

Reconhecimento de padrões de orações passivas por meio de Mapas Auto-organizáveis

Identifying Passive Clause Patterns Based on Self-Organizing Maps

*Gesieny Laurett Neves Damasceno**
gesienylaurett@gmail.com
Universidade Federal do Espírito Santo

*Violeta Virginia Rodrigues***
violetarodrigues@uol.com.br
Universidade Federal do Rio de Janeiro

RESUMO: O presente artigo busca lançar as bases para o desenvolvimento de um modelo descritivo capaz de distinguir, por intermédio da Rede Neural de Mapas Auto-organizáveis (Rede SOM), as diversas configurações que as orações passivas formadas com verbo auxiliar *ser* + particípio passado do verbo principal assumem em contextos específicos de interação. As 220 orações constitutivas do *corpus* foram alocadas, pela Rede SOM, em 46 neurônios, que representam diferentes exemplares de orações passivas, considerando, concomitantemente, as classes dos oito parâmetros selecionados por esta pesquisa. Esses 46 exemplares refletem, de forma bastante criteriosa, as variações das orações passivas em termos de configurações gramatical e pragmático-discursiva, considerando o aparato teórico da Gramática Sistêmico-Funcional. A definição do número ótimo de grupos possibilitou, ainda, a apreensão dos dados a partir de, basicamente, três importantes agrupamentos, que

* Possui doutorado em Letras Vernáculas (Língua Portuguesa) pela Universidade Federal do Rio de Janeiro - UFRJ (2012 - 2016), mestrado em Estudos Linguísticos pela Universidade Federal do Espírito Santo - UFES (2009 - 2011) e graduação em Língua Portuguesa e Literatura de Língua Portuguesa pela Universidade Federal do Espírito Santo - UFES (2004 - 2008). Atua, principalmente, nos seguintes temas: Gramática da Língua Portuguesa, Funcionalismo Linguístico, Gramática Sistêmico-Funcional. É professora Adjunta do Departamento de Línguas e Letras da Universidade Federal do Espírito Santo (UFES) e Professora Permanente do Programa de Pós-Graduação em Linguística da mesma universidade.

** Possui graduação em Português-Literaturas (1990), Especialização em Literatura Infanto-Juvenil (1991), Mestrado em Letras Vernáculas (1994) e Doutorado em Letras Vernáculas (2001) pela Universidade Federal do Rio de Janeiro (UFRJ). Fez estágio Pós-Doutoral (de 01 de agosto 2012 a 10 de abril de 2013 e de 07 de março a 07 de dezembro de 2019) sob supervisão da Professora Doutora Maria Beatriz Nascimento Decat (FALE / POS-LIN / UFMG). Atualmente é professor titular da Universidade Federal do Rio de Janeiro, integrando o quadro de docentes permanentes do Departamento de Letras Vernáculas - Setor de Língua Portuguesa -, atuando nos cursos de Graduação e no Programa de Pós-Graduação em Letras Vernáculas. Tem experiência na área de Letras, atuando principalmente nos seguintes temas: Língua Portuguesa, Sintaxe (Processos sintáticos: Subordinação, Coordenação, Correlação, Justaposição); Funcionalismo (Gramaticalização de conectores; Desgarramento e insubordinação de cláusulas), Ensino de Língua Portuguesa e Redação (correção e avaliação).

retratam os principais significados evocados pelas orações passivas no âmbito das 102 notícias jornalísticas analisadas.

PALAVRAS-CHAVE: Orações passivas. Gramática Sistêmico-Funcional. Rede Neural de Mapas Auto-organizáveis.

ABSTRACT: The aim of the current article is to lay the foundations for the development of a descriptive model capable of differentiating several configurations embodied by passive clauses formed with auxiliary verb *ser* (to be) + past participle of the main verb in specific interaction contexts, based on using the Neural Network of Self-Organizing Maps (SOM Network). All 220 constitutive clauses forming the herein analyzed *corpus* were allocated by the SOM Network in 46 neurons, which represent different samples of passive clauses, by concomitantly taking into consideration the classes of all eight parameters selected in the current research. These 46 samples reflect variations in the grammatical and pragmatic-discursive configurations of passive clauses, in a quite thorough manner, by taking into consideration the theoretical apparatus of Systemic-Functional Grammar. The definition of the optimal number of groups allowed us to deduce data from three important groupings that have portrayed the major meanings elicited by the analyzed passive clauses within the scope of all 102 analyzed journalistic news.

KEYWORDS: Passive clauses. Systemic-Functional Grammar. Neural Network of Self-Organizing Maps.

Introdução

Em virtude do interesse pelas diversas configurações que as orações passivas formadas pelo grupo verbal passivo *ser + participio do verbo principal* podem assumir no uso efetivo da linguagem, buscou-se, por meio deste estudo, a seleção mais sistemática de critérios descritivos capazes de estabelecer as similaridades e dissimilaridades das construções linguísticas, tendo em conta seus aspectos sintáticos, semânticos e pragmático-discursivos. As escolhas efetuadas na codificação das orações passivas em análise consideraram os três significados da oração, segundo a Gramática Sistêmico-Funcional (HALLIDAY, 1994; HALLIDAY; MATTHIESSEN, 2004; 2014): oração como representação (Sistema de TRANSITIVIDADE), oração como mensagem (Sistema de TEMA e INFORMAÇÃO) e oração como troca (Sistema de MODO). Os resultados encontrados endossam a ideia de que um tratamento mais coerente para as passivas deve levar em conta a plurifuncionalidade gramatical das orações de um modo geral e, assim, analisá-las a partir dos três significados mais gerais, ou seja, das chamadas *metafunções da linguagem*.

O *corpus* desta pesquisa constitui-se de 220 orações retiradas de 102 notícias pertencentes ao século XX (1901 a 2000). Os dados foram analisados a partir da Rede Neural de Mapas Auto-organizáveis (Rede SOM – *Self Organizing Maps*), que representa um modelo de inteligência artificial, cujo objetivo principal é o reconhecimento de padrões.

1 As orações passivas no quadro da Gramática Sistêmico-Funcional (GSF)

Para a teoria sistêmico-funcional, as línguas, de um modo geral, são organizadas em torno de três tipos principais de significados: o ideacional, o textual e o interpessoal. Esses componentes, chamados de metafunções, são a manifestação no sistema linguístico dos três objetivos principais que fundamentam todos os usos da linguagem: representação, mensagem e interação. Na abordagem sistêmico, cada uma das três metafunções relaciona-se com um sistema diferenciado: (i) a metafunção ideacional (oração como representação) é codificada pelo Sistema de TRANSITIVIDADE; (ii) a metafunção textual (oração como mensagem) é codificada pelo Sistema de TEMA e de INFORMAÇÃO; e (iii) a metafunção interpessoal (oração como troca) é codificada pelo Sistema de MODO.

