

Identificação da unidade central de textos dissertativo-argumentativos em contexto de avaliação

Identification of the central unit of argumentative texts in the context of examinations

Juliano Desiderato Antonio¹

jdantonio@uem.br

Universidade Estadual de Maringá

Kátia Roseane Cortez dos Santos¹

katiacortez_@hotmail.com

Universidade Estadual de Maringá

Daniella Paes da Silva¹

daniellapaes@live.com

Universidade Estadual de Maringá

Camila Cristiane Moreschi¹

camilacmoreschi@gmail.com

Universidade Estadual de Maringá

RESUMO - Este trabalho tem como objetivo investigar critérios para a identificação da unidade central de textos argumentativos em contexto de vestibular. Comparam-se os resultados de trabalho anterior sobre textos do gênero resposta argumentativa produzidos por candidatos ao vestibular com os resultados obtidos a partir da aplicação dos mesmos critérios a um corpus formado por textos dissertativo-argumentativos produzidos por candidatos do Enem. O aparato teórico-metodológico utilizado é a *Rhetorical Structure Theory*, teoria de caráter descritivo que visa estudar a organização textual por meio das relações implícitas que se estabelecem entre as partes do texto. Ao contrário do gênero resposta argumentativa, em que a unidade central ocorre em posição inicial na maioria dos textos, no texto dissertativo-argumentativo, a posição mais comum da unidade central é na porção que vai de 20% a 30% da extensão do texto. Essa diferença organizacional se deve ao fato de, no gênero resposta argumentativa, o produtor dever iniciar seu texto respondendo à pergunta que o motivou. No texto dissertativo-argumentativo, é necessária a criação de um fundo com informações que situem o destinatário do texto com relação à temática. Outro traço importante na busca de um padrão é a frequência de determinadas palavras utilizadas nas unidades centrais. Substantivos e adjetivos que compõem o comando da proposta de redação do texto dissertativo-argumentativo apresentam alta frequência, uma vez que os candidatos tendem a utilizar essas palavras como forma de garantir o atendimento do tema solicitado. O caráter mais subjetivo que a resposta

ABSTRACT - This study aims to investigate criteria for the identification of the central unit of argumentative texts in the context of university entrance examinations. The results of a previous study on argumentative answers produced by entrance examinations candidates are compared with the results obtained from the application of the same criteria to a corpus formed by argumentative essays produced by Brazilian National High School Exam (Exame Nacional do Ensino Médio - Enem) candidates. The theoretical and methodological apparatus used was the *Rhetorical Structure Theory*, a descriptive theory that studies textual organization through implicit relations established between the parts of the text. Unlike argumentative answers, in which the central unit appears at the beginning of most texts, the most common position of the central unit in argumentative essays is between 20% and 30% of the length of the text. This organizational difference is due to the fact that, in argumentative answers, the text producer must start the text answering the question, whereas in argumentative essays, creating a background with relevant information is necessary to situate the reader with respect to the theme of the text. Another important feature in the search for a pattern is the frequency of certain words used in the central units. Nouns and adjectives used in the question show high frequency, since candidates tend to use these words in order to ensure that they meet the theme requested. The more subjective character of the answer in the argumentative answer genre favors the use of evidential verbs

¹ Universidade Estadual de Maringá. Av. Colombo, 5790, Jd. Universitário, 87020-900, Maringá, PR, Brasil.

pode assumir no gênero resposta argumentativa favorece a emergência de verbos evidenciais que expressam atitude proposicional (“acreditar”, “crer” e “pensar”), atuando como matriz da sentença na qual a resposta é dada. Nos textos dissertativo-argumentativos, o caráter mais objetivo da expressão do ponto de vista não favorece o uso de verbos desse tipo.

Palavras-chave: unidade central, texto dissertativo-argumentativo, gênero resposta argumentativa.

expressing propositional attitude (“believe” and “think”), acting as the matrix of the sentence in which the answer is given. In argumentative essays, the point of view is more objective, which does not favor the use of such verbs.

Keywords: central unit, argumentative essays, argumentative answers.

