de sua forma particular. Por exemplo, uma lista de palavras demanda uma quantidade menor de etapas na análise lexical (*tokenização*) do que uma lista de Sujeito Indeterminado, ou de Tipo de Processo na análise sintática/gramatical (*parsing*) (cf. CÂNDIDO-JR, 2013).

Os itens gramaticais são generalizáveis, uma vez que constituem recursos empregados em uma quantidade maior de situações na produção de significado. Com a exceção de funções gramaticais específicas – como, por exemplo, a Pluralidade e o Tempo Verbal – os itens gramaticais não possuem forma ou expressão únicas. Ao contrário, as funções gramaticais tendem a ser realizadas por um agrupamento de formas, tais como estruturas, partículas, prosódias ou pela recorrência de itens em conjunto (HALLIDAY, 2002).

Diante dessas considerações, uma abordagem mais efetiva deve “selecionar os sistemas fundamentais para uma teoria gramatical probabilística”^{*1} (HALLIDAY, 1992, p. 64). De determinada maneira, a análise gramatical pode ser tomada como um processo guia para a LC, tanto na etapa de compilação do *corpus*, como também na etapa de análise dos dados. A noção de ‘*corpus*’ então se amplia, abrangendo a expressão e os itens lexicais

* Todas as traduções apresentadas são de autoria própria.

¹ [...] select systems that could be critical for a probabilistic grammatical theory.

(conjuntos de caracteres no computador), bem como as funções gramaticais realizadas pela expressão (as etiquetas dos conjuntos de caracteres).

Com o objetivo de aplicar a perfilação gramatical ao *corpus* de pesquisa, passa-se aos pressupostos teóricos desta, a partir da LSF.

3. Organização do sistema linguístico e funções gramaticais

A arquitetura sistêmica da língua (HALLIDAY, 2002) é uma proposta abrangente na qual a descrição de cada fenômeno contribui para a descrição do sistema como um todo. A justificativa de um modelo sistêmico é dada por HALLIDAY e MATTHIESSEN (2004, p. 20):

São muitas as razões para a adoção de uma perspectiva sistêmica. Uma delas é que as línguas não são projetadas, mas evoluem. Ademais, os sistemas evolutivos não podem ser explicados simplesmente como uma soma de partes. Nosso pensamento tradicional sobre a linguagem, que é composicional, deve ser, se não substituído, pelo menos complementado por um pensamento ‘sistêmico’, por meio do qual buscamos compreender a natureza e a dinâmica do sistema semiótico como um todo.²

Nesses moldes, a LSF propõe que o sistema seja concebido a partir das dimensões necessárias para que o mesmo seja explicado, bem como os princípios pelos quais se organiza. Dentre as dimensões, destacam-se pela relevância para o presente trabalho: estratificação, metafunção e eixo.

3.1 Dimensões do sistema linguístico

Estratificação: na condição de sistema semiótico, a língua possui dois estratos – o conteúdo e a expressão. No entanto, a divisão do estrato do conteúdo em dois – um gramático-organizacional e outro semântico-discursivo – possibilitou ao sistema não apenas transmitir, mas também criar significado. Pelo fato de o conteúdo possuir um estrato responsável pela organização (i.e., a gramática, que realiza os itens semânticos e o texto), qualquer escolha gramatical cria, necessariamente, uma mudança na produção do significado

² There are many reasons for adopting this systemic perspective; one is that languages evolve – they are not designed, and evolved systems cannot be explained simply as the sum of parts. Our traditional compositional thinking about language needs to be, if not replaced by, at least complemented by a ‘systems’ thinking whereby we seek to understand the nature and the dynamic of a semiotic system as a whole.

e no desenvolvimento do texto. Contudo, ressalta-se que essa criação não pode acontecer independente do contexto (retórico).

Halliday (2002) afirma que o fato de o sistema ser estratificado permite que a gramática comporte ao mesmo tempo mais de um modelo de realidade, podendo, com isso, equilibrar os recursos de cada modelo dependendo da experiência a ser concebida, o que se faz em ocasiões diferentes de formas diferentes entre os estratos.

Com isso, é possível encontrar padrões na utilização dos recursos; isso porque os eventos do mundo que são concebidos pelos falantes como fenômenos similares tendem a ser categorizados de forma similar e, conseqüentemente, seu significado ser formalmente produzido na gramática de forma similar. Esses padrões se formam pela pressão exercida pelo contexto e a necessidade de se produzirem novos significados para novas situações.

Por conseguinte, além de criar os significados, a língua cria também as variáveis contextuais relevantes a partir das quais deve ser interpretada. As variáveis compreendem (i) o estabelecimento e manutenção das relações entre os falantes, (ii) a organização da realidade e do conhecimento e (iii) a constituição de cada significado individual como parte da unidade básica da língua – o texto (discurso em contexto). A variação contextual promove uma diversificação dos recursos gramaticais, de forma que os sistemas se agrupem em conjuntos de funções distintos, denominados metafunções.

Metafunção: é a dimensão da língua responsável por agrupar os sistemas segundo a sua codependência para a produção de determinado tipo de significado, que pode ser (i) estabelecer e manter as relações entre os falantes, (ii) organizar a realidade e o conhecimento, e (iii) variar a frequência de itens segundo a situação para assim criar o discurso em contexto. A metafunção interpessoal é responsável pelo tipo de relação ou interação que o falante (ou escritor) estabelece com seu ouvinte (ou leitor). Assim, relações de poder entre emissor e receptor, bem como as de polidez, subserviência, exigência, pedido, entre outras, são codificadas pelos sistemas desta metafunção. A metafunção ideacional representa nossa percepção das coisas que existem no mundo. É responsável pela transitividade, ou o tipo de interação existente entre as coisas (participantes) e os eventos (processos). Assim, os processos, isto é, o conjunto de realização dos eventos, os participantes e as circunstâncias nas quais estes processos acontecem compreendem esta metafunção. A organização das informações na construção do texto é de responsabilidade da metafunção textual. Segundo Halliday (1976), o texto é uma unidade operacional da língua, que pode ser falado ou

expresso através da escrita, independente de seu tamanho. É na composição do texto que o falante escolherá quais partes de seu texto são mais importantes e merecem destaque, quais ideias mantêm relação com outras, e assim por diante.

Eixo paradigmático: segundo a dimensão do eixo paradigmático, a língua é examinada como o conjunto de opções sistêmicas. Por um lado, isto implica em dar proeminência à escolha como forma principal de se produzirem significados; por outro lado, o valor de cada opção só pode ser conhecido na oposição às outras opções (i.e., na relação de agnação). O princípio que ordena o sistema linguístico do ponto de vista do eixo paradigmático é a delicadeza, segundo o qual a escolha de uma opção se dá em níveis cada vez mais refinados desde a condição de entrada mais geral (indelicada) até a escolha, no final do sistema, por um item lexical, que seria a opção mais delicada.

Os sistemas são organizados em redes e cada item de uma classe constitui um termo a ser selecionado. Dentre os itens de uma classe, os sistemas determinam a propriedade que permite a seleção de um item possível, e não de outro, igualmente possível. Como forma de notação do sistema, utilizam-se as redes dos sistemas. A título de ilustração, apresenta-se a seguir a rede do sistema de QUANTIFICAÇÃO do PB (Figura 1).



FIGURA 1 – O sistema de QUANTIFICAÇÃO em português brasileiro.

A unidade da condição de entrada para a QUANTIFICAÇÃO sistema é o grupo nominal, aberta para três subsistemas: ORDEM, PRECISÃO e MODO. Em cada subsistema, há duas opções e apenas uma delas deve ser selecionada. Por exemplo, se for selecionada uma opção da seguinte maneira: ORDEM → ordenação; PRECISÃO → particularização; MODO → indicativo, então a opção em PB é realizada pelos numerais ordinais, como em ‘terceiro’ ou ‘décimo’.

Um passo importante da investigação gramatical, é relacionar as funções às outras opções que a ela são opostas, dispondo os dados da pesquisa formalmente. Logo, faz parte da metodologia desta pesquisa dispor formalmente os dados, apresentando-os como redes de sistemas.

Especificamente no caso deste trabalho, o foco serão os principais sistemas da oração, interpessoal, ideacional e textual (respectivamente, os sistemas de MODO, TRANSITIVIDADE e TEMA).

3.2 Dinâmica textual

A produção de significado é caracterizada como a atividade de seleções sucessivas de funções dos sistemas, de forma que, cumulativamente, gerem um texto. O acúmulo de significado no texto pode ser observado segundo dois pontos de vista (LEMKE, 1991, p. 25):

(i) Sinóptico: o sistema é visto como o conjunto geral de todas as funções, disposto de forma organizada pelo *valeur* e pela delicadeza. O sistema são todas as funções à disposição para selecionar dentre elas os inúmeros significados, conjuntos de significados e redes de relações diferentes. O termo ‘sinóptico’ implica em uma descrição geral de todas as possibilidades do sistema.