O Sistema de TRANSITIVIDADE é responsável pela codificação e representação de algum processo na experiência humana. Sob essa perspectiva, a linguagem é utilizada para organizar, compreender e expressar as nossas percepções do mundo e de nossa própria consciência. A Gramática Sistêmico-Funcional (GSF) reconhece três tipos principais de processos pelos quais o ser humano representa suas experiências: *Material* (processo do 'fazer' e do 'acontecer'), *Mental* (processo do sentir) e *Relacional* (processo do ser, estar e ter). Na fronteira entre esses três há outros secundários: *Verbal* (processo do dizer), *Comportamental* (processo do comportar-se) e *Existencial* (processo do existir). A esses processos, encontram-se associados Participantes específicos (como os Participantes inerentes Ator, Experienciador, Portador, Dizente, Comportante e Existente, dos processos Material, Mental, Relacional, Verbal, Comportamental e Existencial, respectivamente) e, eventualmente, Circunstâncias (como Extensão, Localização, Modo, Causa, Acompanhamento etc.). Dentro do quadro da GSF, os conceitos de Processo, Participante e Circunstância são categorias semânticas que explicam, de forma

bastante geral, como os fenômenos do mundo real são representados em termos de estruturas linguísticas.

O Sistema de TEMA (Tema + Rema) e de INFORMAÇÃO (Dado + Novo) diz respeito ao ponto de partida para a mensagem, ou seja, ao elemento que o falante seleciona para fundamentar o que ele vai dizer. Sob essa perspectiva, a linguagem é utilizada para relacionar o que é dito (ou escrito) para o mundo real. Embora estejam relacionados, Dado + Novo e Tema + Rema não são a mesma coisa. O Tema é o que o falante escolhe tomar como ponto de partida, e o Dado é o que o ouvinte já conhece ou tem acessível. Tema + Rema é, portanto, orientado ao falante, enquanto o Dado + Novo é orientado ao ouvinte. Mas ambos são, é claro, selecionados pelo falante (HALLIDAY; MATTHIESSEN, 2014, p. 119).

Simultaneamente a sua organização como representação e mensagem, a oração também é organizada como um evento interativo envolvendo falante¹ (ou escritor) e ouvinte (HALLIDAY; MATTHIESSEN, 2014, p. 134). No ato de falar, o falante adota para si um determinado papel de fala e, ao fazê-lo, atribui ao ouvinte um papel complementar que ele deseja que ele adote. Por exemplo, ao fazer uma pergunta, o falante está assumindo o papel de buscador da informação e exigindo que o ouvinte assuma o papel de fornecedor da informação demandada. Dessa feita, a metafunção interpessoal diz respeito à interação entre falante e ouvinte. Sob essa perspectiva, a linguagem é utilizada para nos permitir participar de atos comunicativos com outras pessoas, para assumir papéis e para expressar e entender os sentimentos, atitudes e julgamentos. Como foi dito anteriormente, a parte da oração que desempenha a metafunção interpessoal é chamada sistema de MODO, em que a oração se organiza em torno de dois componentes básicos: o Modo² (Sujeito e operador Finito) e o Resíduo (Predicador, Complemento(s) e Adjunto(s)).

A respeito dos grupos verbais ativos e passivos, Halliday e Matthiessen (2014) explicam que, do ponto de vista experiencial, as orações são as mesmas, ou seja, tanto as orações ativas como as passivas representam uma configuração de Ator + Processo + Meta, quando o Processo é Material; Experienciador + Processo +

¹ Assim como Halliday e Matthiessen (2014), usaremos o termo 'falante' como um termo que abrange tanto o falante como o escritor.

² Modo (com inicial maiúscula) diz respeito ao nome de um elemento da estrutura interpessoal da oração (Modo + Resíduo) e MODO (todas as letras maiúsculas) diz respeito ao sistema interpessoal primário da oração. Essa representação segue a convenção geral segundo a qual nomes de funções estruturais são escritos com inicial maiúscula e nomes de sistemas com letras maiúsculas.

Fenômeno, quando o Processo é Mental; Dizente + Processo + Verbiagem, quando o Processo é Verbal; e Comportante + Processo + Fenômeno, quando o Processo é Comportamental³. No entanto, quando se consideram as funções interpessoais na estrutura modal da oração, as orações ativas e passivas se diferem no modo como esses papéis são mapeados. Na variante ativa, os Participantes inerentes (Ator, Experienciador, Dizente e Comportante) são codificados como o Sujeito da oração, por isso, é atribuída a eles a responsabilidade modal. Já na variante passiva (ou receptiva), os Participantes adicionais (Meta, Fenômeno e Verbiagem) é que são codificados como o Sujeito da oração, por isso, é atribuída a eles a responsabilidade modal, conforme se constata no exemplo a seguir, construído com um Processo Material (HALLIDAY; MATTHIESSEN, 2014, p. 228):

(1) Oração Material transitiva operativa:

<i>O leão</i>	<i>pegou</i>		<i>o turista</i> ⁴
Ator	Processo: Ativo		Meta
Sujeito	Finito	Predicador	Adjunto
Modo		Resíduo	
Tema	Rema		

(2) Oração Material transitiva receptiva, com Processo realizado por grupo verbal passivo:

<i>O turista</i>	<i>foi</i>	<i>pego</i>	<i>pele leão</i> ⁵
Meta	Processo: Passivo		Ator
Sujeito	Finito	Predicador	Adjunto
Modo		Resíduo	
Tema	Rema		

Os exemplos anteriores permitem constatar que, na variante ativa (ou operativa), o Ator (*O leão*) é codificado como Sujeito da oração, e a Meta (*o turista*) é codificada como Complemento. No entanto, na variante passiva (ou receptiva), é a Meta que é codificada como Sujeito da oração, e o Ator tem o *status* de um Adjunto (*pele leão*) e,

³ Expandimos a noção apresentada em Halliday e Matthiessen (2014, p. 227) para os outros processos que, além dos processos Materiais, podem realizar orações na voz passiva: os Mentais, os Verbais e os Comportamentais.

⁴ Tradução livre de *the lion caught the tourist* (HALLIDAY; MATTHIESSEN, 2014, p. 228).

⁵ Tradução livre de *the tourist was caught by the lion* (HALLIDAY; MATTHIESSEN, 2014, p. 228).

como Adjunto, pode de fato ser deixado de fora: *o turista foi pego pelo leão*; *o turista foi pego*. Portanto, dentro do quadro analítico da GSF, assume-se que o termo tradicionalmente chamado de “agente da passiva” se encontra numa relação de adjunção com o núcleo da predicação.

