

Busca de medida de complexidade textual numa abordagem funcionalista*

Search for textual complexity measure in a functionalist approach

João Bosco FIGUEIREDO-GOMES (UERN)
boscofigueiredo@yahoo.com.br

Wellington Vieira MENDES (UERN)
wvmendes@gmail.com

Recebido em: 27 de fev. de 2018.

Aceito em: 08 de jun. de 2018.

*Agradecimento ao CESPE/CEBRASPE, órgão financiador do projeto de pesquisa Padrões de complexidade textual: uma proposta funcionalista para avaliação de desempenho de alunos do ensino médio, de nossa autoria, concluído em 2017.

FIGUEIREDO-GOMES, João Bosco; MENDES, Wellington Vieira. Busca de medida de complexidade textual numa abordagem funcionalista. *Entrepalavras*, Fortaleza, v. 8, n. 2, p. 419-441, maio/ago. 2018.

Resumo: Desde o final dos anos 70, a busca por medidas de compreensão, de inteligibilidade e de complexidade de textos vêm interessando, principalmente, aos estudiosos de várias áreas, às instituições de avaliação de ensino e a professores de línguas no Brasil. Surgiram várias fórmulas e técnicas baseadas na estrutura do texto, além da avaliação subjetiva baseada em faixa etária e em preferências pessoais. No entanto, desconhece-se uma medida de complexidade textual que se baseie em aspectos verbais e visuais numa abordagem funcionalista. Este artigo relata uma investigação que objetiva buscar um instrumento de avaliação do nível de complexidade de textos verbais e visuais, por meio do emprego de recursos tecnológicos computacionais, baseado em categorias e princípios funcionalistas. Para tanto, tem, sobretudo, a Linguística Funcional e a Linguística de *Corpus* como orientação teórica e metodológica para a análise dos recursos verbais e visuais. Toma como amostra de estudo textos e provas do Exame Nacional de Ensino Médio (ENEM), submetidos a uma comparação quantitativa e estatística com dados de textos de referência do *Corpus* Lácio-Ref (ALUÍSIO *et al.*, 2003). A comparação quantitativa entre os *corpora* forneceu

dados para a elaboração de fórmulas de medida que avaliam, de forma escalar, textos de língua portuguesa, a partir de três níveis de complexidade: estrutural, cognitiva e visual. Conclui-se que a proposta é de fácil acesso e que, por meio do emprego de recursos tecnológicos computacionais, pode-se prover a análise da complexidade de apenas um texto como de textos em grande monta.

Palavras-chave: Funcionalismo linguístico. Complexidade Textual. Avaliação.

Abstract: Since the end of the 1970s, the search for measures of comprehension, intelligibility and complexity of texts has been interesting, mainly, to scholars from several areas, to institutions of teaching evaluation and to Language teachers in Brazil. Many formulas and techniques based on structural features of the text came up, besides the subjective evaluation based on age and personal preferences. However, it is unknown a textual complexity measure based on verbal and visual aspects in a functionalist approach. This article reports an investigation that aims to search an instrument of evaluation of the level of complexity of verbal and visual texts by using computational technological resources, based on categories and functionalist principles. For that, it has, above all, Functional Linguistics and Linguistics of Corpus as theoretical and methodological orientation for the analysis of the verbal and visual resources. It takes as study sample texts and tests of the National High School Exam (ENEM), submitted to a quantitative and statistical comparison with data of reference texts of the Corpus Lácio-Ref (ALUÍSIO *et al.*, 2003). The quantitative comparison between *corpora* has provided data for the elaboration of formulas of measures that evaluate, in a scalar way, Portuguese language texts, from three levels of complexity: structural, cognitive and visual. It is concluded that the proposal is of easy access and that by using computational technological resources, it can be provided the analysis of the complexity of only one text as well as of texts in large amounts.

Keywords: Linguistic Functionalism. Textual Complexity. Evaluation.

Introdução

O interesse pela temática da complexidade textual não é novo na literatura internacional nem na nacional, sobretudo nos estudos sobre leitura. Embora, nos anos 80, houvesse a ênfase de pesquisas sobre compreensão e estratégias de leitura, sobre tipificação de leitores e sobre elementos linguísticos associados ao nível de compreensão leitora, a preocupação com a legibilidade já era bastante antiga em, pelo menos 13 línguas (MOLINA, 1979; DANIELSON, 1987; FIGUEIREDO-GOMES, 1989), cujas fórmulas se baseavam, em sua maioria, no número de sentenças e de palavras.

Essas fórmulas de legibilidade (cf. revisões de KLARE, 1975) e técnicas baseadas na estrutura do texto, como o *cloze* (TAYLOR, 1953), baseavam-se, geralmente, além da avaliação subjetiva do examinador, na idade do leitor e em preferências pessoais. No entanto, desconhecemos um instrumento de medida de complexidade textual que se baseie em aspectos verbais e na composição verbo-imagética de textos numa abordagem funcionalista e que utilize recursos tecnológicos computacionais que facilitem o acesso aos interessados.

Considerando essa lacuna, este artigo objetiva apresentar um instrumento de avaliação do nível de complexidade textual, resultante de uma pesquisa empírica com textos reais, contemplando categorias e princípios funcionalistas. Para tanto, utilizamos a Linguística Funcional, segundo as vertentes norte-americana e sistêmico-funcional, além da Linguística de *Corpus*, como base de orientação teórica e metodológica para a análise dos recursos verbais e visuais, tendo como *corpus* de estudo os textos utilizados nas questões do ENEM da área Linguagens e Códigos e suas Tecnologias.

O artigo está dividido em mais quatro seções: a primeira contextualiza estudos sobre complexidade textual; a segunda apresenta uma síntese das teorias que abalizam esta pesquisa, tendo como foco a Linguística Funcional e, em especial, a Gramática do *Design Visual*; a terceira seção traz os procedimentos metodológicos vinculados à Linguística de *Corpus*; a quarta descreve, com base na comparação de *corpora*, uma ficha com escalas de avaliação do nível de complexidade textual a partir de categorias e de princípios funcionalistas.

A complexidade textual e os modelos de análise

Dada a falta de um conceito de *complexidade textual* (CT) na literatura disponível, e partindo da noção funcionalista de língua enquanto sistema, ou seja, como um conjunto de opções intrincadas, em que cada relação no conjunto implica sempre um movimento que mobiliza o complexo na sua inteireza (HALLIDAY; MATTHIESSEN, 2004; CAPRA; LUISI, 2014; MENDES, 2016), estamos definindo CT como *uma rede de sistemas de signos linguísticos e visuais integrados passíveis de determinar o grau de inteligibilidade de textos por diferentes usuários*.