(ii) Dinâmico: o sistema é visto como o conjunto de relações entre as funções selecionadas durante a produção de um texto. O termo ‘dinâmico’ implica em observar, dentre todas as possibilidades, apenas aquelas que de fato foram selecionadas para produzir determinado texto.

Lemke (1991) caracteriza a relação entre os sistemas e as perspectivas afirmando que uma rede sistêmica é sempre sinóptica. Com isso, ela não consegue apreender o processo de produção de significado. Esse autor introduz a noção de sistema dinâmico, o qual descreve, para cada escolha no sistema, conjuntamente a opção “próxima escolha”. Lemke (1991, p. 25-26) conclui:

[...] é fundamental compreender “como” e “por que” se dá a variação. Dado que a descrição dos sistemas dinâmicos é, basicamente, uma questão de pesar a probabilidade de cada escolha, então cabe entender como as escolhas “até este ponto” condicionam as probabilidades para “próxima escolha”.³

³ [...] it is very important for us to be able to say something about “how” and “why” it varies as it does. if we imagine the description of dynamic systems to be mainly a matter of the dynamic weightings of selection

A perspectiva da produção dinâmica altera o conceito de significado, que pode variar dependendo de quais outras funções se opõem em uma mesma rede, formando assim um fluxo de significados. Igualmente, a produção dinâmica cria uma história da rede. Na perspectiva dinâmica, a história da rede armazena o acúmulo de informação nos sistemas.

É justamente essa história da rede que a perfilação gramatical captura e, a partir daí, possibilita a compilação de *corpora* das histórias das redes. Esse processo tem início na dimensão do eixo e, subsequentemente, integra as outras dimensões segundo a relação entre as disposições tipológica e topológica da gramática.

3.3 Tipologia e Topologia

A tipologia pode ser definida como o conjunto das relações de oposição nos sistemas de acordo com as categorias que os diferenciam. Na descrição gramatical, a tipologia se associa à organização paradigmática, uma vez que captura o aspecto da escolha das funções, dividindo-as, assim, em tipos (i.e., classes).

A topologia, por sua vez, se define como a representação multiaxial que estabelece um espaço no qual as categorias são dispostas de forma gradual e em posição relativa (SOUZA, 2013). Na descrição gramatical, a topologia busca apreender as relações entre as funções que não são definidas pela oposição ou pela escolha, mas que, de outra forma se associam por outras formas, tais como a coocorrência nos textos, a frequência, ou a proximidade semântica (MARTIN e MATTHIESSEN, 1991).

O modelo de organização sistêmica, construído a partir das relações paradigmáticas, dá ênfase ao aspecto tipológico do sistema. Contudo, o modelo tipológico não consegue apreender algumas relações entre as funções, por elas não se relacionarem como oposição em um determinado paradigma, mas por serem produzidas a partir de relações dinâmicas que variam com os textos. Com isso, é possível perceber que a organização tipológica se associa à visão sinóptica do sistema.

Dessa maneira, uma forma de complementar o modelo é tentar apreender tais relações. Se a necessidade se relaciona à visão dinâmica do sistema, faz sentido empregar um modelo topológico. É isto que a perfilação gramatical apresentada neste artigo busca fazer.

probabilities, then we wish to know how the selections “up to now” condition the probabilities for selections “now”.

4. Perfilação gramatical: uma metodologia baseada em *corpus*

Dado que o significado produzido pela gramática está no contraste das opções paradigmáticas, o trabalho realizado pela dimensão do eixo é alterar a organização dos sistemas para produzir variação e, assim, gerar textos para as diferentes situações (HALLIDAY, 1996). Nessas condições, faz-se necessária uma abordagem multidimensional, que integre estratificação, metafunção e eixo, segundo a modelagem das redes sistêmicas (MARTIN, 2013; QUIROZ, 2013). Além disso, é igualmente necessário explicar o funcionamento da dinâmica textual (LEMKE, 1991) – ou a seleção sucessiva de opções nos sistemas para a produção de um texto.

Tendo como base a LC e o aporte teórico da LSF, apresenta-se, a seguir, a forma como a metodologia para a perfilação gramatical se desenvolve. A perfilação parte da organização sistêmica e das redes para a construção de mapas topológicos e do espaço gramatical do qual serão extraídos os perfis gramaticais do *corpus*.

4.1 Mapas topológicos

O mapa topológico da gramática consiste em converter as oposições sistêmicas em espaços entre as funções (MARTIN e MATTHIESSEN, 1991). Como dito anteriormente, o mapa topológico abrange as seguintes etapas da perfilação gramatical: (a) o mapeamento das funções gramaticais, (b) o estabelecimento da distância topológica entre as funções e, por fim, (c) o movimento do emprego das funções no espaço gramatical.

Como ilustração do desenho dos mapas, será apresentado o sistema de MODO IMPERATIVO em PB, o qual possui quatro opções.

1) A primeira opção mais delicada do imperativo se refere ao Sujeito das orações imperativas. Este é sempre o ouvinte no Modo imperativo jussivo, mas pode também incluir o falante nas orações sugestivas. Por exemplo: “**Tomemos** como exemplo o segmento de frase”.

2) Uma outra opção mais delicada é o recurso para aumentar a chance de se obter uma resposta esperada ao comando. Esta é realizada pela possibilidade de se expressar uma cópia do Predicador, igualmente com forma verbal imperativa, ao fim da proposta, funcionando como o Ecopredicador. Por exemplo: “**Conta** seu sonho pra ele, **conta**”. Ressalta-se que a utilização do Ecopredicador se restringe às opções sugestivo ou jussivo: neutro no sistema de tipo de imperativo.

3a) Imperativo: singularizado: Ocorre quando o falante determina exatamente a quem será atribuído o comando. Nesse caso, a especificidade do comando é realizada pela posição na estrutura dos elementos do Negociador: Sujeito seguido pelo Predicador.

3b) Imperativo: negociado: este ocorre quando a responsabilidade por se obedecer a um comando é negociada entre os interlocutores. Nesse caso, o Sujeito segue o Predicador na estrutura Predicador[^]Sujeito. Por exemplo: “**Você abre** a porta.” “Eu não, **abre você**”.

Além do TIPO DE IMPERATIVO, o PB apresenta ainda um outro subsistema, denominado FORÇA. Em PB a FORÇA é realizada fonologicamente pelos movimentos tônicos 1, 1+ e 5. Uma oração imperativa pode codificar um comando (demanda de bens-e-serviços), que seria a opção não-marcada, opção esta realizada pelo Movimento Tônico 1 (queda). Pode ainda codificar uma opção marcada: (i) codificando um pedido, realizada pelo Movimento Tônico 5 (subida; queda); ou (ii) codificando uma ordem, realizada pelo Movimento Tônico 1+ (queda forte).

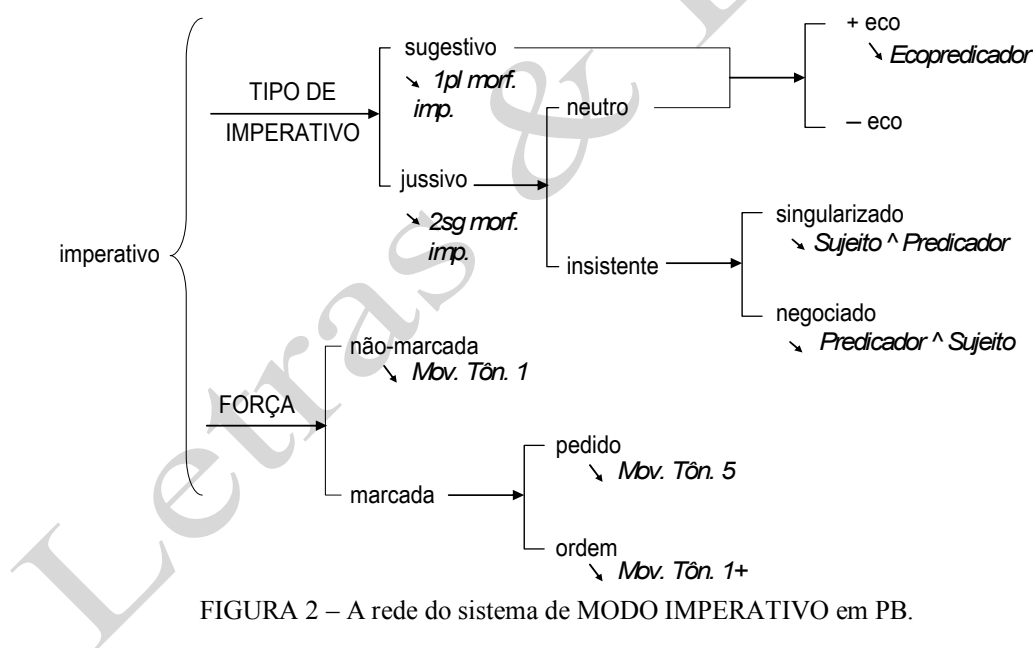


FIGURA 2 – A rede do sistema de MODO IMPERATIVO em PB.