2 A constituição do *corpus* e a metodologia de análise dos dados

O *corpus* desta pesquisa é composto por 220 orações passivas formadas pelo grupo verbal passivo *ser + particípio do verbo principal*, extraídas de 102 notícias jornalísticas. Essas notícias fazem parte do *Corpus VARPORT*, um banco de dados criado pela Faculdade de Letras da Universidade Federal do Rio de Janeiro em parceria com pesquisadores de Portugal, e foram escritas em quatro fases distintas, ao longo do século XX: Fase 1 (de 1901 a 1924); Fase 2 (de 1925 a 1949); Fase 3 (de 1950 a 1974); e Fase 4 (de 1975 a 2000).

Dentro do panorama teórico da GSF, o significado de um texto só pode ser explicado a partir da relação do sistema linguístico como um todo, ou seja, para se alcançar a interpretação do texto, é preciso que haja a interpretação do contexto, da situação, da cultura e da sistemática relação entre texto e contexto. De acordo com a teoria sistemicista, um texto ocorre em dois contextos, um dentro do outro: o *contexto de cultura* e o *contexto de situação* (HALLIDAY; MATTHIESSEN, 2004; BUTT, 2000).

Quando considerado o contexto mais geral das notícias jornalísticas, ou seja, do contexto de cultura, tem-se que alguns importantes eventos históricos foram determinantes para a construção dos significados instanciados nesse gênero de texto, tanto no cenário mundial como no nacional. Dentre esses eventos marcantes, citam-se: os efeitos políticos e sociais da Proclamação da República (ocorrida em 1889); a epidemia da Gripe Espanhola (1918); os conflitos bélicos, como a 1ª Guerra Mundial (1914-1918) e a 2ª Guerra Mundial (1939-1945); a ascensão de Hitler ao poder de chefe dos nazistas (1930); no Brasil, o Movimento Diretas Já (1983-1984); o fim do Regime Militar (1985); a promulgação da Constituição (1988), o *Impeachment* do presidente Fernando Collor (1992) e o assassinato do ativista Chico Mendes (1988). Como é próprio do gênero notícia, cumpriu aos 102 textos analisados a função de informar ao leitor, de forma objetiva, acerca desses e de outros fatos, considerados mais relevantes na sociedade.

Se as notícias forem consideradas a partir das variáveis que compõem o contexto de situação, os seguintes aspectos caracterizam esse gênero: (i) *Campo* (natureza da prática social) – relato de uma série de fatos, a partir do fato considerado mais importante ou interessante; *Relação* (ligação entre os falantes) – o autor é alguém que possui a informação e o poder de decidir quais são os elementos mais importantes dos acontecimentos; o leitor é o público em geral; a relação entre eles é de natureza semiformal; (iii) *Modo* (meio de transmissão) – o canal é o jornal, o meio é escrito e a linguagem tem papel constitutivo.

Com vistas à discriminação das orações passivas à luz de critérios gramaticais e discursivos, como requer uma pesquisa funcionalista, após a separação das 220 orações constitutivas do *corpus*, por meio de leitura criteriosa, procedeu-se ao levantamento de parâmetros descritivos capazes de evidenciar os diferentes sentidos evocados pelas orações passivas. Num primeiro momento, tomando como referência o modelo desenvolvido por Damasceno (2016) e Damasceno e Rodrigues (2017a; 2017b) para as orações Materiais ativas e as leituras previamente efetuadas acerca do fenômeno em análise, elencou-se um total de 15 parâmetros descritivos. Todas as orações foram analisadas com base nesses 15 parâmetros e, na sequência, os dados da pesquisa foram padronizados, com o objetivo de conceder a mesma importância a cada uma das classes dos parâmetros durante a análise na Rede SOM.

Após a padronização da planilha, os dados foram submetidos ainda à verificação da magnitude de multicolinearidade, através do número de condição (NC), e à análise da correlação simples (STEEL; TORRIE, 1980). Essas medidas fizeram-se necessárias, a fim de que fossem descartados os parâmetros altamente correlacionados ou aqueles que não estivessem contribuindo, efetivamente, para a discriminação das orações. Constatada a existência de severa multicolinearidade entre alguns dos 15 parâmetros selecionados, submeteu-se a planilha de dados ao método do cosseno quadrado, visando à redução do efeito de multicolinearidade por intermédio do descarte de parâmetros. O método do cosseno quadrado sugeriu a exclusão de sete parâmetros. Foram eles: (i) período em que a notícia foi escrita; (ii) assunto da notícia; (iii) tipo de Processo Material; (iv) formas de expansão das orações; (v) estatuto sintático da passiva; (vi) tempo e modo do verbo auxiliar da perífrase verbal; (vii) polaridade da oração.

Realizadas as sete exclusões sugeridas pelo método do cosseno quadrado, chegou-se, portanto, a um modelo analítico para as orações passivas que considera

oito parâmetros descritivos. Esses parâmetros, suas abreviaturas e suas classes constitutivas estão listados a seguir:

1. *Tipos de Processo* (TIP): Material, Verbal, Mental e Comportamental.
2. *Número de participantes* (NUP): Zero, Um (Sujeito), Um (SPrep), Dois (Sujeito e Objeto Indireto), Dois (Sujeito e SPrep), Três (Sujeito, Objeto Indireto e SPrep).
3. *Ordem dos constituintes* (OSI): [Per. Verbal, Per. Verbal + Sprep], [Per. Verbal + SN (Suj.)], [Per. Verbal + SN (Suj.) + SPrep], [Per. Verbal + Sprep + SN (Suj.)], [SN (Suj.) + Per. Verbal], [SN (Suj.) + Per. Verbal + SPrep], [SPrep + Per. Verbal + SN (Suj.)], [Verbo auxiliar + SN (Suj) + Particípio verbo principal + SPrep], [Verbo auxiliar + SN (Suj.) + Particípio verbo principal].
4. *Forma de realização do Participante responsável pelo processo* – SPrep (REC): Expresso e Não Expresso.
5. *Núcleo do SPrep que codifica o responsável pelo processo* (NSP): Por, Com, Em, Não se aplica.
6. *Estatuto informacional do Participante responsável pelo processo* – SPrep (INF): Expresso, Referência ancorada no texto, Referência inferida, Referência desconhecida.
7. *Papel semântico do Participante responsável pelo processo* – SPrep (PAS): Agente, Causa, Modo, Metonímia (referência mais genérica), Desconhecido.
8. *Sistema de Informação do Participante responsável pelo Processo (Tema ou Rema)* (SIF): Tema, Rema, Não se aplica.