Tomando como ponto de referência a noção de usos específicos e de variação de terminologias (sob a forma de sinônimos, paráfrases e explanações), Ciapuscio (1998) avaliou a complexidade em diferentes tipos de textos que tratavam de uma mesma temática, dirigidos a diferentes perfis de leitores (cientistas, público semi-leigo e leigo). Em Ciapuscio (2003), a autora propõe, a partir dessas noções iniciais, uma tipologização de quatro níveis (ou multinível): a) o nível funcional do texto (que tem relação com a função ou propósito); b) o situacional (que se relaciona aos interlocutores e ao tipo de comunicação envolvidos); c) o de conteúdo semântico (que inclui modos de tratamento e de apresentação do tema); d) o formal-gramatical, que inclui aspectos

gramaticais, lexicais e terminológicos. A cada um desses níveis corresponde uma gradação, tendo, na análise conjunta de todas essas variáveis, a identificação dos tipos de textos e, portanto, de seu nível de complexidade.

Nessa mesma direção, Biber (1988) propõe uma análise multidimensional (AMD) que possibilita a associação de análises de nível geral (do *corpus*) com análises de nível textual (um texto no *corpus* e dele com suas frases ou expressões) em função de diferentes dimensões. O objetivo da AMD é identificar padrões de coocorrência de características linguísticas e situacionais, pois, com um caráter essencialmente quantitativo e computacional, descreve seus objetos por meio de uma grande quantidade de características.

Finatto (2011) sumariza a proposta de AMD, a partir do trabalho de Shergue (2003), especialmente indicando a noção de dimensão, que, no dizer de Berber-Sardinha (2004), seria o *status* assumido pelo conjunto de características interpretado do ponto de vista de sua função comunicativa, conforme exemplificado no Quadro 1.

Quadro 1 – Características linguísticas e funções

CARACTERÍSTICA LINGUÍSTICA	FUNÇÕES
Conjunção coordenada	Conexão entre orações, fragmentar o texto (PACHECO, 1997, p. 95)
Conjunção subordinada	Conexão entre orações, complexidade estrutural (PACHECO, 1997, p. 95)
Pronomes pessoais de 1a. e 2a. pessoas	Interação e envolvimento (BIBER, 1988, p. 225)
Salvaguardas	Marcar incerteza do autor ou apresentar o conteúdo de forma mais generalizada, distanciamento (CHAFE; DANIELEWICZ, 1986 em BIBER, 1988, p. 106, 240; SALAGER-MEYER, 1994, p. 154)
Passiva	Distanciamento e abstração (BIBER, 1988, p. 228)
Look e See	Interação com o ouvinte em chamadas de atenção para o tópico que está sendo apresentado (SERAFINI; SHERGUE, 2002)
Densidade e nominalizações	Organizar o texto não em função de nós mesmos mas em função de ideias, razões, causas, distanciamentos (EGGINS, 1994, p. 59)

Fonte: Finatto (2011, p. 13).

As características linguísticas listadas na primeira coluna correspondem à dimensão; logo, as características resultam da análise e interpretação feita nos textos dos *corpora*, não sendo o ponto de partida apriorístico.

A compreensão de diferentes escalas de CT e sua mensuração (tanto em nível do *corpus* como de textos que o constituem) pode ser objetivamente favorecida pelo uso de aplicações eletrônicas, capazes de processar grandes volumes textuais a partir da execução de instruções por processador(es), abrangendo a entrada, a verificação, o armazenamento, a recuperação, a transformação e a produção de novas informações a partir dos dados fornecidos pelo pesquisador, conforme a Linguística de *Corpus*.

Um exemplo de aplicação eletrônica desse tipo é o sistema Coh-Metrix (*cohesion metrics*), desenvolvido por pesquisadores da Universidade de Memphis (GRAESSER; McNAMARA; LOUWERSE; CAI, 2004), nos Estados Unidos, para análise de textos em inglês. O programa calcula marcas de coesão e de coerência textuais a partir de itens lexicais, de categorias sintáticas, semânticas e referenciais, que são chamados de métricas, na terminologia da Linguística Computacional. Em sua versão para o português¹, apresenta sessenta métricas, agrupadas em seis conjuntos, que indicam a complexidade de um texto, quais sejam: identificação geral e informação de referência, índices de inteligibilidade (calculados com as fórmulas *Flesch Reading Ease* e *Flesch Kincaid Grade Level*, as quais consideram o tamanho de sentença, o número de palavras por sentença e o número de palavras diferentes por sentença), palavras gerais e informação do texto (contagens básicas, frequências, concretude, hiperônimos), índices sintáticos (constituintes, pronomes, tipos e *tokens*, conectivos, operadores lógicos e similaridade sintática de sentenças), índices referenciais e semânticos (anáfora, correferência e análise semântica latente) e dimensões do modelo de situações (causal, intencional, temporal e espacial).

Embora Finatto (2011) aponte para a base eletrônica em que a ferramenta está hospedada (vide nota), não conseguimos, até o momento presente, processar nenhum texto (*error Gateway Time-out*), a fim de atestar o modelo de processamento do Coh-Metrix para o português, mas, mesmo assim, verificamos a possibilidade de utilização e de simplificação de algumas dessas métricas na construção da nossa proposta funcionalista de avaliação escalar da CT.

Ainda que o Coh-Metrix apresente funcionalidades semelhantes ao *WordSmith Tools* (SCOTT, 2008), que utilizamos nesta pesquisa, eles diferem pelo fato de o primeiro se processar no ambiente Web, enquanto o segundo submete o processamento no disco rígido da máquina física do pesquisador.

¹ Disponível em: <www.nilc.icmc.usp.br/coh-metrix-port>.

Como podemos observar, os modelos até então existentes, em busca de uma precisão maior, elegem muitas variáveis para aferir a complexidade de um texto, tornando o procedimento não tão prático, ou até mesmo de difícil manuseio para, por exemplo, um professor que queira adequar um texto aos objetivos específicos de um exame mensal. No entanto, sabemos que nem só de categorias exclusivamente linguísticas constitui-se um texto; há outras variáveis de ordem semântico-discursiva e pragmática envolvidas na construção de sentidos em sua dimensão comunicativa que não são mensuráveis. Daí surge a motivação de pesquisarmos e propormos um instrumento de cunho funcionalista que possa atender às demandas mais básicas como também as de grande monta em termos de avaliação.

Base teórica da pesquisa

Nesta seção, apresentamos o suporte teórico de nossa pesquisa, cujo objetivo consiste em apresentar uma proposta de avaliação escalar de textos, baseada no Funcionalismo Linguístico norte-americano e na Linguística Sistêmico-Funcional, sobretudo, em uma de suas vertentes, a Gramática do *Design Visual*.