A partir da rede, dividem-se as escolhas conforme seu nível de delicadeza, e cada uma delas é numerada. Cabe ressaltar que os números representam valores categóricos e, com isso, irão revelar a posição relativa das funções no mapa topológico (Figura 3).

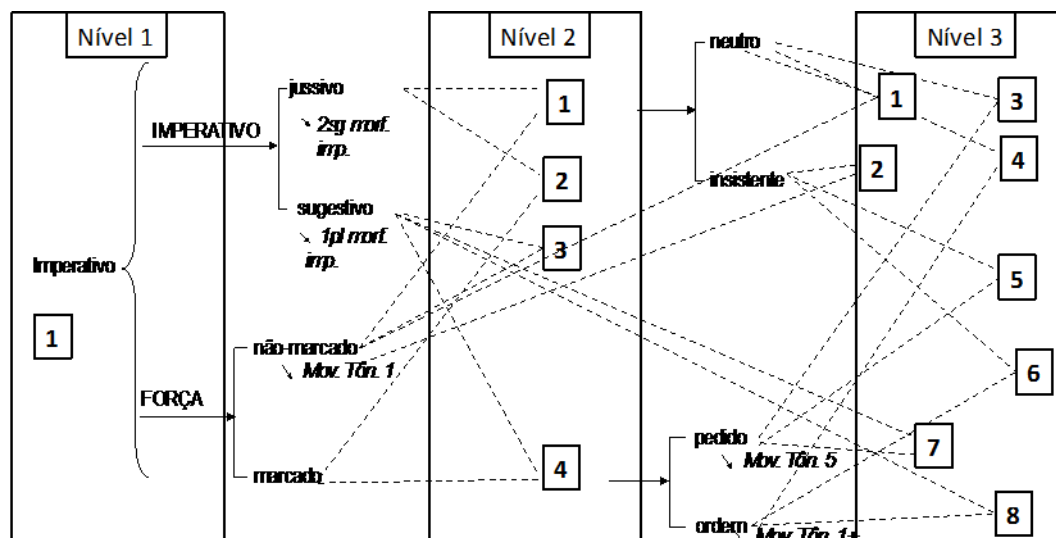


FIGURA 3 – Categorização por nível de delicadeza do IMPERATIVO.

A partir daí, é possível conferir a cada cosseleção uma etiqueta que representa o seu lugar no sistema, conforme um conjunto de escolhas. O Quadro 1 mostra as etiquetas de cada função, bem como a escolha feita para cada nível de delicadeza. Por exemplo, a função 1124 mostra que, no primeiro nível de delicadeza (N1), seleciona-se 1 para a função Imperativo. No segundo nível de delicadeza (N2), seleciona-se 1 para a função Jussivo & Não-Marcado. Já no terceiro nível de delicadeza (N3), seleciona-se 2 para a função Insistente. Por fim, para o nível mais delicado (N4), seleciona-se 4, a função Negociado. Por exemplo, “Assiste o filme você”.

QUADRO 1 – Etiquetamento das cosseleções do MODO IMPERATIVO.

N1	função	N2	Função	N3	função	N4	#	exemplo	
1	jussivo & não-marcado	1	neutro insistente	1	não-eco eco	1	1111	Assiste o filme.	
							1112	Assiste o filme, assiste.	
		2	ordem & neutro pedido & neutro ordem & insistente pedido & insistente	2	singularizado negociado	3	3	1123	Você assiste o filme.
								1124	Assiste o filme você.
								1231	Assiste o filme. (+1)
								1232	Assiste o filme, assiste. (+1)
	sugestivo & não-marcado	3	sugestivo & ordem sugestivo & pedido	7	8	4	1241	Assiste o filme. (+5)	
							1242	Assiste o filme, assiste. (+5)	
							1253	Vc assiste o filme (+1)	
							1254	Assiste o filme vc (+1)	
							1263	Vc assiste o filme (+5)	
							1264	Assiste o filme vc (+5)	
sugestivo & marcado	4					1311	Vamos ver o filme.		
						1312	Vamos ver o filme, vamos.		
						1471	Vamos ver o filme (+1)		
						1472	Vamos ver o filme, vamos (+1)		
						1481	Vamos ver o filme (+5)		
						1482	Vamos ver o filme, vamos (+5)		

O etiquetamento das funções é um recurso analítico importante por conseguir capturar as cosseleções entre os subsistemas, bem como o seu desenvolvimento na delicadeza. Ele permite desenhar um mapa topológico com as distâncias relativas entre as funções.

Para a perfilação, os mapas topológicos servem como uma “descrição dinâmica” da gramática, seguindo Lemke (1991), uma vez que permitem a visualização do panorama gramatical para um determinado fenômeno. A Figura 4 mostra o mapa topológico do MODO em PB. No caso da Figura 4, tem-se um panorama da gramática interpessoal da oração em PB. O mapa compreende o sistema de MODO e seus subsistemas, MODO INDICATIVO e MODO IMPERATIVO, incluindo todos os seus subníveis de delicadeza.

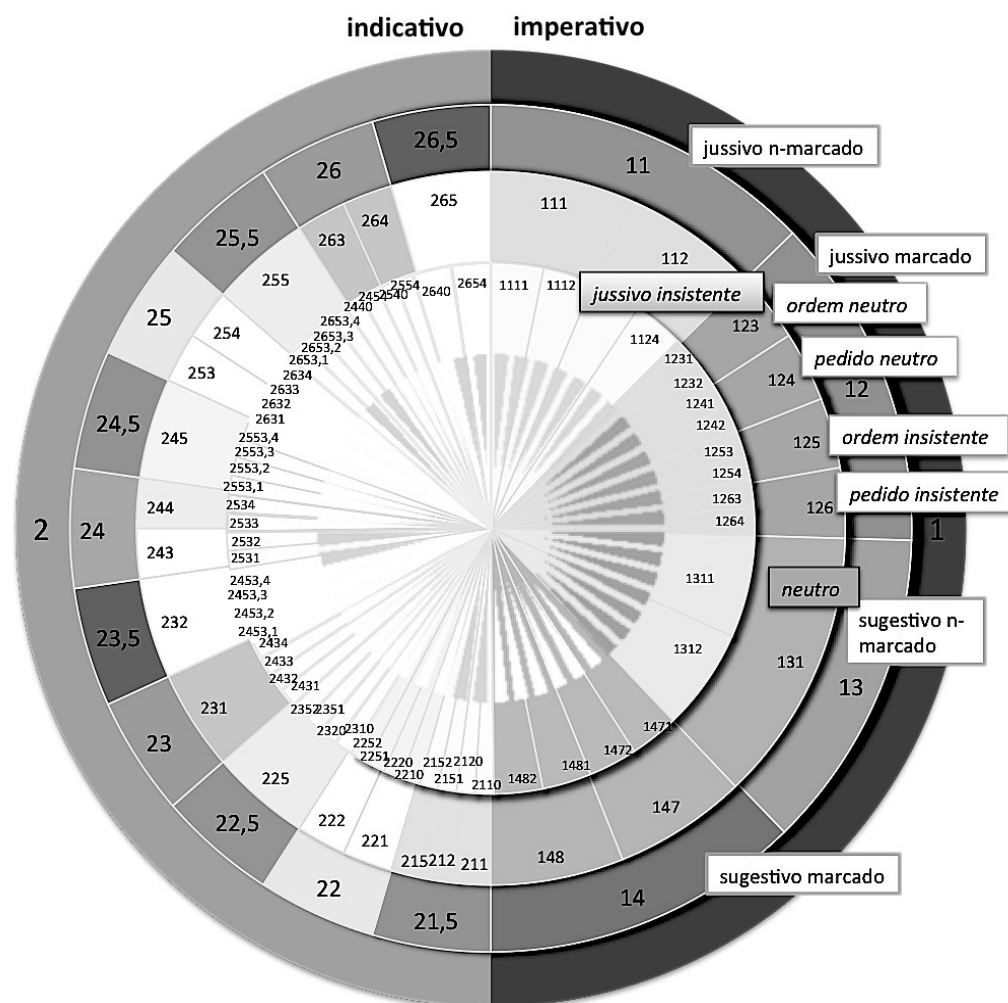


FIGURA 4 – Mapa topológico do MODO em PB.