Conforme é possível constatar, cada um dos oito parâmetros visa a colocar em evidência um aspecto diferente das orações passivas, tanto em suas dimensões gramaticais como em suas dimensões semântico-discursivas. Ao nível gramatical, pertencem os seguintes parâmetros: *Número de participantes* (NUP), *Ordem dos constituintes* (OSI), *Forma de realização do Participante responsável pelo processo* (REC) e *Núcleo do SPrep que codifica o responsável pelo processo* (NSP). Ao nível semântico-discursivo, os seguintes parâmetros: *Tipos de Processo* (TIP), *Estatuto informacional do Participante responsável pelo processo* (INF), *Papel semântico do Participante responsável pelo processo* (PAS), *Sistema de Informação do Participante responsável pelo Processo* (SIF).

A identificação dos padrões de orações passivas foi feita por intermédio da Rede Neural de Mapas Auto-organizáveis (Rede SOM – *Self Organizing Maps*), que representa um modelo de inteligência artificial. Criados por Teuvo Kohonen em 1982,

os Mapas Auto-organizáveis são mais um dos importantes algoritmos disponíveis para organização de informações, com a notável característica de representarem visualmente a classificação e a distribuição dos dados através de seu posicionamento espacial em uma rede ou grade (DAMASCENO, 2016; DAMASCENO; RODRIGUES, 2017a; 2017b).

Para o processamento estatístico dos dados dessa etapa, foi utilizado o pacote “kohonen” (WEHRENS; BUYDENS, 2007). Todas as análises estatísticas foram realizadas no Programa R versão 3.1.2 (R DEVELOPMENT CORE TEAM, 2011). As definições quanto ao tamanho do mapa de grade (quantidade de neurônios) e quanto ao número ótimo de macroagrupamentos foram feitas com base na metodologia estipulada por Kohonen (1982; 2001; 2013): com base nessa metodologia, definiu-se a Rede SOM com uma tipologia (*grid map*) constituída por 15 x 14 neurônios (portanto, 210 neurônios).

Seguida à estruturação do *grid map*, a Rede SOM foi treinada, com o objetivo de obter a melhor precisão possível com relação aos modelos localizados dentro dos neurônios. Nesse treinamento, buscou-se a estimativa do menor erro topológico durante o treinamento: no caso dos dados desta pesquisa, o menor erro foi estimado em 0,014.

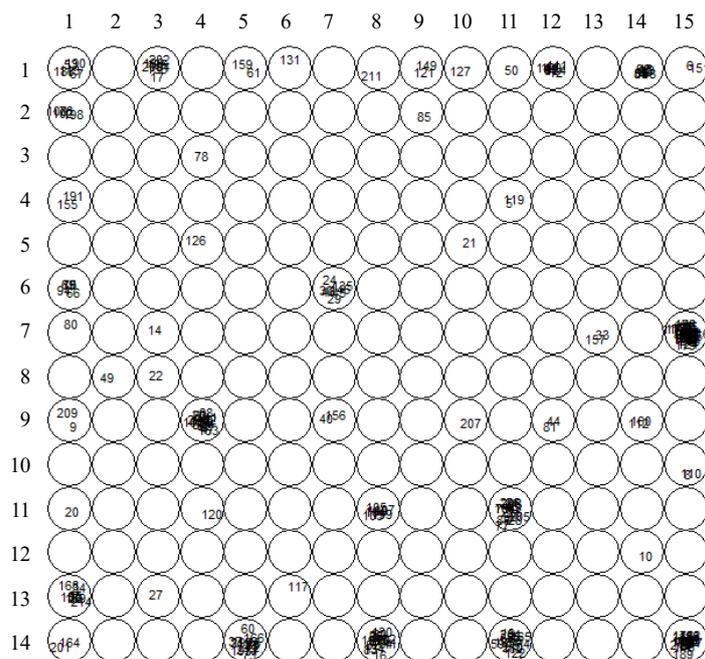
Por intermédio do pacote “kohonen” (WEHRENS; BUYDENS, 2007), estimaram-se os modelos de orações passivas presentes nas notícias jornalísticas, os macroagrupamentos das orações, o número de orações por neurônio (*Counts Plots*) e o peso dos parâmetros em cada neurônio (*Weight vectors*). Com o intuito de descrever os padrões de orações passivas mais significativos nos textos analisados, estimou-se, por meio do pacote *cValid* (BROCK et al., 2008), o número ótimo de grupos.

3 Os padrões das orações passivas e os significados construídos no contexto das notícias jornalísticas

Considerando o grau de semelhança entre as classes que compõem os oito parâmetros selecionados, as 220 orações constitutivas do *corpus* deste estudo foram alocadas pela Rede SOM em 46 neurônios distintos, num *grid map* formado por 210 neurônios (15 x 14). A Figura 1, a seguir, de natureza bidimensional, permite a visualização quanto à distribuição das orações sobre o mapa topológico, bem como a

identificação dos modelos formados a partir da semelhança entre essas orações. A finalidade da Figura 1 é exibir, de forma mais abrangente, o modo de organização dos dados pela Rede SOM: os neurônios efetivamente preenchidos retratam os exemplares de orações passivas encontrados nas notícias jornalísticas analisadas.

Figura 1: Exemplos das orações passivas, a partir da relação de similaridade entre as classes que compõem os oito parâmetros descritivos selecionados.



Fonte: Gerado pelo Programa R versão 3.1.2

Ao lançar mão dos Mapas Auto-organizáveis, tem-se como objetivo principal aproximar as orações constitutivas do *corpus* que são semelhantes entre si, a fim de reconhecerem-se agrupamentos e padrões. As sete orações que compõem o neurônio 82 da Figura 1 (Linha 6, Coluna 7), por exemplo, possuem o mesmo padrão gramatical-discursivo, quando considerados os oito parâmetros de descrição arrolados. São eles: (i) o Processo codificado é do tipo Material – expressa a noção de que alguma entidade ‘faz’ algo – que pode ser feito ‘para’ alguma outra entidade; (ii) há apenas um participante expresso na oração: a Meta (na função de sujeito da oração); (iii) a ordem dos constituintes segue o padrão *Per. Verbal + SN (Suj.)* – há, portanto, uma inversão concernente à ordem canônica, preferencial (SV[C]); (iv) o SPrep que codifica o responsável pelo processo (participante inerente) não se encontra linguisticamente realizado – há, portanto, um apagamento desse constituinte

sintático; (v) devido ao apagamento do SPrep que codifica o “fazedor” da ação, não é possível recuperar o núcleo do SPrep; (vi) Apesar de o responsável pelo processo (o Ator) não estar linguisticamente expresso, essa informação pode ser recuperada no próprio texto (Informação ancorada no texto); (vii) o participante responsável pelo processo é um agente prototípico, ou seja, é uma entidade [+instigadora, +controladora]; (viii) devido ao apagamento do SPrep que codifica o “fazedor” da ação, não é possível classificá-lo quanto ao *status* informacional (Tema ou Rema).