424

Linguística funcional

Considerando o nosso objetivo para o estudo da CT, optamos pelo Funcionalismo Linguístico, posto que seus estudiosos trabalham essencialmente com a linguagem verbal e não verbal, nas modalidades falada ou escrita, retiradas de contextos reais de interação humana, observando as relações entre os constituintes linguísticos e seus significados ou funções, ou entre a língua e seu meio.

Por considerarmos a língua como instrumento de interação social entre os seres humanos, adotamos, para esta pesquisa, duas vertentes funcionalistas, a Linguística funcional norte-americana, praticada, sobretudo, por Givón (1990, 1991, 1995), Hopper e Traugott (2003) e Bybee (2010), na análise dos aspectos funcionais da língua (morfo sintáticos, semânticos e pragmáticos), e a Linguística Sistêmico-Funcional (LSF), mais especificamente, na análise dos recursos multimodais sob a ótica da Gramática do Design Visual (GDV), de Kress e van Leeuwen (2006), que reelaboram o modelo sistêmico-funcional desenvolvido por Halliday (1985).

Assim, tendo como pressuposto o fato de que a língua desempenha funções que são externas ao sistema linguístico e o de que essas funções externas influenciam a organização interna do sistema linguístico, a Linguística funcional norte-americana busca explicar a forma da língua através do uso que se faz dela. Segundo Hopper e Traugott (2003) e Bybee (2010), a língua é usada para satisfazer necessidades interativas, e suas estruturas devem ser explicadas com base no uso real a que elas se prestam. Nesse sentido, a língua é tida como um “sistema adaptativo” (DU BOIS, 1985), uma “estrutura maleável” (BOLINGER, 1977) e “emergente” (HOPPER, 1987), posto que está sujeita às pressões do uso, ou seja, a codificação linguística é resultante do uso (da língua).

Cabe aos funcionalistas, ao lado da descrição da sintaxe (forma), investigar as circunstâncias discursivas que envolvem as estruturas da língua e os seus contextos específicos de uso (funções). Resulta, daí, a necessidade de investigar a sintaxe abrangendo a semântica e a pragmática.

Nessa perspectiva, destacamos, para este estudo, numa perspectiva funcionalista, os princípios da **iconicidade** e da **marcação**, além do estudo dos **planos discursivos** e do **estatuto informacional** da estrutura textual (verbal e visual).

Nos estudos linguísticos, define-se **iconicidade**, em sua versão forte, como a correlação natural entre forma (código linguístico/ expressão) e função (*designatum*/conteúdo). Os funcionalistas, por sua vez, defendem a ideia de que a estrutura da língua reflete, de algum modo, a estrutura da experiência humana (conceptualização humana do mundo). Em sua formulação branda, a relação de um-para-um entre expressão/forma e conteúdo se insere num contínuo entre os polos transparência e opacidade; sendo esta última interpretada como decorrente da trajetória de gramaticalização (cf. GIVÓN, 1990).

Segundo Givón (1990), três subprincípios são desdobrados da propriedade geral icônica, que se relacionam à quantidade de informação (volume), ao nível de proximidade entre os constituintes (integração) e à ordenação sequencial (linear) desses mesmos constituintes. Para esta investigação, interessa o da **quantidade de informação**, segundo o qual **quanto maior, mais imprevisível (mais nova) e saliente for uma informação a ser transmitida ao interlocutor, maior será a quantidade de forma a ser utilizada na codificação morfossintática dessa informação**.

Acreditando que as “estruturas sintáticas não devem ser muito diferentes, na forma e na organização, das estruturas semântico-cognitivas subjacentes”, Furtado da Cunha *et al.* (2003, p. 34) advogam

que, teoricamente, o princípio da iconicidade, em sua versão branda, permite detalhar as condições que governam o uso dos recursos de codificação morfossintática da língua, uma vez que há razões estritamente humanas de importância e complexidade que se refletem nos traços estruturais da língua.

Interessa também, para o nosso estudo da CT, o princípio meta-icônico da marcação, já que, segundo Givón (1991, p. 106), de forma idealizada (um construto teórico), “categorias que são cognitivamente marcadas (i.e., complexas) tendem a ser marcadas estruturalmente”.

Dessa forma, Givón (1990, 1995) apresenta três critérios para se distinguir categorias **marcadas** e categorias **não-marcadas**, em um contraste gramatical binário:

- a) **complexidade cognitiva:** a categoria marcada tende a ser cognitivamente mais complexa – em termos de esforço mental, demanda de atenção ou tempo de processamento – do que a não-marcada;
- b) **complexidade estrutural:** a estrutura marcada tende a ser mais complexa ou maior do que sua correspondente não-marcada;
- c) **distribuição de frequência:** a categoria marcada tende a ser menos frequente nos textos e, assim, cognitivamente mais perceptível do que a categoria correspondente não-marcada.

Esses três subprincípios tendem a representar um correlato mais geral entre padrões de marcação e motivações substantivas (comunicativas e cognitivas) que governam a natureza icônica das relações entre conteúdo e expressão.

Givón (1995) defende, ainda, que uma mesma estrutura pode ser marcada num contexto e não-marcada em outro, acrescentando que, por esse ângulo, a marcação é um fenômeno dependente do contexto, devendo, pois, ser explicada com base em fatores comunicativos, socioculturais, cognitivos ou biológicos.

Como sabemos, a **informatividade** se manifesta em todos os níveis da codificação linguística e diz respeito ao que falante e ouvinte compartilham, ou supõem compartilhar, em situação de interação. Do ponto de vista cognitivo, as pessoas se comunicam para informar o

ouvinte sobre alguma coisa, relativa ao mundo externo, ao seu próprio mundo interior, ou algum tipo de manipulação que queira exercer sobre esse ouvinte (PRINCE, 1981).

A hipótese geral que explica a questão do estatuto informacional dos constituintes e a organização destes na sentença/imagem é a de que a estrutura informativa do discurso, seja oral seja escrito ou mesmo imagético, deve apresentar elementos **dados** (conhecidos) e **novos**. A informação dada em geral corresponde àquele conhecimento partilhado pelos interlocutores. É ela que cria condições para que o falante introduza o novo, que, por seu turno, vai ampliar e/ou reformular os conhecimentos do ouvinte. Desse modo, há uma maior tendência de a informação dada anteceder a informação nova, resultando na opção não-marcada da ordem dado/novo.

Também nessa perspectiva, o modo como o falante organiza seu texto é determinado tanto por seus objetivos comunicativos, como pela percepção das necessidades do seu interlocutor, para que a comunicação se processe satisfatoriamente. Resulta daí o que é central e o que é periférico no texto (verbal e não verbal). Em termos de estrutura textual, ou de **planos discursivos**, essa divisão entre central e periférico é correspondente à distinção das noções de **figura e fundo**.