4.2 Corpora gramaticais

Os mapas topológicos permitem que cada categoria – i.e., funções, ou cosseleções de funções, para qualquer nível de delicadeza – seja anotada por um código simples de apenas alguns dígitos (no caso do MODO, quatro ou cinco). Torna-se possível converter a anotação gramatical em um conjunto de *tokens*, sendo que cada etiqueta do código se torna um *token*. Por conseguinte, os mapas topológicos possibilitam a compilação de *corpora* gramaticais que, por sua vez, podem ser investigados pelas etiquetas, as quais representam “*tokens* gramaticais”.

A Figura 5 traz um exemplo de um *corpus* de *tokens* gramaticais elaborado a partir do *corpus* da pesquisa. O número entre parênteses indica a sequência das orações e cada ponto final indica o início de uma nova oração. Cada código de seis dígitos indica as funções gramaticais da oração. Na primeira célula, as funções se apresentam na forma como foram empregadas no texto; na segunda célula, apresentam-se em um exemplo de busca por linhas de concordância.

Token	Frequency	Percentage
14 21 10. (3) 16 21 41. (4) 16 21 41. (5) 16 21 41. (6)	16	15 13% 03%
31 21 10. (3) 33 21 41. (4) 16 21 41. (5) 16 21 41. (6)	16	14 17% 05%
19 21 10. (2) 14 21 10. (3) 16 21 41. (4) 16 21 41. (5)	16	11 17% 05%
9 (1) 16 21 10. (2) 52 21 10. (3)	16	2 03% 02%
21 10. (27) 14 21 41. (28) 19 21 41.	19	109 17% 07%
21 41. (25) 14 21 10. (26) 19 21 10. (27) 14 21 41.	19	101 17% 07%
4 (1) 19 21 10. (2) 33 21 10. (3)	19	3 03% 03%
3 (1) 19 21 10. (2) 14 21 10. (3)	19	3 03% 04%
8 (1) 19 21 10. (2) 31 21 10. (3)	19	2 03% 04%
6 (1) 19 21 10. (2) 14 21 45. (3)	19	2 03% 04%
5 (1) 19 21 10. (2) 11 21 41. (3)	19	2 03% 02%
1 (1) 19 21 10. (2) 14 21 45. (3)	19	2 03% 03%
(11) 14 21 10. (12) 39 21 32.	32	48 13% 03%
21 41. (19) 59 21 10. (20) 31 21 41. (21) 33 21 10.	31	77 13% 03%
21 45. (10) 14 21 41. (11) 31 21 10. (12) 14 21 41.	31	42 12% 04%
14 21 10. (8) 14 21 10. (9) 31 21 10. (10) 11 21 41.	31	35 13% 03%
11 21 10. (8) 14 21 41. (9) 31 21 45. (10) 14 21 41.	31	34 17% 03%
8 (1) 19 21 10. (2) 31 21 10. (3) 33 21 41. (4)	31	6 13% 01%
21 10. (17) 42 21 41. (18) 32 21 10. (19) 14 21 10.	32	71 14% 05%
21 41. (11) 14 21 10. (12) 32 21 10. (13) 14 21 10.	32	47 17% 02%
11 21 10. (8) 39 21 10. (9) 32 21 45. (10) 39 21 10.	32	34 14% 07%
21 10. (23) 14 21 10. (24) 33 21 10. (25) 11 21 41.	33	95 17% 07%
21 10. (23) 14 21 10. (24) 33 21 41.	33	94 17% 07%
21 10. (22) 14 21 41. (23) 33 21 10. (24) 39 21 41.	33	89 13% 09%
21 10. (20) 31 21 41. (21) 33 21 10. (22) 14 21 41.	33	81 12% 02%

FIGURA 5 – Exemplo de um *corpus* de *tokens* gramaticais.

Quando o movimento do emprego das funções gramaticais é acrescentado à perfilação, faz-se necessário acompanhar a forma pela qual a dinâmica textual se desenvolve. Assim, não

basta apenas entender as distâncias topológicas entre as funções em um texto, mas, igualmente, cabe levar em conta a variável ‘tempo’ – ou a sequência em que ocorrem as mudanças, dinamicamente. Isso é feito estabelecendo a distância entre as funções gramaticais das orações, na ordem em que aparecem no texto.

Tomando como exemplo o TEXTO 10 do *corpus* da presente pesquisa, as quatro primeiras orações são classificadas conforme suas funções gramaticais para os sistemas de TRANSITIVIDADE, MODO e TEMA e recebem os seguintes códigos de acordo com sua localização nos mapas topológicos:

QUADRO 2 – Análise na perfilação gramatical

oração		FUNÇÃO	CÓDIGO
1	Évariste Galois nasceu nas proximidades de Paris, na aldeia de Bourg la-Reine, onde seu pai era prefeito.	TRANS.: Material: Intransitivo (11) MODO: Declarativo (21) TEMA: Default (10)	112110
2	Aos 12 anos mostrava pouco interesse por Latim, Grego e Álgebra mas a Geometria de Legendre o fascinava.	TRANS.: Material: Transitivo (14) MODO: Declarativo (21) TEMA: Perspectiva (41)	142141
3	Aos 16 anos, julgando-se em condições, procurou entrar na Escola Politécnica mas foi recusado por falta de preparo e isto marcou seu primeiro fracasso.	TRANS.: Mental: Cognitivo (57) MODO: Declarativo (21) TEMA: Perspectiva (41)	572141
4	Aos 17 anos escreveu um artigo onde expunha suas descobertas fundamentais entregando-o a Cauchy para que o apresentasse na Academia.	TRANS.: Material: Transitivo (14) MODO: Declarativo (21) TEMA: Perspectiva (41)	142141

Relativamente a essas quatro orações, pode-se perceber que 1, 2 e 4 são próximas entre si e distantes de 3 na TRANSITIVIDADE; as quatro estão no mesmo ponto com relação ao MODO; e 1 está distante de 2, 3 e 4 com relação ao TEMA. Quando esse tipo de medida é feito para o texto completo, é possível desenhar um gráfico que represente essas distâncias no espaço gramatical (Figura 6).

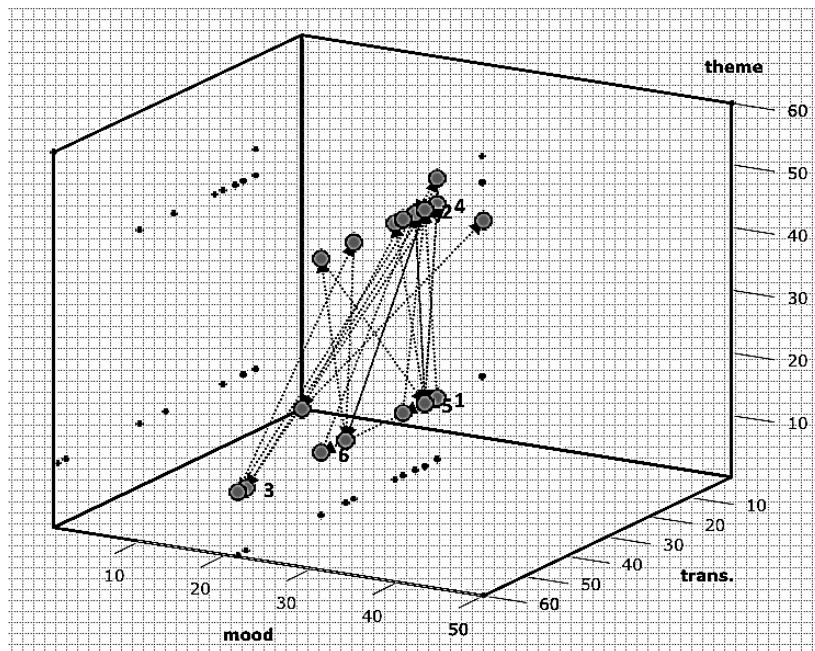


FIGURA 6 – Dinâmica no espaço topológico da gramática para o TEXTO 10

Como mostra a Figura 6, cada ponto representa uma oração e, com isso, um conjunto de coordenadas das funções oracionais relativas à TRANSITIVIDADE (eixo x), ao MODO (eixo y) e ao TEMA (eixo z). Esse tipo de gráfico que representa o espaço topológico da gramática consegue contribuir para a análise do texto por permitir que ela seja visualizada, por completo, de uma só vez. Além disso, consegue também estabelecer as distâncias, do ponto de vista gramatical, entre as orações e suas posições relativas. Por fim, as linhas que ligam as orações convertem-se, na verdade, em vetores. Consequentemente, depreendem movimento em função da variável tempo e, assim, capturam a dinâmica do desenvolvimento do texto por completo.

A coleção de espaços gramaticais de diversos textos também funciona como a compilação de *corpora*. Neste caso específico, não são *corpora* de tokens lexicais, ou gramaticais, mas sim de “tokens textuais”.