A oração a seguir, constitutiva do neurônio 82, é um exemplo dessa configuração gramatical-discursivo:

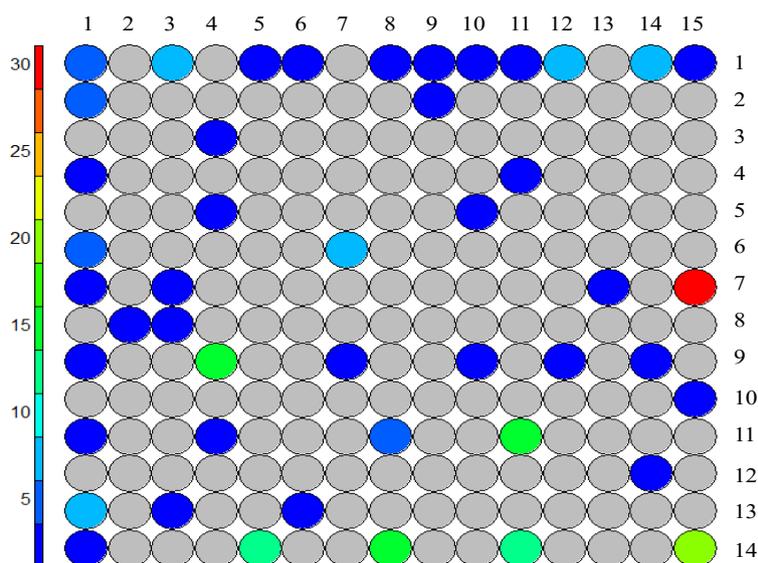
(3) Foi extinta a praticagem dos portos do Sul

O Dr. Frederico Burlamaqui, Director-Presidente do Lloyd Brasileiro, tendo em vista a inutilidade de manter a praticagem de alguns portos do Sul, determinou aos agentes de Santos, Paranaguá, Florianópolis e S. Francisco que extinguissem esse serviço creado por ocasião da guerra. (VARPORT, E-B-91-Jn-015)

Tal como ocorre com o neurônio 82, os demais neurônios formados pela Rede SOM (Figura 1) reúnem orações com as mesmas configurações léxico-gramaticais e pragmático-discursivas. Dessa feita, depreende-se que as notícias jornalísticas analisadas possuem 46 exemplares diferentes de orações passivas, considerando, concomitantemente, as classes selecionadas por esta pesquisa. Esses diferentes padrões e os sentidos evocados por eles serão mais bem apresentados em parágrafos subsequentes. A Figura 2, a seguir, denominada pela Rede SOM *Counts plot*, informa, por intermédio de cores, a quantidade de orações alocadas dentro dos exemplares de oração efetivamente realizados no *corpus* da pesquisa:

Figura 2: Frequência dos padrões encontrados: quantidade de orações por neurônio⁶.

⁶ A cor cinza indica os neurônios vazios, ou seja, aqueles exemplares que não tiveram ocorrência no *corpus*.



Fonte: Gerado pelo Programa R versão 3.1.2

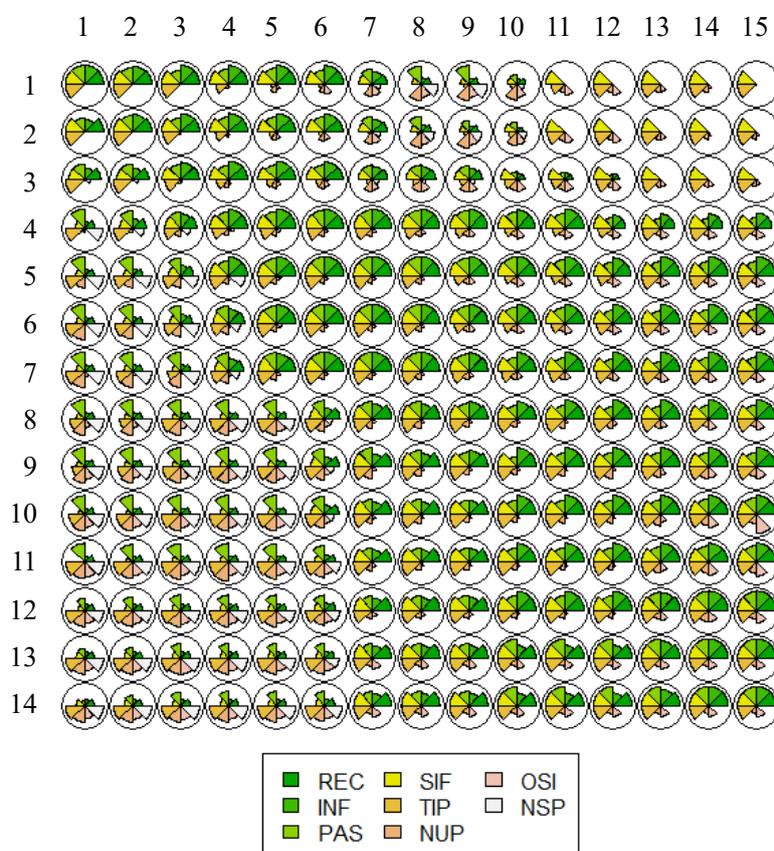
O modelo de oração passiva mais recorrente nas notícias jornalísticas, representado na Figura 2 pelo neurônio em vermelho, possui as seguintes características: (i) o Processo codificado é do tipo Material – expressa a noção de que alguma entidade ‘faz’ algo – que pode ser feito ‘para’ alguma outra entidade; (ii) há apenas um participante expresso na oração: a Meta (na função de sujeito da oração); (iii) a ordem dos constituintes segue o padrão *SN (Suj.) + Per. Verbal* – portanto, o sujeito-paciente é o Tema da oração; (iv) o SPrep que codifica o responsável pelo processo (participante inerente) não se encontra linguisticamente realizado – há, portanto, um apagamento desse constituinte sintático; (v) devido ao apagamento do SPrep que codifica o “fazedor” da ação, não é possível recuperar o núcleo do SPrep; (vi) apesar de o responsável pelo processo (o Ator) não estar linguisticamente expresso, essa informação pode ser recuperada no próprio texto (Informação ancorada no texto); (vii) o participante responsável pelo processo possui uma natureza mais genérica, metonímica (especialmente do tipo *Instituição pelo Responsável*); (viii) devido ao apagamento do SPrep que codifica o “fazedor” da ação, não é possível classificá-lo quanto ao *status* informacional (Tema ou Rema). A oração a seguir, constitutiva do neurônio 105, é um exemplo dessa configuração gramatical-discursivo:

(4) Os motoristas mostram-se alarmados com o novo sistema de sinalização imposto pelo Serviço de Trânsito em Botafogo, no trecho compreendido entre as Ruas Voluntários da Pátria e passagem. Nada menos de cinco sinais luminosos foram ali

instalados e distribuídos estrategicamente, mas de tal forma que gera a confusão. (VARPORT, E-B-93-Jn-003)