Segundo Furtado da Cunha (2003, p. 40), o fundamento cognitivo para o plano discursivo provém da psicologia gestáltica. Segundo essa concepção, identificam-se mais prontamente as entidades que se apresentam em primeiro plano, como figuras bem-recortadas e focalizadas, contrastando com o todo periférico, que passa a ser percebido em segundo plano ou em plano de fundo. A GDV tem esse entendimento, como veremos adiante.

Gramática do Design Visual

Sabemos que os textos, que estão presentes no cotidiano das pessoas, apresentam-se nos mais diversos gêneros e são caracterizados por possuírem, em sua constituição e apresentação, uma estruturação significativa. O próprio texto escrito já traz significados sob diferentes perspectivas: a disposição, o tamanho, o formato e as cores das fontes já são indícios da presença de diferentes modos de representação, ou seja, constroem sentidos e estabelecem relações através dos conteúdos ou discursos neles veiculados. Essas diferentes perspectivas de visualização do texto escrito e sua disposição interessam aos estudos empreendidos

pela Semântica Social e pela Multimodalidade Discursiva, que as chamam de **multimodalidade**, noção esta que pode ser ampliada pela presença de elementos como imagens, sons, animações, porque pode tornar mais explícita a relação entre os diferentes modos de representação, e é, portanto, um recurso bastante válido para a aferição de aprendizagem.

Segundo Dionísio (2005, p. 178), os textos demonstram em sua multimodalidade, de alguma forma, as relações entre a sociedade e o que ela representa; ou seja, tanto as ações sociais quanto os gêneros, que as tornam explícitas, são multimodais, considerando a sua capacidade de produzirem os sentidos pretendidos pelo enunciador/interlocutor, através de “no mínimo dois modos de representação: palavras e gestos, palavras e entonações, palavras e imagens, palavras e marcas tipográficas, palavras e sorrisos, palavras e animações, etc.” (DIONÍSIO, 2005, p. 178). E, com o advento das novas tecnologias da informação e comunicação, as formas de se produzir e de se processar conhecimentos, sentidos, significados ganharam uma nova configuração, em que a palavra e a imagem têm, cada vez mais, uma relação integrada, necessitando, portanto, de que seja dada a devida importância à multimodalidade dos textos escritos no contexto educacional ou mais especificamente na teoria linguística ou até mesmo no senso comum.

A multimodalidade tem seu aporte no Funcionalismo Linguístico, sobretudo na LSF. As conhecidas noções de funções da linguagem ideacional, interpessoal, textual, introduzidas por Halliday (1985), estabelecem a articulação entre recursos semióticos e as formas linguísticas. Kress e van Leeuwen (2006) reelaboram essas mesmas funções na GDV, sob a descrição, respectivamente, dos termos das metafunções: **representacional**, **interativa** e **composicional**.

A **metafunção representacional** comunica a relação entre os participantes internos de uma imagem. Nessa metafunção, os participantes podem ser classificados como **interativos** e como **representados**. Os participantes interativos produzem ou visualizam imagens; já os representados são (re)tratados nas imagens. Segundo Kress e van Leeuwen (2006), na composição das imagens, podem ser expressas as relações vetoriais, ou seja, a conexão entre os participantes de uma determinada representação imagética através de **processos narrativos** ou através de **processos conceituais**. Os narrativos apresentam participantes ligados por vetores que indicam ações, que, por sua vez, podem se dar através de dois processos: **processos de ação**, que implicam um Ator praticando uma ação sobre uma Meta; e

processos de reação, que implicam um Reagente que pratica a ação de olhar para um Fenômeno. Já os processos conceituais ocorrem quando não há vetores e os participantes são apresentados de maneira estática e atemporal. Esse tipo de processo classifica (tido como classificacional), estrutura (analítico) ou confere significado (simbólico).

A **metafunção interativa** estabelece entre o leitor/observador e a imagem observada uma relação que apresenta os seguintes aspectos: **contato**, **distância social**, **perspectiva** e **modalidade**. **Contato** é a representação imagética do participante que “olha” (ou não) nos olhos do leitor. A **distância social** corresponde ao nível de interação estabelecido entre a imagem e o leitor, podendo, assim, estabelecer relações de maior ou menor proximidade entre eles. A **perspectiva** denota o ângulo em que são representados os participantes, e a **modalidade** diz respeito à verossimilhança entre o real e o imagético.

A **metafunção composicional** diz respeito à composição do conjunto todo coeso, a partir de três princípios de composição inter-relacionados, que são o **valor de informação**, **a saliência** e **a estruturação**. O **valor de informação** é verificado por meio do local dos elementos (participantes e sintagmas que relatam uns aos outros e ao Espectador) que tem valores informativos específicos anexados às várias zonas da imagem: a) os elementos dispostos ao lado direito correspondem ao **Novo** (o desconhecido), ao passo que os dispostos à esquerda remetem ao **Dado** (aquilo supostamente conhecido pelo leitor); b) os elementos dispostos na parte superior da imagem têm valor de **Ideal**, enquanto aqueles dispostos na parte inferior, correspondem ao **Real**; c) os elementos dispostos no centro e nas margens denotam a **informação central** (principal) e as **informações subordinadas** (ou secundárias), respectivamente. O princípio de **saliência** estabelece uma hierarquia de importância entre os elementos, cuja função é atrair a atenção do espectador em diferentes graus, quais sejam: primeiro plano ou plano de fundo; tamanho; contrastes de tons e cores; diferença de nitidez, entre outros. Segundo o princípio da **estruturação**, a presença ou ausência de planos de estrutura desconecta ou conecta elementos da imagem, significando que eles pertencem ou não ao mesmo sentido. Isso pode ser realizado ou não por elementos que criam linhas divisórias, ou por linhas de estruturação reais.

Embora as metafunções não sejam excludentes, para esta pesquisa, interessa a metafunção **composicional**, posto que objetivamos analisar os recursos multimodais relativos à composição das provas do ENEM.

Procedimentos metodológicos

Para a elaboração da proposta de análise funcionalista do nível de CT, empregamos um *corpus* de controle como parâmetro para possíveis textos com maior nível de formalidade, portanto mais complexos. Utilizamos, então, como material de referência, o *Corpus Lácio-Ref* (ALUÍSIO *et al.*, 2003), que é composto por 4.046 textos científicos, em português brasileiro, representativos de diferentes campos do saber: ciências agrárias, biológicas, exatas e da terra, sociais aplicadas, da saúde, humanas, entre outros. A nossa escolha pelo uso do Lácio-Ref se deu pelo fato de o *corpus* ser composto de textos mais formais e variados quanto aos temas e gêneros.