Somado à compilação de *corpora* de “tokens textuais” está o movimento entre as orações e, por conseguinte, entre suas funções, extraindo dados relativos apenas à dinâmica do desenvolvimento do texto. A partir do gráfico, ainda tomando o TEXTO 10 como exemplo, cada vetor permite calcular a mudança gramatical entre as orações ao longo do tempo. Com isso, elaboram-se uma lista da sequência de equações vetoriais – em outras palavras, compila-se um *corpus* de “tokens discursivos” (Quadro 3).

QUADRO 3 – *corpus* da dinâmica textual: equações vetoriais do desenvolvimento do texto.

Oração	trans.(x)	modo (y)	tema (z)	amostra do <i>corpus</i> de equações vetoriais	linhas de concordância do <i>corpus</i> de equações vetoriais
01	11	21	10		$r = \langle 11, 21, 10 \rangle + t \langle 3, 0, 31 \rangle$
				$r = \langle 11, 21, 10 \rangle + t \langle 3, 0, 31 \rangle$	$r = \langle 57, 21, 10 \rangle + t \langle -46, 0, 31 \rangle$
02	14	21	41		$r = \langle 39, 21, 10 \rangle + t \langle 23, 0, 31 \rangle$
				$r = \langle 14, 21, 41 \rangle + t \langle 43, 0, -31 \rangle$	$r = \langle 14, 21, 10 \rangle + t \langle 7, 0, 31 \rangle$
03	57	21	10		
				$r = \langle 57, 21, 10 \rangle + t \langle -46, 0, 31 \rangle$	
04	11	21	41		
				$r = \langle 11, 21, 41 \rangle + t \langle 3, 0, -31 \rangle$	
05	14	21	10		
				$r = \langle 14, 21, 10 \rangle + t \langle 15, 0, 0 \rangle$	
06	39	21	10		
				$r = \langle 39, 21, 10 \rangle + t \langle 23, 0, 31 \rangle$	
07	16	21	41		
				$r = \langle 16, 21, 41 \rangle + t \langle -2, 0, -31 \rangle$	
08	14	21	10		
				$r = \langle 14, 21, 10 \rangle + t \langle 7, 0, 31 \rangle$	
09	21	21	41		
				$r = \langle 21, 21, 41 \rangle + t \langle -7, 0, 0 \rangle$	
10	14	21	41		

O conjunto das análises e dos dados gerados a partir delas nessa seção compõe a perfilação gramatical. A próxima seção apresenta um Exemplo de Pesquisa da perfilação em um *corpus* de minibiografias.

4.3 Perfilação gramatical das minibiografias – Exemplo de Pesquisa

Esta seção tem por objetivo apresentar um exemplo de Pesquisa da perfilação. Especificamente, esta traz o perfil gramatical de um *corpus* de minibiografias em PB e, devido ao escopo reduzido, irá se concentrar apenas na função de maior número absoluto de ocorrências.

Para o desenho do perfil gramatical, fazem-se necessárias as seguintes etapas: (a) compilação do *corpus*; (b) análise sistêmica do estrato gramatical; (c) distribuição topológica das funções gramaticais; (d) constituição do *corpus* gramatical; (e) busca de padrões no espaço gramatical e no desenvolvimento do texto (LEMKE, 1991).

(a) **Compilação do corpus:** o *corpus* da pesquisa foi compilado a partir da seleção de dez minibiografias de matemáticos importantes, apresentadas no início dos capítulos de livros didáticos de matemática. Conforme a tipologia da língua no contexto de cultura (HALLIDAY e MATTHIESSEN, 2004), os textos do *corpus* são classificados como: Relatar/Escrito/Monólogo. Especificamente, foram retirados dos livros da coleção *Fundamentos de Matemática Elementar*, de Gelson Iezzi (IEZZI *et al.*, 1977/78). A descrição do *corpus* pode ser vista no Quadro 4.

QUADRO 4 – O *corpus* da pesquisa.

TEXTO	NOME	TIPOLOGIA	VOL.	TOKENS
GAUSS 01	<i>De Plebeu a Príncipe</i>	Relatar/Escrito/Monólogo	1	391
ABEL 02	<i>Jovem Luta para Ser Ouvido</i>	Relatar/Escrito/Monólogo	1	335
FERMAT 03	<i>As Margens dos Livros Falam</i>	Relatar/Escrito/Monólogo		320
EULER 04	<i>Cego Enxerga Longe</i>	Relatar/Escrito/Monólogo	2	454
LEIBNIZ 05	<i>Autodidata Cria a Análise</i>	Relatar/Escrito/Monólogo	2	403
FOURIER 06	<i>Condução do Calor: Nova Teoria</i>	Relatar/Escrito/Monólogo	3	227
LAPLACE 07	<i>Napoleão Demite Ministro do Interior</i>	Relatar/Escrito/Monólogo	4	335
V. NEUMANN 08	<i>Computadores – As Máquinas com Memória</i>	Relatar/Escrito/Monólogo	5	279
NEWTON 09	<i>Nem só de Física Vive um Gênio</i>	Relatar/Escrito/Monólogo	5	477
GALOIS 10	<i>Intelectual Morre em Duelo</i>	Relatar/Escrito/Monólogo	6	366
TOTAL				3.587

(b) **Análise sistêmica do estrato gramatical:** A análise gramatical do presente estudo focalizou a oração como unidade de investigação, especificando seus três principais sistemas, a partir da diversificação metafuncional. Assim, analisaram-se as funções ideacionais do sistema de TRANSITIVIDADE; interpessoais do sistema de MODO; e textuais do sistema de TEMA. Para esta análise, foram contemplados os dois primeiros níveis de delicadeza. Por exemplo:

QUADRO 5 – Exemplo de análise sistêmica.

	Évariste Galois	nasceu	nas proximidades de Paris, na aldeia de Bourg la-Reine
TRANS.	<i>Ator</i>	<i>Material: Intransitivo e Criativo</i>	<i>Circ. Localização</i>
MODO	<i>Sujeito</i>	<i>Finito</i>	<i>Adj. Circunstancial</i>
	<i>Modo: Indicativo: Declarativo</i>		<i>Resto</i>
TEMA	<i>Tema: Default</i>	<i>Rema</i>	

(c) **Distribuição topológica das funções gramaticais:** De posse de todas as funções dos sistemas principais da oração, estas foram localizadas nos mapas topológicos. Com isso, foi

possível obter do *corpus* sua distribuição topológica. Como exemplo, apresentam-se os valores nos dois primeiros níveis de delicadeza para as funções do texto GAUSS 01.

QUADRO 6 – Distribuição topológica do texto GAUSS 01

oração	Trans. (X)		modo (y)		tema (z)	
GAUSS 01	Material: Intransitivo e Criativo	19	Indicativo: Declarativo Temporal	21	Default	10
GAUSS 02	Material: Transitivo e Transformativo	14	Indicativo: Declarativo Temporal	21	Proeminente: Intensivo	45
GAUSS 03	Material: Transitivo e Transformativo	14	Indicativo: Declarativo Temporal	21	Proeminente: Perspectiva	41
GAUSS 04	Mental: Fenômeno e Cognitivo	59	Indicativo: Declarativo Temporal	21	Default	10
GAUSS 05	Material: Transitivo e Transformativo	14	Indicativo: Declarativo Temporal	21	Proeminente: Perspectiva	41
GAUSS 06	Material: Intransitivo e Transformativo	11	Indicativo: Declarativo Temporal	21	Default	10
GAUSS 07	Material: Intransitivo e Transformativo	11	Indicativo: Declarativo Temporal	21	Default	10
GAUSS 08	Relacional: Identificativo e Intensivo	39	Indicativo: Declarativo Temporal	21	Default	10
GAUSS 09	Relacional: Atributivo e Possessivo	32	Indicativo: Declarativo Temporal	21	Proeminente: Intensivo	45
GAUSS 10	Relacional: Identificativo e Intensivo	39	Indicativo: Declarativo Temporal	21	Default	10
GAUSS 11	Material: Transitivo e Transformativo	14	Indicativo: Declarativo Temporal	21	Proeminente: Intensivo	45
GAUSS 12	Verbal: Semiose e Não-Recepção	41	Indicativo: Declarativo Temporal	21	Default	10
GAUSS 13	Material: Transitivo e Transformativo	14	Indicativo: Declarativo Temporal	21	Proeminente: Perspectiva	41
GAUSS 14	Material: Transitivo e Transformativo	14	Indicativo: Declarativo Temporal	21	Default	10
GAUSS 15	Material: Intransitivo e Transformativo	11	Indicativo: Declarativo Temporal	21	Default	10
GAUSS 16	Material: Transitivo e Criativo	16	Indicativo: Declarativo Temporal	21	Proeminente: Perspectiva	41
GAUSS 17	Material: Transitivo e Criativo	16	Indicativo: Declarativo Temporal	21	Proeminente: Perspectiva	41
GAUSS 18	Relacional: Identificativo e Intensivo	39	Indicativo: Declarativo Temporal	21	Default	10

(d) *Constituição do corpus gramatical*: A partir dos mapas topológicos, são realizadas as distribuições tais como aquela apresentada no Quadro 6. A partir desse trabalho feito para todos os textos do *corpus*, apresenta-se a seguir o “*corpus gramatical*” do presente estudo (Quadro 7).