Introdução

A descrição linguística que extrapola o nível de análise sentencial pode contribuir com o desenvolvimento de aplicações automáticas como a sumarização, a segmentação, a tradução e até mesmo a avaliação. De acordo com van Dijk (1980), os falantes são capazes de sumarizar o conteúdo de um texto por meio dos tópicos principais daquele texto. Nas palavras do linguista holandês (p. 41, tradução nossa), discursos são “organizados em torno de um ‘núcleo’ semântico que intuitivamente chamamos de tópico”. Para van Dijk (1980), os tópicos discursivos são propriedades do sentido global do texto, necessárias para que o texto seja globalmente coerente.

Segundo Iruskieta *et al.* (2015), alguns termos têm sido utilizados para nomear a síntese do sentido global do texto: declaração da tese (Burstein *et al.*, 2001, tradução nossa), proposição central (Pardo *et al.*, 2003, tradução nossa), subconstituente central (Egg e Redeker, 2010, tradução nossa), unidade central (Stede, 2008, tradução nossa). Como este trabalho está embasado na *Rhetorical Structure Theory* (doravante RST), mesmo modelo teórico-metodológico utilizado no trabalho de Stede (2008), optou-se por utilizar o termo “unidade central” (doravante UC).

Neste trabalho, objetiva-se investigar critérios para a identificação da UC de textos dissertativo-argumentativos. Serão apresentados os resultados obtidos a partir da aplicação dos mesmos critérios utilizados em pesquisa anterior com textos do gênero resposta argumentativa (Antonio, 2015), produzidos por candidatos ao vestibular de uma universidade pública, a um corpus formado por textos dissertativo-argumentativos elaborados por candidatos do Exame Nacional do Ensino Médio (doravante Enem) de 2013. Pelo fato de se utilizarem os mesmos critérios, serão retomados, para fins de comparação, os resultados dessa pesquisa anterior.

Estabelecer critérios para a identificação da UC é metodologicamente importante para o desenvolvimento de análises baseadas na RST. De acordo com (Iruskieta *et al.*, 2014), a detecção da UC é um passo decisivo para a anotação da estrutura retórica de um texto. Do ponto de vista acadêmico, a identificação da UC é essencial para o treinamento das bancas de professores que realizam a

avaliação de textos em concursos vestibulares e no Enem. Produzir um texto conscientemente a partir do desenvolvimento da UC também pode ser crucial para o futuro de alunos que disputam uma vaga no ensino superior, seja por meio de concurso vestibular, seja por meio do Enem.

Um exemplo de UC encontrada em um texto do corpus é apresentado a seguir:

A Lei Seca, desde sua implantação, gerou resultados positivos tanto nas estatísticas quanto no comportamento dos motoristas.

No exemplo, as características que permitem identificar a UC são os seguintes: (i) a UC retoma vocábulos que fazem parte do comando da produção textual (implantação, Lei Seca) como forma de garantir o atendimento à temática proposta; (ii) o vocábulo “resultado” é tomado como sinônimo de “efeito”, vocábulo também presente no comando; (iii) a UC encontra-se na porção que vai de 20% a 30% da extensão do texto.

Para que os objetivos sejam atingidos, este trabalho está dividido em mais quatro seções além desta introdução. Na fundamentação teórica, apresentam-se os principais pressupostos da teoria que embasa o trabalho, bem como os conceitos subjacentes ao texto dissertativo-argumentativo. O corpus e os parâmetros de análise são descritos na metodologia. Na análise, são exibidos os resultados da aplicação dos critérios ao corpus de textos dissertativo-argumentativos. Por fim, expõem-se as conclusões deste trabalho.

Fundamentação teórica

RST

De acordo com Matthiessen (2005), a RST surgiu das pesquisas do grupo liderado por William C. Mann, no início dos anos 1980, no *Information Sciences Institute*, da Universidade da Califórnia, que tinham como objetivo investigar a organização dos textos. A finalidade maior da pesquisa estava relacionada à geração automática de textos. Dois linguistas de renome, Christian Matthiessen e Sandra Thompson, se uniram ao grupo, que também contava com a consultoria de Michael Halliday, grande nome da Linguística Sistemico-Funcional. Ainda segundo Matthiessen (2005), esse grupo não imaginava que a teoria

que eles estavam criando despertaria interesse tanto na Linguística Computacional quanto na Linguística.

A RST tem como principal conjectura o fato de que as orações de um texto veiculam mais do que apenas conteúdo proposicional explícito. Da combinação entre as orações e as partes de um texto surgem proposições implícitas, as chamadas *proposições relacionais*, que recebem outros rótulos como “relações retóricas”, “relações discursivas”, “relações de coerência” (Taboada, 2009, p. 127).