Embora saibamos que a textualidade guarda muitos fatores, por exemplo, de ordem cognitiva e comunicativa não mensuráveis, ou seja, que se encontram subjacentes à superfície textual, limitamo-nos, para simplificar e dar maior praticidade à avaliação da CT, numa perspectiva funcional, a analisar indicadores linguísticos e visuais que podem ser facilmente quantificados por aplicação eletrônica.

Para a definição do inventário dos indicadores funcionalistas para a elaboração do instrumento de avaliação do nível de CT, partimos dos textos motivadores das Propostas de Redação da Prova de Linguagens, Códigos e suas Tecnologias, aplicadas no período de 2009-2015. As propostas exploram 29 textos motivadores com saliência da linguagem verbal, ou da linguagem não verbal (imagem), ou das duas modalidades integradas em um todo.

Como já são estipuladas, pela GDV, as categorias de análise da metafunção composicional dos textos visuais ou integrados, tivemos de calcular as frequências dos elementos linguísticos para a análise de textos predominantemente verbais. Primeiramente, transformamos os textos armazenados em extensão *.pdf em extensão *.txt, para a aplicação do programa computacional. Empregamos uma das ferramentas da Linguística de *Corpus* (LC), o suíte denominado *WordSmith Tools* (WST), desenvolvido por Scott (2008), da Universidade de Liverpool. É um programa concordanceador de análise de *corpus* usado nas Ciências Sociais. Por se tratar de um *software* muito rápido, acaba acelerando o processo de levantamento dos termos da linguagem especializada com mais eficiência, velocidade e precisão do que uma busca manual, que fatalmente levaria mais tempo. Para Berber-Sardinha (2004, p. 112), “[...] é fácil perceber que se trata de um programa que reúne as

principais ferramentas que o linguista de *corpus* precisa para identificar e comparar frequências e listar palavras no seu contexto original” (BERBER-SARDINHA, 2004, p. 112).

O WST é composto das seguintes ferramentas: *Wordlist*, *Keyword* e *Concord*. Vejamos:

- a) **Wordlist** – é a ferramenta básica do WST que fornece ao pesquisador uma lista de palavras do texto. Essa lista pode aparecer em ordem alfabética, ordem de frequência e dados gerais sobre o *corpus*. Por meio dessa ferramenta, é possível criar listas de palavras dos textos que compõem o *corpus*. O uso da ferramenta possibilita conhecer o número de parágrafos dos textos carregados no aplicativo, a quantidade de orações, o número de palavras por oração e razão *type/token*.
- b) **Keyword** – esta ferramenta faz a comparação entre a lista de palavras consideradas mais importantes geradas dos textos selecionados e um *corpus* de referência. Essa comparação é importante, visto que, através dela, pode-se ter acesso aos termos típicos do *corpus* de análise.
- c) **Concord** – é a ferramenta que produz concordâncias. Concordâncias, segundo Berber-Sardinha (2009, p. 87), “são listagens das ocorrências de um item específico (chamado termo de busca ou nóculo, que pode ser formado por uma ou mais palavras) acompanhado do texto ao seu redor (o co-texto)” (BERBER-SARDINHA, 2009, p. 87). Através da *Concord*, é possível analisar em que contexto um termo está situado e também a busca de termos complexos que se constituem de sequências fixas de palavras (*clusters*). Ela apresenta todas as ocorrências a partir do parâmetro proposto para a busca sob a forma de lista. Apresenta também as ocorrências com a expressão da busca e a construção imediatamente anterior em destaque. Na barra de *status* do aplicativo, é possível ver o número de entradas (ocorrências localizadas), além de informações estatísticas que apontam, desde o percentual de frequência de realizações com a configuração léxico-gramatical apresentada até a disposição no texto como um todo. Além disso, é possível saber de qual arquivo

foi classificada a ocorrência, sendo possível, portanto, quantificar em quais textos as ocorrências se mostraram mais recorrentes. Portanto, foi-nos uma ferramenta muito útil, entre outras, para a confirmação da exatidão do elemento pesquisado através do contexto de uso.

Então, com base no exposto na seção sobre o funcionalismo linguístico, definimos indicadores de CT que podem ser facilmente quantificados por aplicação eletrônica. A partir das propostas de redação, fazendo a correlação com o *corpus* de referência, buscamos, com base na frequência, indicadores segundo três tipos de complexidade: estrutural, cognitiva e visual.

Também com base na correlação dos dois *corpora*, definimos, para cada tipo de complexidade, quatro graus, relativos aos intervalos de frequência, ora relativa ora absoluta, ou mesmo da ausência dos indicadores, partindo do maior nível, na ordem do (+) complexo > (-) complexo, ou seja, do que denominamos de **Complexo**, seguido, em ordem decrescente, dos níveis **Difícil**, **Mediano** e **Fácil**, dispostos nas escalas propostas na seção seguinte.

Escalas do nível de complexidade textual sob a perspectiva funcionalista

Apresentamos, nesta seção, os resultados da nossa proposta para se calcular o nível de CT, ao tempo em que mostramos, passo a passo, como o usuário deve proceder para classificar os intervalos dos indicadores nas escalas de CT.

Salientamos que os cálculos dos indicadores dos níveis de complexidade estrutural e cognitiva são realizados, aplicando-se as ferramentas *WordList* e *Concord* do *WST* (SCOTT, 2012) para contagem de todas as ocorrências dos itens léxico-gramaticais em seu contexto de uso. Já para o levantamento dos indicadores da complexidade visual do texto, verifica-se a presença ou ausência deles nos textos.

Complexidade Estrutural

Seguindo a tradição dos estudos sobre complexidade e sobre inteligibilidade de textos, utilizamos, como critério, a extensão de frase conforme o número de palavras. Justificamos essa tomada de decisão, por meio de uma abordagem funcionalista, pelo subprincípio meta-icônico da *quantidade*, de Givón (1990). Com base nisso, importa-nos

verificar a extensão do período e a escolha léxico-gramatical de um item (*type*) e de suas repetições (*tokens*), por meio da razão *Type/Tokens*, cujo resultado revela que, quanto mais alto é quociente, mais complexo é o texto. O mesmo acontece com a extensão do período que, quanto mais volume codificado, mais complexo é o texto.