QUADRO 7 – *corpus* gramatical da pesquisa

TEXTO	Posições topológicas
GAUSS 01	(1) 19 21 10. (2) 14 21 45. (3) 14 21 41. (4) 59 21 10. (5) 14 21 41. (6) 11 21 10. (7) 11 21 10. (8) 39 21 10. (9) 32 21 45. (10) 39 21 10. (11) 14 21 45. (12) 41 21 10. (13) 14 21 41. (14) 14 21 10. (15) 11 21 10. (16) 16 21 41. (17) 16 21 41. (18) 39 21 10.
ABEL 02	(1) 33 21 10. (2) 41 21 41. (3) 14 21 10. (4) 14 21 41. (5) 14 21 41. (6) 33 21 10. (7) 11 21 10. (8) 14 21 41. (9) 33 21 10. (10) 14 21 10. (11) 33 21 10. (12) 59 21 41. (13) 33 21 10. (14) 11 21 10. (15) 14 21 41. (16) 11 21 10. (17) 42 21 41. (18) 32 21 10. (19) 14 21 10. (20) 58 21 10. (21) 58 21 10. (22) 14 21 10. (23) 14 21 10. (24) 33 21 10. (25) 11 21 41. (26) 21 41. (27) 14 21 41.
FERMAT 03	(1) 19 21 10. (2) 14 21 10. (3) 16 21 41. (4) 16 21 41. (5) 16 21 41. (6) 33 21 41. (7) 14 21 41. (8) 16 21 45. (9) 33 21 10. (10) 39 21 10. (11) 16 21 45. (12) 39 21 41. (13) 33 21 10. (14) 16 21 41. (15) 14 21 41. (16) 33 21 10. (17) 59 21 10. (18) 16 21 41.
EULER 04	(1) 19 21 10. (2) 33 21 10. (3) 11 21 10. (4) 11 21 45. (5) 14 21 41. (6) 33 21 10. (7) 14 21 10. (8) 14 21 10. (9) 31 21 10. (10) 11 21 41. (11) 14 21 10. (12) 32 21 10. (13) 14 21 10. (14) 14 21 10. (15) 39 21 10. (16) 16 21 41. (17) 16 21 10. (18) 11 21 45. (19) 14 21 10. (20) 11 21 10. (21) 14 21 10. (22) 49 21 45.
LEIBNIZ 05	(1) 19 21 10. (2) 11 21 41. (3) 33 21 41. (4) 11 21 41. (5) 14 21 10. (6) 11 21 10. (7) 39 21 41. (8) 16 21 10. (9) 16 21 10. (10) 39 21 10. (11) 39 21 41. (12) 16 21 10. (13) 14 21 45. (14) 33 21 10. (15) 39 21 10. (16) 14 21 10. (17) 16 21 10. (18) 14 21 41. (19) 16 21 10. (20) 39 21 10. (21) 14 21 10. (22) 14 21 41. (23) 14 21 45. (24) 59 21 45.
FOURIER 06	(1) 19 21 10. (2) 14 21 45. (3) 14 21 41. (4) 14 21 45. (5) 42 21 45. (6) 21 21 41. (7) 11 21 10. (8) 14 21 41. (9) 31 21 45. (10) 14 21 41. (11) 31 21 10. (12) 14 21 41. (13) 14 21 41. (14) 21 21 41.
LAPLAC E 07	(1) 11 21 10. (2) 11 21 45. (3) 14 21 41. (4) 14 21 41. (5) 33 21 10. (6) 39 21 10. (7) 41 21 41. (8) 16 21 41. (9) 16 21 41. (10) 39 21 10. (11) 14 21 10. (12) 39 21 32.
V. NEUM. 08	(1) 19 21 10. (2) 31 21 10. (3) 33 21 41. (4) 16 21 41. (5) 16 21 41. (6) 16 21 10. (7) 16 21 41. (8) 21 21 41. (9) 33 21 10. (10) 39 21 10. (11) 11 21 45. (12) 11 21 10. (13) 16 21 10. (14) 14 21 41.
NEWTO N 09	(1) 16 21 10. (2) 52 21 10. (3) 14 21 41. (4) 11 21 41. (5) 14 21 10. (6) 14 21 41. (7) 42 21 10. (8) 14 21 10. (9) 14 21 41. (10) 16 21 41. (11) 14 21 41. (12) 16 21 41. (13) 59 21 41. (14) 14 21 41. (15) 11 21 10. (16) 14 21 41. (17) 41 21 41. (18) 39 21 10. (19) 14 21 45. (20) 41 21 10. (21) 16 21 41. (22) 41 21 10. (23) 14 21 10. (24) 33 21 41.
GALOIS 10	(1) 11 21 10. (2) 14 21 41. (3) 57 21 10. (4) 11 21 41. (5) 14 21 10. (6) 39 21 10. (7) 16 21 41. (8) 14 21 10. (9) 21 41. (10) 14 21 41. (11) 11 21 45. (12) 16 21 41. (13) 14 21 10. (14) 11 21 41. (15) 14 21 10. (16) 11 21 10. (17) 14 21 10. (18) 16 21 41. (19) 59 21 10. (20) 31 21 41. (21) 33 21 10. (22) 14 21 41. (23) 33 21 10. (24) 39 21 41. (25) 14 21 10. (26) 19 21 10. (27) 14 21 41. (28) 19 21 41.

1(e) Busca de padrões – a construção de uma microteoria do corpus: Tendo como origem os pressupostos de investigação apresentados anteriormente (TOGNINI-BONELLI, 2001; HALLIDAY, 1992), as pesquisas gramaticais realizas seguindo uma metodologia de *corpus* partem da abordagem baseada para a direcionada pelo *corpus*, construindo a partir dos dados uma microteoria da língua. Com isso, a pesquisa tem início no estrato da expressão e se move para o estrato da organização, e das ocorrências de itens específicos para os padrões mais gerais de organização.

A metodologia da perfilação gramatical permite a condução desse tipo de pesquisa porque extrai da expressão um *corpus* gramatical e outro do desenvolvimento do texto. Tendo

em mãos o *corpus* gramatical, torna-se possível realizar análises do funcionamento gramatical em conjunto com sua frequência e padrões de ocorrência. Dentro do escopo da LC, busca-se por “linhas de concordância”, “listas de palavras”, “colocações”, dispersão e agrupamento, de forma análoga às pesquisas lexicais e da expressão.

Neste Exemplo de Pesquisa, começando a análise pela lista de concordâncias, esta mostra que a maior variação de funções no *corpus* se dá nos sistemas da TRANSITIVIDADE (11, 14, 57, 39, 16, 21, etc.), enquanto o sistema de MODO emprega sempre a mesma função (21), e o sistema de TEMA varia menos (10, 41 ou 45) (Figura 7).

TRANS. MODO TEMA		
Concordance		Seja
14 21 10. (3) 16 21 41.	16 21 41. (6)	16
31 21 10. (3) 33 21 41. (4) 10 21 41. (5) 16 21 41. (8)		16
19 21 10. (2) 14 21 10. (3) 16 21 41. (4) 18 21 41. (5)		16
9 (1) 16 21 10. (2) 52 21 10. (3)		16
21 10. (27) 14 21 41. (28) 19 21 41.		19
21 41. (25) 14 21 10. (26) 19 21 10. (27) 14 21 41.		19
4 (1) 19 21 10. (2) 33 21 10. (3)		19
3 (1) 19 21 10. (2) 14 21 10. (3)		19
8 (1) 19 21 10. (2) 31 21 10. (3)		19
6 (1) 19 21 10. (2) 14 21 45. (3)		19
5 (1) 19 21 10. (2) 11 21 41. (3)		19
1 (1) 19 21 10. (2) 14 21 45. (3)		19
(11) 14 21 10. (12) 39 21 32.		32
21 41. (19) 59 21 10. (20) 31 21 41. (21) 33 21 10.		31
21 45. (10) 14 21 41. (11) 31 21 10. (12) 14 21 41.		31
14 21 10. (8) 14 21 10. (9) 31 21 10. (10) 11 21 41.		31
11 21 10. (8) 14 21 41. (9) 31 21 45. (10) 14 21 41.		31
8 (1) 19 21 10. (2) 31 21 10. (3) 33 21 41. (4)		31

FIGURA 7 – Lista de concordâncias gramaticais.