Até agora, as informações geradas pela Rede SOM foram consideradas a partir dos exemplares representados por cada neurônio, ou seja, de forma mais individualizada. No entanto, conforme se explicou na seção destinada à metodologia, no mapa topológico, as localizações espaciais dos neurônios desempenham uma importante função, pois são indicativas das características estatísticas intrínsecas contidas nos padrões de entrada (HAYKIN, 2001). Uma maneira bastante eficiente para a visualização mais geral acerca desses diferentes modelos das orações constantes do nosso *corpus* encontra-se apresentado na Figura 3, a seguir (*Weight vectors*). Por intermédio das representações constantes de cada neurônio, é possível depreender as informações a respeito dos oito parâmetros elencados para análise dos dados, já que cada parte desses mapas diz respeito a um parâmetro, como é possível verificar na legenda de cores da referida imagem:

Figura 3: Representatividade dos parâmetros em cada neurônio



Fonte: Gerado pelo Programa R versão 3.1.2

Legenda: **TIP:** Tipo de Processo; **NUP:** Número de participantes; **OSI:** Ordem dos constituintes na oração; **REC:** Forma de realização do SPrep responsável pelo processo; **NSP:** Núcleo do SPrep responsável pelo processo; **INF:** Estatuto informacional do SPrep responsável pelo processo; **PAS:** Papel semântico do SPrep responsável pelo processo; **SIF:** Sistema de Informação do Participante responsável pelo Processo – Tema ou Rema (Tema/Rema).

A partir da leitura da Figura 3, pode-se perceber, por exemplo, que as orações representativas dos neurônios 196 (Linha 14, Coluna 1) possuem uma configuração gramatical-discursiva, em termos dos parâmetros selecionados, muito diferente daquelas orações representativas do neurônio 210 (Linha 14, Coluna 15). Ao serem consideradas as classes constitutivas dos oito parâmetros arrolados por esta pesquisa, tem-se que os exemplares gramatical-discursivos representados pelos neurônios 196 e 210 se aproximam somente pelo quesito *Tipo de Processo*, já que ambos codificam experiências do tipo Material. Nos demais quesitos, os dois modelos de oração passiva veiculam significados bastante distintos, como o fato de, nas orações representativas do neurônio 196, ser atribuído ao Participante Ator, linguisticamente realizado, um papel temático não agentivo (de Meio), ao passo que, nas orações representativas do neurônio 210, esse participante é do tipo agentivo (essa informação encontra-se ancorada no texto). As seguintes orações exemplificam esses dois modelos, respectivamente:

(5) Ao final de mais de 60 discursos, numa das mais longas (16 horas) e tensas sessões de sua história, o Congresso Nacional rejeitou as primeiras horas de hoje, por não ter alcançado quorum constitucional, a emenda Dante de Oliveira que previa eleições diretas já para a Presidência da República. [...] Por todo o país o povo reuniu-se nas praças para acompanhar a contagem de votos em painéis chamados de placares das diretas. Em alguns lugares os votos dos parlamentares eram anunciados por sistemas de alto-falantes. (VARPORT, E-B-94-Jn-015) (Neurônio 196)

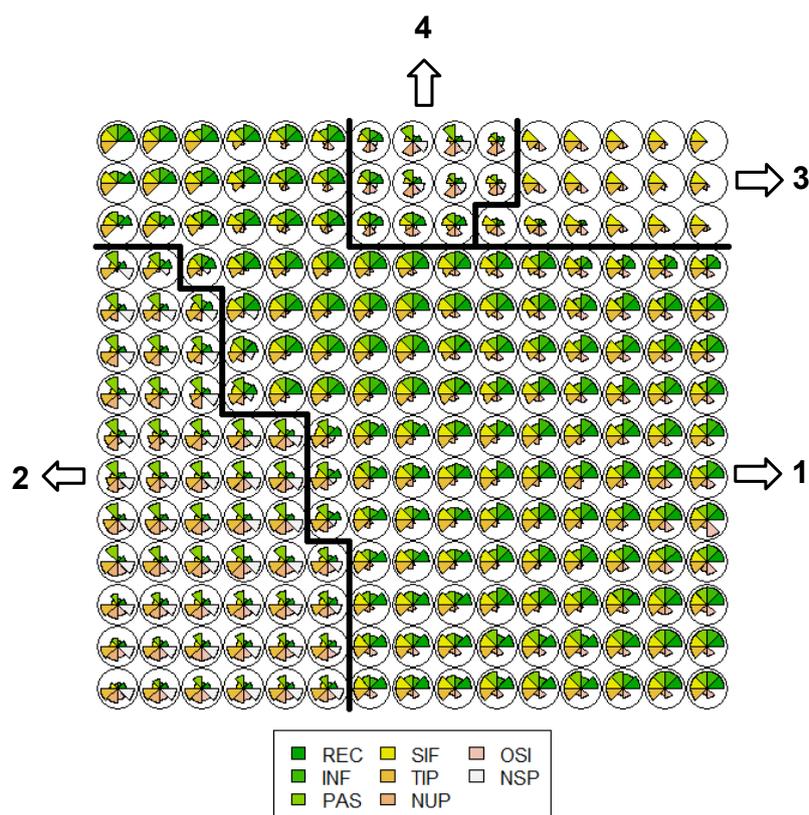
(6) O autor do atentado, o turco Mohamed Ali Agca, estava a cinco metros do jipe em que João Paulo percorria a praça. Os tiros foram disparados logo após a primeira volta, quando o Papa acabava de apertar as mãos de algumas das 30 mil pessoas que estavam na praça. (VARPORT, E-B-94-Jn-012) (Neurônio 210)

Na Figura 3, a representação mais geral dos modelos de oração passiva efetivamente realizados em nosso *corpus* e de outras configurações possíveis

(representadas pelos neurônios vazios) revela-se como um importante recurso para a análise dos padrões oracionais presentes no contexto das notícias analisadas, tendo em vista que esse mapa permite que sejam aproximados, intuitivamente, alguns neurônios a partir de suas semelhanças. Por exemplo, os neurônios dispostos no canto inferior direito da Figura 3 possuem alto grau de similaridade, o que faz supor que, de forma mais geral, eles pertencem a um mesmo padrão.

Mas, apesar da possibilidade de identificação de alguns agrupamentos por intermédio da Figura 3, a definição do número ótimo de grupos das orações passivas analisadas foi estatisticamente estimada, por meio do pacote cValid (BROCK et al., 2008). De acordo com essa estimativa, foram quatro os padrões oracionais mais significativos nas notícias analisadas. A partir de agora, as discussões considerarão esses padrões identificados, bem como os significados evocados por eles. Os quatro agrupamentos estão representados na Figura 4, a seguir:

Figura 4: Padrões das orações passivas nas notícias jornalísticas.