Então, para verificarmos a complexidade estrutural, devemos nos valer dos seguintes indicadores:

a) **Razão *type/token***

Para classificação dos textos nesse indicador, utilizamos a aplicação *WordList* do suíte *WST* (SCOTT, 2012). O cálculo da razão *type/token* é feito pelo próprio programa, que divide o número total de palavras (*tokens*) pela ocorrência única de cada item. Na tela da lista de palavras, esse valor pode ser visto na sétima coluna da guia “*statistics*”, que fica na barra de *status*. O cálculo pode ser feito da seguinte forma: $RTT = [TYP * 100 \div TOK]$, em que:

RTT = Razão *Type/Token*

TYP = *Types* (palavras distintas)

TOK = *Token* (todas as palavras)

Definimos, para a RTT, escala com os intervalos: (i) Complexo, densidade textual maior que 86%; (ii) Difícil, no intervalo entre 71% a 85%; (iii) Mediano, no intervalo de 51% a 70%; (iv) Fácil, quando a densidade textual for menor que 50%.

b) **Extensão do período**

A classificação de extensão do período, ou seja, o número médio de palavras no período, é também feita pela utilização da aplicação *WordList*. O programa calcula o número de períodos do texto e faz a média de palavras de cada um deles. Na tela da lista de palavras, esse valor pode ser visto na décima quarta coluna da guia “*statistics*” – onde se lê, “*mean (in words)*”, que fica na barra de *status*. A fórmula do cálculo seria: $MPP = [TOK \div NTS]$, em que:

TOK = *Token* (todas as palavras)

NTS = Número Total de Sentenças

A escala que definimos para a extensão do período identifica os períodos com maior e menor quantidade de palavras, a fim de obter o intervalo possível de ser aplicado na escala padrão, cujo intervalo corresponde: (i) Complexo, para períodos com mais de 50 palavras; (ii) Difícil, no caso de períodos com extensão entre 20 e 49 palavras; (iii) Mediano, para períodos entre 9 e 19 palavras; e, (iv) Fácil, quando a extensão do período for igual ou menor que 8 palavras.

Complexidade Cognitiva

Elegemos, para avaliar a complexidade cognitiva, os fatores de textualidade conhecidos na literatura como **coesão** e **coerência**, que, apesar de constituírem fenômenos diferentes, posto que a coesão atua na superfície textual, portanto observável e quantificável; e a coerência configura, na mente dos interlocutores, o sentido por meio de elementos subjacentes à superfície textual, imbricam-se na ocasião do processamento do texto. Nessa relação, escolhemos aferir os recursos de **junção** (tradicionalmente, conjunções), responsáveis pela articulação de frases e sequenciação textual; a **referenciação**, que, por meio de recursos léxico-gramaticais (quantificáveis, portanto) realiza o encadeamento coesivo do texto, auxiliando o interlocutor na construção do sentido, aferida pela presença de pronomes demonstrativos e pessoais; e a **pontuação**, que funciona como pista para orientar o interlocutor, por meio do número de pausas, o processamento e compreensibilidade do período. Já a **informatividade** se manifesta em todos os níveis da codificação linguística e diz respeito às informações que o falante/ouvinte compartilham, ou supõem compartilhar, em situação de interação. Como vimos, a estrutura informacional de um texto exige a presença de elementos conhecidos (dados) e de elementos novos. Para aferirmos os referentes **Dados** (velhos), elegemos os sintagmas iniciados por artigo definido; e para aferirmos os referentes **Novos**, isto é, introduzidos pela primeira vez no discurso, identificamos os usos codificados pelo sintagma iniciado por artigo indefinido².

Então, vejamos como verificar a complexidade cognitiva por meio dos quatro indicadores:

² A escolha pelos artigos definidos e indefinidos se deu por serem mais fáceis de mensuração, dentro das características possíveis de informações novas e velhas. Apesar da inexatidão do parâmetro, o uso do aplicativo *Concord* (SCOOT, 2012) permite ao analista visualizar e identificar a ocorrência desses indicadores.

a) Junção

Para classificação dos textos nesse indicador, utilizamos a aplicação *Concord*. O cálculo da frequência de junção é feito da seguinte forma: $MEJ = [TMJ * 100 \div TOK]$, em que:

MEJ = Marcadores Explícitos de Junção

TMJ = Total de Marcadores de Junção

TOK = Token (todas as palavras)

A identificação da quantidade de marcadores de junção pode ser obtida do *corpus* em análise, a partir da entrada de marcadores explícitos de junção na caixa de pesquisa do *Concord*. No Quadro 2, dispomos do inventário de MEJ, separados em coordenativos e subordinativos, segundo manuais de gramática e pesquisas na área (MENDES, 2016), somente para facilitar a busca na aplicação.

Quadro 2 – Marcadores explícitos de junção

Juntores de coordenação	Juntores de subordinação
além de/e/não só * mas também/nem/mas/ porém/contudo/todavia/entretanto/no entanto/ora * ora/ou/ou * ou/quer * quer/seja * seja/..., pois, .../então/logo/por consequinte/portanto/, porque/, pois.	já que/visto que/assim como/como/maior * do que/mais * do que/melhor * do que/menor * do que/ menos * do que/pior * do que/tanto * que/tão * que/ainda que/apesar de/ embora/mesmo que/por mais que/se bem que/a não ser que/ caso/, contanto que/ quanto mais * mais/quanto mais * menos desde que/se/conforme/ consoante/ segundo/tanto que/tamanho que/para que/a fim de/sem que/à medida que/na medida em que/à proporção que/ao passo que/antes que/assim que/depois que/ enquanto/logo que/mal/quando/sempre que/onde/por onde.

Fonte: Figueiredo-Gomes e Mendes (2017).

Para medirmos a CT, por meio do indicador junção, utilizamos a seguinte escala: (i) Complexo, quando a ocorrência de juntores explícitos for menor que 2%; (ii) Difícil, para os textos em que a junção esteja no intervalo entre 2,1% e 4%; (iii) Mediano, nos casos em que os juntores ocorrem no intervalo entre 4,1% e 6,0%; (iv) Fácil, quando a presença de juntores for maior que 6,1%.

b) Referenciação

Para classificação dos textos segundo esse indicador, utilizamos a aplicação *Concord*. A fórmula do cálculo da frequência de referenciação é: $MRE = [TMR * 100 \div TOK]$, em que:

MRE = Marcadores de Referenciação

TMR = Total de Marcadores de Referenciação

TOK = *Token* (todas as palavras)

A identificação da quantidade de marcadores de referenciação pode ser obtida do *corpus* em análise, a partir da entrada de marcadores explícitos de junção na caixa de pesquisa do *Concord*. No Quadro 3, dispomos do inventário de MRE, baseado em manuais de gramática e pesquisas na área.

Quadro 3 – Marcadores de Referenciação

Referência pessoal	Referência demonstrativa
ele/eles/ela/elas/ lhe/lhes/-o/-os/-a/-as/-lo/-los/-la/-las/-no/-nos/-na/-nas	seu/seus/sua/suas/dele/deles/dela/delas/nele/neles/nela/nelas/dele/deles/dela/delas/este/estes/esta/estas/isto/esse/esses/essa/essas/isso/aquele/aqueles/aquela/aquelas/aquilo/aqui/aí/ali/lá/o/os/a/as/tal/tais/mesmo/mesmos/mesma/mesmas/ambos/ambas/um e outro/uma e outra/o primeiro * o segundo/tudo.