A lista de concordâncias também mostra que o maior número de ocorrências é dos subtipos de Processo Material. Esse fato é corroborado por outros tipos de padrão que podem ser buscados.

A dispersão, efetuada pelo *plot* – com o auxílio da ferramenta Concord do *software* WordSmith Tools (SCOTT, 2007) – mostra que, no sistema de TRANSITIVIDADE, os Processos Materiais (Material: Transitivo e Transformativo [14]; Material: Transitivo e Criativo [16] e Material: Intransitivo e Transformativo [11]) apresentam maior dispersão, contribuindo assim para todas as fases do texto. De outra forma, os outros tipos de Processo, em particular os Relacionais (Relacional: Atributivo e Possessivo [32]; Relacional: atributivo e Intensivo [33] e Relacional: Identificativo e Intensivo [39]) ocorrem pontualmente, sugerindo a realização de um trabalho distinto no texto. Por exemplo:

- (1) NEWTON 09 (4) 11 21 41. Por ocasião da peste, Newton **voltou** <MATERIAL> para casa
- (2) VON NEUMANN 08 (2) 31 21 10. Foi <RELACIONAL> professor em Berlim e Hamburgo
- (3) NEWTON 09 VERBAL (20) 41 21 10. Wallis lhe **comunica** <VERBAL> que, na Holanda, o Cálculo é considerada descoberta de Leibniz
- (4) LEIBNIZ 05 (24) 59 21 45. Otimista ao extremo, sempre **acreditou** <MENTAL> numa universalização da linguagem

A Figura 8 mostra a dispersão das funções de TRANSITIVIDADE (Material: 11, 14, 16 e 19; Relacional: 33 e 39; Verbal: 42; e Mental: 52 e 59) no texto NEWTON 09.

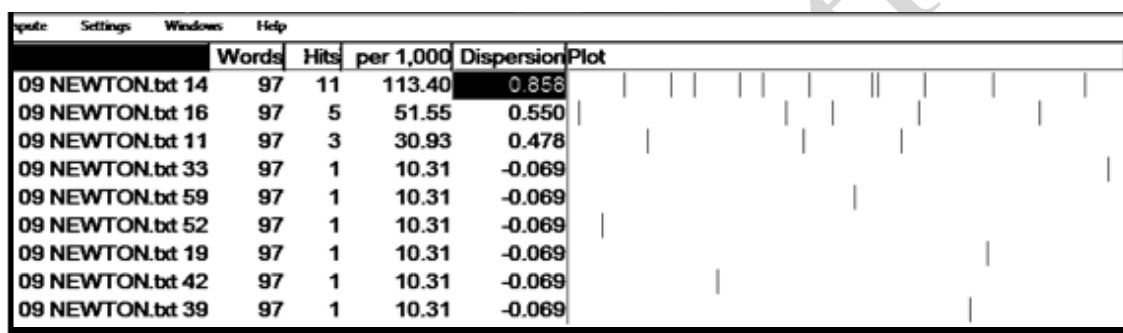


FIGURA 8 – Dispersão (*plot*) dos Tipos de Processo no texto NEWTON 09.

Dentre os subtipos de Processo Material, o mais frequente é o Transitivo e Transformativo [14]. A essa função cabe realizar uma mudança na Meta do Processo. Por exemplo:

- (5) FOURIER 06 (2) 14 21 45. Órfão aos 8 anos, Fourier **foi colocado** <MATERIAL> no Colégio Militar
- (6) GALOIS 10 (8) 14 21 10. Cauchy **perdeu** <MATERIAL> seu trabalho

Quando se observa o trabalho dessa função ao longo do *corpus*, encontra-se, de maneira semelhante, o trabalho importante desempenhado por ela nos textos. A seguir, a Figura 9 mostra como a função está dispersa.

File	Words	Hits	per 1,000	Dispersion	Plot
Co ódigo 01 GAUSS.bt 14	73	7	95.89	0.514	
Co ódigo 02 ABEL.bt 14	109	11	100.92	0.858	
Co ódigo 03 FERMAT.bt 14	74	4	54.05	0.596	
Co ódigo 04 EULER.bt 14	90	9	100.00	0.867	
Co ódigo 05 LEIBNIZ.bt 14	97	8	82.47	0.650	
Co ódigo 06 FOURIER.bt 14	57	8	140.35	0.714	
Co ódigo 07 LAPLACE.bt 14	49	3	61.22	0.478	
Co ódigo 08 VON NEUMANN.bt 14	57	2	35.09	-0.069	
Co ódigo 09 NEWTON.bt 14	97	11	113.40	0.858	
Co ódigo 10 GALOIS.bt 14	112	11	98.21	0.795	

FIGURA 9 – Dispersão (*plot*) Material: Transitivo e Transformativo no *corpus*.

A busca pelos agrupamentos do *corpus* revela ainda um outro aspecto importante dessa função, a saber, a maneira como ela forma padrões oracionais. Em primeiro lugar, os dois padrões oracionais mais frequentes empregam o Processo Material: Transitivo e Transformativo. O padrão 142141 é o mais frequente, com 106 ocorrências. Em seguida, 142110 aparece com 74 ocorrências (Figura 10).

N	Cluster	Freq.
1	14 21 41	106
2	14 21 10	74
3	16 21 41	69
4	33 21 10	47
5	11 21 10	39
6	39 21 10	34
7	11 21 41	25
8	16 21 10	25
9	14 21 45	19
10	19 21 10	14
11	11 21 45	14
12	39 21 41	13
13	33 21 41	11
14	41 21 41	10
15	59 21 10	9
16	31 21 10	8
17	21 21 41	8
18	59 21 41	8
19	41 21 10	7
20	16 21 45	6
21	32 21 10	6

FIGURA 10 – Agrupamentos (*clusters*) mais frequentes no *corpus*.

Na primeira linha (maior ocorrência) o código ‘142141’ significa orações Materiais Transitivas e Transformativas [14] Indicativas Declarativas Temporais [21] Tema Proeminente: Perspectiva [41].

O Modo Indicativo Declarativo, que ocorre em todas as orações do *corpus*, é empregado na biografia para estabelecer uma relação de especialista/leigo entre o produtor e o

leitor do texto, uma vez que não abre espaço para negociação. Soma-se a isso o fato de todas serem Temporais, não havendo modalização/possibilidade/dúvida. Além disso, O tempo empregado é sempre o passado (ou às vezes o “presente histórico”, ou o presente simples em orações materiais tipicamente realizadas pelo presente-no-presente), o que contribui para a construção da narrativa, fundamental para o tipo de texto da biografia.

O Tema Proeminente: Perspectiva é realizado por Circunstâncias de Intensificação: Localização Espaço-Temporal, Modo, Causa e Contingência. Na biografia, as Circunstâncias fazem com que o texto tenha sempre como ponto de partida os lugares onde os matemáticos nasceram, estudaram, trabalharam, etc., bem como as datas. Do ponto de vista do discurso, criam uma linha do tempo para os eventos da vida dos matemáticos, que passa pelos lugares onde viveram.

O agrupamento com os Processos do tipo 14 mostra como as atividades dos matemáticos formam um padrão com a linha do tempo da biografia quando em conjunto com Modo 21 e Tema 41. Por exemplo:

- (7) GAUSS 01 (5) 14 21 41. Ainda nesta obra <TEMA> Gauss apresenta <MATERIAL> a lei da reciprocidade quadrática <META>
(8) GAUSS 01 (13) 14 21 41. No começo do séc. XIX <TEMA> abandonou <MATERIAL> a Aritmética <META>
(9) ABEL 02 (4) 14 21 41. Nesta época <TEMA>, Abel conseguiu generalizar <MATERIAL> o teorema binomial <META>

Retomando a ideia da construção de uma microteoria sobre o *corpus*, é possível, com este pequeno Exemplo de Pesquisa apresentar uma descrição de como as funções gramaticais operam. Devido ao seu número maior de ocorrências, o foco se deu sobre a função Processo Material: Transitivo e Transformativo [14]. Foi possível mostrar a sua frequência, a distribuição nos textos, a dispersão ao longo dos textos, bem como os agrupamentos que forma, nos padrões oracionais, com outras funções.

Apesar de limitada a esta função, determinada pelo escopo deste Exemplo de Pesquisa, a análise consegue revelar o potencial da perfilação gramatical. Ela aponta, assim, para a possibilidade desse tipo de análise ser feita para quaisquer funções gramaticais de interesse.

5. Conclusões

Demonstrando como as metodologias de *corpus* podem contribuir para o desenvolvimento das pesquisas do estrato da gramática, este artigo ofereceu uma metodologia de perfilação gramatical, a qual possibilita compilar e analisar *corpora* gramaticais para além do que podem oferecer as buscas e contagens computacionais, restritas às realizações por palavras, compreendendo igualmente funções encobertas, realizadas de forma criptotípica.