Fonte: Gerado pelo Programa R versão 3.1.2

No padrão 1, o mais recorrente do *corpus*, com 142 orações (64,5% do total de orações do *corpus*), os processos que codificam as experiências são do tipo Material (95,1%), Verbal (2,8%) e Mental (2,1%). Os participantes expressos nas orações são do tipo Meta, Verbiagem e Fenômeno – portanto, não há realização linguística, por meio do SPrep, dos participantes inerentes aos processos, ou seja, do Ator (processo Material), do Dizente (processo Verbal) e do Experienciador (processo Mental). Nesse padrão, a configuração gramatical mais recorrente é a do tipo *SN (Suj.) + Per. Verbal* (60,6%), ou seja, a ordem canônica, em que o sujeito é, também, o Tema da oração – isto é, o ponto de partida da mensagem. A segunda configuração mais recorrente, porém, com um percentual bem menor (26,8%), é a do tipo *Per. Verbal + SN (Suj.)*, em que ocorre a inversão da ordem preferencial SV[C]. Apesar de os participantes responsáveis pelos processos não virem expressos nas orações, sua referência pode ser recuperada na própria notícia (informação ancorada no texto) ou pode ser inferida, a partir dos conhecimentos compartilhados (informação inferida). Assim sendo, apesar de os adjuntos agente e experienciador (os tradicionalmente denominados “agentes da passiva”) não serem linguisticamente codificados nas orações pertencentes ao padrão mais frequente das orações passivas analisadas, esse apagamento não afeta a identificação do referente – ainda que, em sua maioria (62,7%), essa referência seja mais genérica, metonímica, do tipo *Instituição pelo Responsável*. A oração a seguir exemplifica o padrão 1:

(7) A polícia apreendeu hoje na sede do Club 5 de Julho 4 fuzis-metralhadoras, 18 fuzis, diversas granadas de mãos e cerca de 500 tiros. O club foi fechado. (VARPORT, E-B-92-Jn-003)

O padrão 2 foi o segundo mais recorrente nas notícias do século XX, com 54 orações (24,6% do total de orações do *corpus*). Nesse padrão, os processos que codificam as experiências são do tipo Material (92,6%) e Verbal (7,4%); os participantes expressos nas orações são do tipo Ator, Meta, Dizente e Verbiagem – portanto, diferentemente do padrão 1, aqui, os participantes inerentes aos processos, ou seja, o Ator (do processo Material) e o Dizente (do processo Verbal), são realizados linguisticamente, por meio do SPrep adjuntivo. Nesse padrão, tal como ocorreu no anterior, a configuração gramatical mais recorrente é a do tipo *SN (Suj.) + Per. Verbal + SPrep* (76%), ou seja, a ordem canônica, em que o sujeito é, também, o Tema da

oração – isto é, o ponto de partida da mensagem –, e o SPrep que codifica o participante inerente do processo é o Rema. No que diz respeito ao núcleo do sintagma preposicionado, a preposição *por* foi, de longe, a mais recorrente (98,1%). O tradicionalmente denominado “agente da passiva” codificou, para além do papel semântico de Agente, os papéis de Causa, Meio e a expressão metonímica do significado (uma referência mais genérica). Todavia, apesar da presença dessas outras funções semânticas, a natureza agentiva do SPrep foi a mais frequente, com 55,5% do total das orações enquadradas no padrão 2 – diferentemente do padrão 1, em que as expressões metonímicas foram as mais recorrentes. A oração subsequente é um exemplo do padrão 2:

(8) O telegramma em que o ministro da Marinha teve participação do grave facto [incêndio da canhoneira "Eber"] foi transmitido pelo commandante Mello Pinna, capitão do Porto de S. Salvador, às 8 e 15 da noite e recebido nesta capital às 8 e 30, e estava redigido nos termos seguintes: [...] (VARPORT, E-B-91-Jn-009)

As orações pertencentes aos agrupamentos 1 e 2 diferenciam-se, principalmente, quanto ao quesito realização linguística do participante responsável pelo processo (participante inerente): enquanto as orações do padrão 2 codificam essa informação por intermédio de um SPrep adjuntivo (com as propriedades semânticas de Agente, Causa e Meio), nas orações do padrão 1, esse constituinte é apagado. No entanto, apesar dessa não realização linguística, a identificação do referente é garantida por meio de retomada anafórica/catafórica ou de inferência.

Concernente ao padrão 3, a primeira constatação que se faz é a de que o seu percentual de realização no *corpus* foi bem abaixo dos percentuais atingidos pelos padrões 1 e 2: representou apenas 8,6% do total das orações passivas analisadas (19 orações). Os processos que codificam as experiências desse padrão são do tipo Material (95%) e Verbal (5%) – apenas uma realização do processo Verbal; os participantes expressos nas orações são, somente, do tipo Meta e Verbiagem – portanto, não há realização linguística, por meio do SPrep, dos participantes responsáveis pelos processos, ou seja, do Ator (processo Material) e do Dizente (processo Verbal). A configuração gramatical das orações pertencentes a esse padrão obteve, também, percentuais bastante distintos dos padrões anteriores: o tipo *SN (Suj.) + Per. Verbal*, que representa a ordem canônica, preferencial, representou

47,4% do total das orações (nove orações), e a ordem não preferencial *Per. Verbal + SN (Suj.)* representou 42,1% do total das orações (oito orações)⁷. Apreende-se, portanto, que, além da significativa presença da ordem preferencial (SV[C]), em que o sujeito da oração é, também, o seu Tema, esse padrão tende a favorecer a realização da ordem não preferencial (VS), ou seja, de uma oração neutra, nos termos de Furtado da Cunha (2000), em que nem o agente nem o paciente são embalados como o ponto de partida da oração (como Tema, portanto).