Fonte: Figueiredo-Gomes e Mendes (2017).

Para a aferição do indicador referenciação, utilizamos a seguinte escala: (i) Complexo, quando a ocorrência de expressões referenciais for menor que 1%; (ii) Difícil, para os textos em que a referência esteja no intervalo entre 1,1% e 1,4%; (iii) Mediano, nos casos em que os referenciadores ocorrem no intervalo entre 1,5% e 1,9%; (iv) Fácil, quando a presença de expressões referenciais for maior que 2%.

c) Pontuação

Para classificação dos textos nesse indicador, utilizamos a aplicação *Concord*. A fórmula do cálculo da quantidade de pontuação é: $IMP = [NTP * MPP \div NTV]$, em que:

IMP = Intervalo Médio de Pausas
NTP = Número Total de Períodos
MPP = Média de Palavras por Período
NTV = Número Total de Vírgulas

A obtenção da quantidade total de pausas no *corpus* analisado se processa a partir do parâmetro *,* (asterisco vírgula asterisco) na caixa de busca do *Concord*.

A avaliação da CT por meio do indicador pontuação é feita segundo a escala: (i) Complexo, quando a quantidade média de pausas por períodos/média de palavras por período for igual ou menor que 3; (ii) Difícil, nos períodos com pausas médias entre 4 e 6; (iii) Mediano, quando o quantitativo for de 7 a 9; e, (iv) Fácil, quando a quantidade média de pausas for maior ou igual 10.

d) **Razão Dado/Novo**

Para classificação dos textos nesse indicador, utilizamos a aplicação *Concord*. A fórmula do cálculo da razão Dado/Novo é: $RDN = [NRV * 100 \div NRN]$, em que:

RDN = Razão Dado/Novo
NRN = Número de Referentes Novos
NRV = Número de Referentes Velhos

A obtenção da quantidade total de referentes novos no *corpus* analisado se processa a partir da entrada (um/uma) na caixa de busca do *Concord*; para os referentes velhos, usa-se a entrada (o/os/a/as).

Os intervalos da CT, segundo a razão Dado/Novo, são os seguintes: (i) Complexo, quando a indicação de novidade for maior que 22%; (ii) Difícil, no intervalo entre 17% e 21%; (iii) Mediano, no intervalo de 11% a 16%; e, (iv) Fácil, quando a indicação de novidade for menor que 11%.

Complexidade Visual

Para medir a complexidade visual, analisamos os recursos visuais dos textos segundo a metafunção composicional, da GDV (KRESS; Van LEEUWEN, 2006), que diz respeito à composição do conjunto todo coeso, a partir dos três princípios de composição inter-relacionados, que

descrevemos na fundamentação teórica: a) em relação à **estruturação**, o usuário deve verificar a presença ou ausência de linhas divisórias, ou por linhas de estruturação reais; b) o **valor de informação** é verificado por meio do local dos elementos que correspondem ao **Novo** e ao **Dado**; ao **Ideal**, ao **Real**; à **informação central** e às **informações subordinadas**; c) no tocante à **saliência**, deve-se observar: plano de *fundo* ou primeiro plano (*figura*); tamanho; contrastes de tons e cores; diferença de nitidez, entre outros.

Para facilitar a pesquisa do usuário, propusemos uma ficha analítica da CT, presente na Figura 1.

Figura 1 – Ficha analítica da escala de complexidade textual

1 Texto nº		2 Tipo () Verbal () Integrado			
3 COMPLEXIDADE ESTRUTURAL	3.1 Razão Type/token	Complexo (> 86%)	Difícil (71% a 85%)	Mediano (51% a 70%)	Fácil (< 50%)
	3.2 Extensão do período	Complexo (> 50)	Difícil (20 a 49)	Mediano (9 a 19)	Fácil (≤ 8)
4 COMPLEXIDADE COGNITIVA	4.1 Junção	Complexo (< 2%)	Difícil (2.1% a 4,0%)	Mediano (4.1% a 6,0%)	Fácil (> 6.1%)
	4.2 Referenciação	Complexo (< 1%)	Difícil (1,1% a 1,4%)	Mediano (1,5% a 1,9%)	Fácil (> 2,0%)
	4.3 Pontuação	Complexo (≤ 3)	Difícil (4 a 6)	Mediano (7 a 9)	Fácil (≥ 10)
	4.4 Razão dado/novo	Complexo (> 22%)	Difícil (17% a 21%)	Mediano (11% a 16%)	Fácil (< 11%)
5 COMPLEXIDADE VISUAL	5.1 Estruturação	() Ausência de linhas divisórias () Ausência de título () Ausência de margem paragrafíca/moldura () Divisão em colunas			
		Complexo (>3)	Difícil (2)	Mediano (1)	Fácil (0)
	5.2 Informatividade	() Posição Direita () Posição Inferior () Ausência de figura de participantes () Ausência de cenário			
		Complexo (>3)	Difícil (2)	Mediano (1)	Fácil (0)
	5.3 Saliência	() Ausência de contraste de cor/ton/nitidez () Indiferença de tamanho () Plano de fundo () Ausência de grifo da fonte			
		Complexo (>3)	Difícil (2)	Mediano (1)	Fácil (0)

Fonte: Figueiredo-Gomes e Mendes (2017).

Para a utilização da ficha analítica da escala de CT, o avaliador deve fazer o recenseamento dos indicadores relativos a cada nível de complexidade estrutural e cognitiva no(s) texto(s), aplicando-se as ferramentas *Wordlist* e *Concord* do *WST*. Já para o levantamento dos indicadores da complexidade visual do texto, assinala-se a presença ou ausência deles nos parênteses e a soma resultante indicará o nível do texto. O texto será classificado de acordo com a incidência do nível escalar pontual ou intervalar.

Conclusão

A correlação entre os *corpora* forneceu dados, em uma perspectiva funcionalista, para a elaboração de uma escala de nível de CT, resultante dos graus de cada tipo de complexidade: *estrutural*, *cognitiva* e *visual*, na ordem escalar do (+) complexo > (-) complexo, ou seja, *Complexo*, *Difícil*, *Mediano* e *Fácil*. Desse modo, esse instrumento permite avaliar textos de língua portuguesa, diferentemente das medidas até então existentes.

Como podemos ver na ficha analítica, o inventário dos indicadores linguísticos e visuais, resultantes desta pesquisa, podem ser facilmente quantificáveis por estudiosos, instituições de avaliação e professores de línguas via aplicação eletrônica tanto em um dado texto ou em uma perspectiva macro – avaliação de textos para exames de grande monta.