A justificativa para uma metodologia própria para os estudos de *corpora* gramaticais se deve ao fato de que a maior parte das funções gramaticais, diferentemente dos itens lexicais, não é realizada de forma aparente pela expressão, mas de forma encoberta a partir de um conjunto de reactâncias. Isto fica visível no caso das funções oracionais, como, por exemplo as funções dos sistemas de TRANSITIVIDADE, MODO e TEMA, objetos da presente pesquisa.

A perfilação gramatical aqui apresentada se mostrou eficaz ao atender as necessidades principais apontadas pelas pesquisas de *corpus* e gramática (cf. HALLIDAY, 1992; TOGNINI-BONELLI, 2001; CONRAD, 2010), a saber, a mudança de foco do léxico para a gramática; da expressão para o conteúdo; e da instância para o sistema.

No que diz respeito à mudança do léxico para a gramática, a perfilação não considera os itens lexicais e suas relações, mas, de outra forma, busca generalizar as funções desses itens para uma unidade determinada e, segundo essas generalizações, constitui o *corpus*. Neste trabalho a unidade selecionada foi a oração e as generalizações foram os sistemas de TRANSITIVIDADE, MODO e TEMA.

Quanto à mudança da expressão para o conteúdo, a perfilação se liberta da obrigatoriedade de a palavra gráfica/*token* ser a unidade de análise. Ao partir da distância topológica entre as funções e o emprego das distâncias como valores categóricos, estes se emparelham aos *tokens*, gerando um novo tipo de unidade de investigação de *corpora*, aqui denominada “*token* gramatical”.

Por fim, ao tratar da relação entre instância e sistema, a perfilação gera resultados que podem alimentar a própria descrição gramatical, advindos da busca por padrões de ocorrência, coocorrência e agrupamentos de funções gramaticais que, sem as frequências não seria possível observar. No que diz respeito a esse tipo de frequência, Biber *et al.* (2004, p. 376) afirmam:

A frequência, por si só, não possui qualquer poder explanatório. Na verdade, acontece o contrário: a frequência mostra padrões que, esses sim, precisam ser explicados. O valor da frequência, e da pesquisa em *corpus* como um todo, está no fato de ela identificar padrões que, de outra forma, escapariam ao analista.⁴

A perfilação gramatical segue exatamente o que afirmam Biber *et al.* (2004), por assim dizer, fazendo o caminho inverso. Ela mostra como a explicação já existente para as funções gramaticais pode ser vista na frequência. É nesse sentido que ela alimenta a descrição.

Ao realizar essas mudanças, a perfilação gramatical apresentada neste artigo aponta como, a partir da expressão, do léxico e da instância – ou o formato em que o *corpus* é coletado – é possível extrair os dados para compilar e anotar um *corpus* para funções gramaticais. A partir deste, retomando a ideia da construção de uma microteoria sobre o *corpus*, realizam-se descrições de como as funções gramaticais operam. Neste artigo foi possível oferecer um Exemplo de Pesquisa realizado com a metodologia em questão.

Focando-se na função com número maior de ocorrências, o artigo mostrou um conjunto de possibilidades de análise advindas da perfilação gramatical, tais como a frequência da função, a distribuição nos textos, a dispersão ao longo dos textos, bem como os agrupamentos que forma, nos padrões oracionais, com outras funções. Para as pesquisas de *corpus* de base gramatical, a perfilação gramatical mostrou como é possível compilar e extrair dados de *corpora*, inclusive para a forma como são empregados na dinâmica textual.

Referências:

BIBER, D.; CONRAD, S.; CORTES, V. **If You Look at...** : Lexical Bundles in University Teaching and Textbooks. *Applied Linguistics*, Oxford, n. 25(3), p. 371–405, 2004.

CÂNDIDO-JR, A. **Análise bidirecional da língua na simplificação sintática em textos do português voltada à acessibilidade digital**. 2013. 225f. Tese (Doutorado em Ciências de Computação e Matemática Computacional). Instituto de Ciências Matemáticas e da Computação, Universidade de São Paulo, São Carlos, 2013.

⁴ We do not regard frequency data as explanatory. In fact we would argue for the opposite: frequency data identifies patterns that must be explained. The usefulness of frequency data (and corpus analysis generally) is that it identifies patterns of use that otherwise often go unnoticed by researchers.

CONRAD, S. What can a *corpus* tell us about grammar? In: O'KEEFFE, A.; McCARTHY, M. (Eds.). **The Routledge Handbook of corpus Linguistics**. Londres e Nova Iorque: Routledge, 2010, p. 227-240.

HALLIDAY, M. A. K. **System and function in language**. London: Oxford University Press, 1976.

HALLIDAY, M. A. K. Language as System and Language as Instance: the corpus as a Theoretical Construct. In: SVARTVIK, J. (Ed.) **Directions in corpus Linguistics**: Berlim e New York: Mouton de Gruyter, 1992, p.61-77. **crossref** <http://dx.doi.org/10.1515/9783110867275.61>

HALLIDAY, M.A.K. On Grammar and Grammaticals. In: HASAN, R.; CLORAN, C.; BUTT, D. (Eds.) **Functional Descriptions: Theory in Practice**. Amsterdam/ Philadelphia: John Benjamins, 1996, p.1-38. **crossref** <http://dx.doi.org/10.1075/cilt.121.03hal>

HALLIDAY, M. A. K. **On grammar**. London: Continuum, 2002. (The collected works of M. A. K. Halliday, v. 1).

HALLIDAY, M. A. K.; JAMES, Z. A Quantitative study of polarity and primary tense in the English finite clause. In: SINCLAIR, J.; HOEY, M.; FOX, J. (Eds.). **Techniques of description: spoken and written discourse**. London: Routledge, 1993.

HALLIDAY, M. A. K.; MATTHIESSEN, C. **An introduction to functional grammar**. 3a ed. London: Edward Arnold, 2004.

IEZZI, G.; MURAKAMI, C.; DOLCE, O.; HAZZAN, S. **Fundamentos de Matemática Elementar**. São Paulo: Atual Editora, vol. 1-10, 1977/78.

KE, S-W. Clustering a translational corpus. In: OAKES, M.; MENG J. (Eds.). **Quantitative Methods in corpus-Based Translation Studies: A Practical Guide to Descriptive Translation Research**. Amsterdam: John Benjamins, 2012. **crossref** <http://dx.doi.org/10.1075/scl.51.06ke>

LEMKE, J. Text Production and Dynamic Text Semantics. In: VENTOLA, E. (Ed.). **Functional and systemic linguistics: approaches and uses**. Berlim: Mouton de Gruyter. 1991. **crossref** <http://dx.doi.org/10.1515/9783110883527.23>

MARTIN, J. R.; MATTHIESSEN, C. Systemic typology and topology. In: CHRISTIE, F. (Ed.). **Literacy in Social Processes: papers from the inaugural Australian Systemic Linguistics Conference, held at Deakin University, January 1990**. Darwin: Centre for Studies in Language in Education, Northern Territory University. 1991.

MARTIN, J. R. **Systemic functional grammar: a next step into the theory – axial relations**. Pequim: Higher Education Press, 2013.

PAGANO, A.; FIGUEREDO, G.; LUKIN, A. **Modelling proximity in a corpus of literary retranslations: a methodological proposal for clustering texts based on systemic-functional**

annotation of lexicogrammatical features. In: International Quantitative Linguistics Conference. Olomouc: Univerzita Palackého v Olomouci, 2014, no prelo.

QUIROZ, B. **The interpersonal and experiential grammar of Chilean Spanish**: towards a principled systemic-functional description based on axial argumentation. 2013. 403f. Tese (Doutorado em Linguística). Universidade de Sydney, Sydney, 2013.

SCOTT, M. 2007. **WordSmith Tools**. Oxford: Oxford University Press.

SOUZA, A. Topologia geral de vários ângulos. 2013. Disponível em <<http://topologia-geral.ourproject.org>>. Acesso em: 10 set. 2014.

TOGNINI-BONELLI, E. corpus **linguistics at work**. Amsterdam e Philadelphia: John Benjamins, 2001. **crossref** <http://dx.doi.org/10.1075/scl.6>

VIANA, V. Linguística de *corpus*: conceitos, técnicas e análises. In: VIANA, V.; TAGNIN, S. (Orgs.). **Corpora no ensino de línguas estrangeiras**. São Paulo: HUB Editorial, 2011.

WHORF, B. L. **Language, thought, and reality**. Cambridge: MIT, 1987.

Artigo recebido em: 15.09.2014

Artigo aprovado em: 27.11.2014