Outra característica que diferencia o padrão 3 dos demais padrões analisados até aqui é o fato de a referência do participante responsável pelo processo ser desconhecida. Conforme verificado no padrão 1, apesar de esses participantes não virem expressos nas orações, sua referência pôde ser recuperada na própria notícia, seja por anáfora/catáfora, seja por inferência. No padrão 3, no entanto, a oração passiva favorece o apagamento linguístico e discursivo do fazedor da ação, sem que, com isso, seja gerada uma oração mal formada do ponto de vista gramatical. No caso desse padrão, o apagamento do SPrep visa à ocultação do agente – seja por desconhecimento desse referente por parte do escritor, seja por preservação da sua face (do escritor e/ou do agente). A oração a seguir é uma amostra do padrão 3:

(9) Caracas, 24 (R.) - Foi divulgado extra-oficialmente que o Governo venezuelano nomeará seu embaixador em Moscou o Sr. José Rafael Pocaterra, atual embaixador em Londres. (VARPORT, E-B-92-Jn-028)

O padrão 4 pode ser considerado um subgrupo do padrão 2, visto que, de um modo geral, ele possui as mesmas características do segundo agrupamento. No entanto, enquanto o padrão 2 codificou os dois tipos de processo mais frequentes em nossos dados (o Material, com 92,3% do total das orações, e o Verbal, com 4,1%), foram agrupados no padrão 4 os processos menos frequentes, ou seja, o Mental, que representou 3,2% do total de orações do *corpus*, e o Comportamental, com apenas uma ocorrência, cujo percentual foi de 0,4%, considerando o *corpus* como um todo. Nesse agrupamento, os participantes expressos nas orações são do tipo Experienciador e Fenômeno, participantes dos processos Mentais, e Comportante e Fenômeno, participantes do processo Comportamental – portanto, assim como no

⁷ Nos outros dois casos, não foi possível recuperar a ordem do sujeito pelo fato de ter havido o apagamento desse constituinte.

padrão 2, aqui, os participantes inerentes aos processos, ou seja, o Experienciador e o Comportante, são realizados linguisticamente, por meio do SPrep adjuntivo. O padrão 4 possui apenas uma configuração gramatical, que é a do tipo *SN (Suj.) + Per. Verbal + SPrep*, ou seja, a ordem canônica, em que o sujeito é, também, o Tema da oração – isto é, o ponto de partida da mensagem –, e o SPrep que codifica o participante inerente do processo é o Rema. No que diz respeito ao núcleo do sintagma preposicionado, a preposição *por* foi também a mais recorrente (60%), apesar da presença das preposições *em* e *com* – cada uma com uma ocorrência. A seguinte oração passiva exemplifica o padrão 4:

(10) O julgamento foi assistido por cerca de 120 jornalistas, 30 dos quais representavam empresas de comunicação estrangeiras. (VARPORT, E-B-94-Jn-023)

Considerações finais

Assim como em outras áreas do conhecimento científico, o estudo da linguagem por intermédio do reconhecimento de padrões tem se revelado uma estratégia altamente promissora, tendo em vista que ela proporciona um panorama sistemático acerca dos arranjos de que o falante lança mão para ativar diferentes significados, considerando seus propósitos comunicativos. Todavia, para que o reconhecimento de padrões seja eficazmente efetivado, faz-se necessária a sistematização de um modelo descritivo-metodológico, cujas classes sejam capazes de distinguir as diversas configurações que as orações assumem nos contextos específicos de interação.

Com base, especialmente, no aparato teórico da Gramática Sistêmico-Funcional, buscou-se o desenvolvimento desse modelo descritivo das orações passivas e, ainda que esse modelo não esteja encerrado, pois outras categorias poderão (ou deverão) lhe ser acrescentadas, os resultados apresentados dão conta de que se trata de um caminho promissor, visto que as análises estatísticas empreendidas captaram padrões muito consistentes. As 220 orações constitutivas do *corpus* foram alocadas, pela Rede SOM, dentro dos neurônios a partir de um elevado grau de similaridade, permitindo a identificação de 46 diferentes exemplares de orações passivas, considerando, concomitantemente, as classes dos oito parâmetros selecionados por esta pesquisa.

Esses 46 exemplares refletem, de forma bastante criteriosa, as variações das orações passivas em termos de configurações gramatical e pragmático-discursiva no *corpus* analisado. E para além da possibilidade de identificação de cada exemplar de oração passiva presente no *corpus*, a definição do número ótimo de grupos (pacote *clValid*) possibilitou a análise dos dados a partir de, basicamente, três importantes agrupamentos, que retratam os principais significados evocados pelas orações passivas no âmbito das notícias jornalísticas.

REFERÊNCIAS

BROCK, G.; PIHUR, V.; DATTA, S.; DATTA, S. *clValid*: an R package for cluster validation. *Journal of Statistical Software*. v. 25. n. 4. Março/2008. p. 1-22.

DAMASCENO, G. L. N. *A transitividade de processos Materiais em notícias jornalísticas*. Rio de Janeiro, 2016. 248 p. Tese (Doutorado em Letras Vernáculas / Língua Portuguesa) – Faculdade de Letras, Universidade Federal do Rio de Janeiro.

DAMASCENO, G. L. N.; RODRIGUES, V. V. Os padrões de transitividade de cláusulas materiais e a construção de sentidos. *Revista de Estudos da Linguagem*, v. 25, n. 4, 2017a.

DAMASCENO, G. L. N.; RODRIGUES, V. V. Os processos materiais em notícias jornalísticas: seleção de parâmetros descritivos e análise de suas correlações. *Fórum Linguístico*, v. 14, n. 2, 2017b.

FURTADO DA CUNHA, M. A. A complexidade da passiva e as implicações pedagógicas do seu uso. *Linguagem & Ensino*, v. 3, n. 1, 2000.

FUZER, C.; CABRAL, S. R. S. *Introdução à gramática sistêmico-funcional em língua portuguesa*. Campinas, SP: Mercado de Letras, 2014.

HALLIDAY, M. A. K.; MATTHIESSEN, C. M. I. M. *An Introduction to Functional Grammar*. 3. ed. London: Arnold, 2004.

HALLIDAY, M. A. K.; MATTHIESSEN, C. M. I. *Halliday's Introduction to Functional Grammar*. 4. ed. London and New York: Routledge, 2014.

HAYKIN, S. *Redes neurais: princípios e prática*. 2. ed. Bookman Companhia, 2001.

KOHONEN, T. Self-Organized Formation of Topologically Correct Feature Maps. *Biological Cybernetics*. n. 43. Springer-Verlag: 1982. p. 59-69.

KOHONEN, T. *Self-Organizing Maps*. 3. ed. Berlin-Heidelberg, Germany: Springer, 2001.

KOHONEN, T. Essentials of the Self-Organizing Map. *Neural Networks*. n. 37. 2013. p. 52-65.

R DEVELOPMENT CORE TEAM. R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*, Vienna, Austria, 2011. Disponível em: <http://www.r-project.org/>. Acesso em: novembro de 2015.

STEEL, R. G. D.; TORRIE, J. H. *Principles and Procedures of Statistics: a Biometrical Approach*. 2. ed. New York: McGraw-Hill Book Company, 1980.

WEHRENS, R.; BUYDENS, L. M. C. Self- and Super-Organizing Maps in R: The Kohonen Package. *Journal of Statistical Software*. v. 21, n. 5, 2007. p. 1-19.

Recebido em 19/04/2023

Aceito em 12/06/2023

Publicado em 15/06/2023