A escala resultante desse inventário pode proporcionar uma maior segurança ao professor de língua portuguesa ou principalmente ao professor elaborador e ao revisor de itens das provas do ensino médio, na escolha de textos *complexos*, *difíceis*, *medianos* e *fáceis*, ou mesmos intervalares, adequados à clientela estudantil, segundo critérios textuais que se queira explorar em termos da proficiência da leitura e da escrita.

Os princípios de iconicidade e de marcação, a informatividade e a saliência, bem como a GDV, de orientação teórica funcionalista, associados à LC, sobretudo com o uso do software *WST*, foram basilares para a elaboração da escala do nível de CT, naquilo que pode ser mensurável em termos de língua.

Considerando a consistência empírica e teórica das escalas de CT, podemos julgar que temos um instrumento de medida confiável e inovador de avaliação do texto em sua completude, tanto verbal quanto visual, por meio de recursos tecnológicos computacionais.

Referências

ALUÍSIO, S.M.; PINHEIRO, G.; FINGER, M.; NUNES, M.G.V. e TAGNIN, S.E.O. 2003. The Lácio-Web Project: overview and issues in Brazilian Portuguese corpus creation. In: CORPUS LINGUISTICS 2003, Lancaster. **Proceedings of Corpus Linguistics**. Lancaster, UCREL: Lancaster University, 2003, p.14-21.

BERBER-SARDINHA, T. **Linguística de Corpus**. São Paulo: Manole, 2004.

_____. **Pesquisa em Linguística de Corpus com WordSmith Tools**. Campinas: Mercado de Letras, 2009.

BIBER, D. **Variation across Speech and Writing**. Cambridge: Cambridge University Press, 1988.

- BOLLINGER, D. **The form of language**. London: Logman, 1977.
- BYBEE, J. **Language, usage and cognition**. Cambridge: Cambridge University Press, 2010.
- CAPRA; F.; LUISI, P. L. **A visão sistêmica da vida**: uma concepção unificada e suas implicações filosóficas, políticas, sociais e econômicas. São Paulo: Cultrix, 2014.
- CIAPUSCIO, G. La terminología desde el punto de vista textual: selección, tratamiento y variación. Porto Alegre, **Organon**, v. 12, n. 26, p. 43-65, 1998.
- _____. **Textos especializados y terminología**. Barcelona: IULA, 2003.
- DANIELSON, K. E. Readability formulas; a necessary evil? **Reading Horizons**, v. 27, n. 3, p. 178-188, apr. 1987.
- DIONÍSIO, A. P. Gêneros multimodais e multiletramento. In: KARWOSKI, A. M. et al. (orgs.) **Gêneros Textuais**: reflexões e ensino. Palmas e União da Vitória: Kay Gangue, 2005.
- DU BOIS, J. W. Competing Motivations. In: HAIMAN, J. (ed.). **Iconicity in syntax**. Amsterdam: John Benjamins Publishing Company, 1985. p. 343-65.
- FIGUEIREDO-GOMES, J. B. **Estudo da legibilidade de livros didáticos**. 1989. 50 f. Monografia (Aperfeiçoamento no Ensino de Português) – Núcleo de Estudos de Língua Materna, Universidade Federal do Ceará, Fortaleza, 1989.
- _____; MENDES, W. V. **Padrões de complexidade textual**: uma proposta funcionalista para avaliação de desempenho de alunos do ensino médio. 2º Relatório apresentado ao CESPE/CEBRASPE. Brasília: CESPE/CEBRASPE, 2017. Documento eletrônico em formato *.pdf. 2,1 MB.
- FINATTO, M. J. B. Complexidade textual em artigos científicos: contribuições para o estudo do texto científico em português. **Organon** (UFRGS), Porto Alegre, v. 50, p. 30-45, 2011.
- FURTADO DA CUNHA, M. A.; COSTA, M. A.; CEZARIO, M. M. Pressupostos teóricos fundamentais. In: FURTADO DA CUNHA, M. A.; OLIVEIRA, M. A.; MARTELOTTA, M. E. **Linguística funcional: teoria e prática**. Rio de Janeiro: DP&A, 2003. p. 29-55.
- GIVÓN, T. **Syntax: A functional-typological introduction**, Vol. II. Amsterdam: Benjamins, 1990.
- _____. Serial verbs and mental reality of event: grammatical vs. cognitive packaging. In: TRAUGOTT, E.; HEINE, B. (eds.) **Approaches to grammaticalization**, v.1. Amsterdam: Benjamins, 1991. p. 81-127.
- _____. **Functionalism and grammar**. Amsterdam: Benjamins, 1995.
- HALLIDAY, M.A.K. **An Introduction to Functional Grammar**. Baltimore: Edward Arnold, 1985.
- _____; MATTHIESSEN, C. M.I.M. **An Introduction to functional grammar**. London: Edward Arnold, 2004.

HOPPER, P. J. Emergent grammar. **BLS** 13, p.139-157, 1987.

_____; TRAUGOTT, E. C. **Grammaticalization**. Cambridge: Cambridge University Press, 2003.

GRAESSER, A. C., MCNAMARA, D. S., LOUWERSE, M. M., CAI, Z. Coh-Metrix: Analysis of text on cohesion and language. **Behavioral Research Methods, Instruments, and Computers**, Arizona, 36, p. 193-202, 2004.

KLARE, G.R. **A manual for readable writing**. Glen Burnie, Maryland: REM Company (119A Roesler Road), 1975.

KRESS, G.; van LEEUWEN, T. **Reading images: the grammar of visual design**. 2. ed. London: Routledge, 2006.

MENDES, W.V. **Mecanismos de junção em textos acadêmicos: uma abordagem sistêmico-funcional**. Tese (Doutorado em Estudos da Linguagem) PPGEL/CCHLA, Universidade Federal do Rio Grande do Norte, Natal, 2016.

MOLINA, O. **Avaliação de inteligibilidade de livros didáticos de 1º e 2º graus por meio da técnica cloze**. 1979. Tese (Doutorado em Psicologia) – Instituto de Psicologia, Universidade de São Paulo, São Paulo, 1979.

PRINCE, H. F. The ZPG Letter: subjects, definiteness, and information-status. In: MANN, William C.; THOMPSON, Sandra A. (Eds.) **Discourse description: diverse linguistic analyses of a fund-raising text**. Amsterdam: John Benjamins, 1992. p. 295-325.

SCOTT, M. **WordSmith Tools**. Oxford: Oxford University Press, [2008]2012.

SHERGUE, O. **Dimensão de Variação no Discurso Médico-Acadêmico: o artigo de pesquisa e a apresentação de trabalhos científicos em congressos**. Dissertação (Mestrado em Linguística Aplicada). PEPLA, Pontifícia Universidade Católica de São Paulo, São Paulo, 2003.

TAYLOR, W. L. Cloze procedure: A new tool for measuring readability. **Journalism Quarterly**, 30, p. 415-433, set. 1953.