Os autores-fundadores da RST destacam que as proposições relacionais são de sentido, e não de forma (Mann e Thompson, 1983). Do ponto de vista da expressão linguística, isso equivale a dizer que não há necessidade de explicar as relações por meio de conectores. Assim, para identificar as relações, parâmetros funcionais e semânticos devem ser utilizados pelo analista na tentativa de identificar a função de cada parte do texto de acordo com o possível efeito no destinatário interlocutor, calculado/planejado/desejado pelo produtor do texto.

É importante observar que uma análise baseada na RST é uma análise de plausibilidade. Essa característica da teoria em foco indica que, embora o analista possa ter conhecimento das condições de produção do texto, falta-lhe conhecimento do estatuto das informações na mente do falante. Por isso, não se deve afirmar que uma determinada relação é absolutamente a que o produtor do texto teve como intenção utilizar, mas sim a relação plausível naquele contexto (Mann e Thompson, 1988).

No texto-fundador da teoria (Mann e Thompson, 1988), os autores propuseram uma lista de vinte e quatro relações após a análise de centenas de textos utilizando a RST. Posteriormente, o *website* da teoria divulgou uma lista com 30 relações (Mann e Taboada, 2010). A lista de relações utilizada neste trabalho é a de Carlson e Marcu (2001), com 136 relações.

A organização táctica das relações pode ocorrer de duas maneiras:

- (i) núcleo-satélite: a porção textual que funciona como satélite traz informações para subsidiar o conteúdo do núcleo. A Figura 1 representa uma relação núcleo-satélite, em que o arco vai do satélite em direção ao núcleo.
- (ii) multinucleares: cada porção de texto tem o mesmo estatuto. Na Figura 2, que representa uma relação multinuclear, pode-se observar que cada porção textual funciona como um núcleo distinto.

Os diagramas que representam as análises da RST têm forma arbórea², como pode ser observado na Figura 3, que exemplifica a análise da macroestrutura de um texto do corpus deste trabalho. Para van Dijk (1992, p. 51), a macroestrutura é “a informação semântica que fornece unidade global ao discurso”.

Além do título do texto (*span*₁), o texto é dividido, em sua macroestrutura, em outros quatro *spans*³. O *span*₃ contém a UC, ou seja, a unidade que apresenta a ideia principal do texto (Uma das consequências imediatas dessa iniciativa do poder público é a diminuição dos perigos relacionados à locomoção viária, uma vez que o número de acidentes tende a ser sensivelmente reduzido nesse sentido). No diagrama, a UC compõe o nódulo central da árvore e não funciona como satélite de nenhum outro *span*. Como os outros *spans* são satélites em relação ao *span*₃, as setas apontam em direção a esse *span*. O *span*₁ estabelece relação de preparação com o núcleo. Essa relação tem como efeito tornar o leitor mais interessado ou mais preparado para ler o conteúdo do núcleo. O *span*₂ estabelece relação de fundo com o núcleo. O efeito dessa relação é fornecer informações que aumentem a capacidade de o leitor compreender o conteúdo do núcleo. O *span*₄ estabelece relação de concessão com o núcleo. Na relação de concessão, o produtor do texto reconhece uma potencial ou aparente incompatibilidade entre o núcleo e o satélite, e reconhecer essa compatibilidade aumenta a



Figura 1. Esquema de relação núcleo-satélite.

Figure 1. Nucleus-satellite scheme.

Fonte: Mann e Thompson (1988, p. 247).

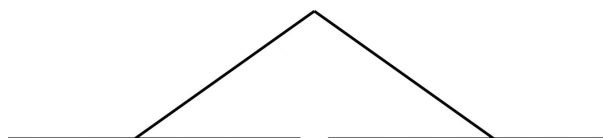


Figura 2. Esquema de relação multinuclear.

Figure 2. Multinuclear scheme.

Fonte: Mann e Thompson (1988, p. 247).

² Os diagramas são elaborados utilizando-se a ferramenta RSTTool (O'Donnell, 2000).

³ “*Spans* são porções de texto. Podem ser atômicas (uma oração ou sentença), ou compostas de outros *spans*” (Taboada e Gómez-González, 2013, p. 68, tradução nossa).

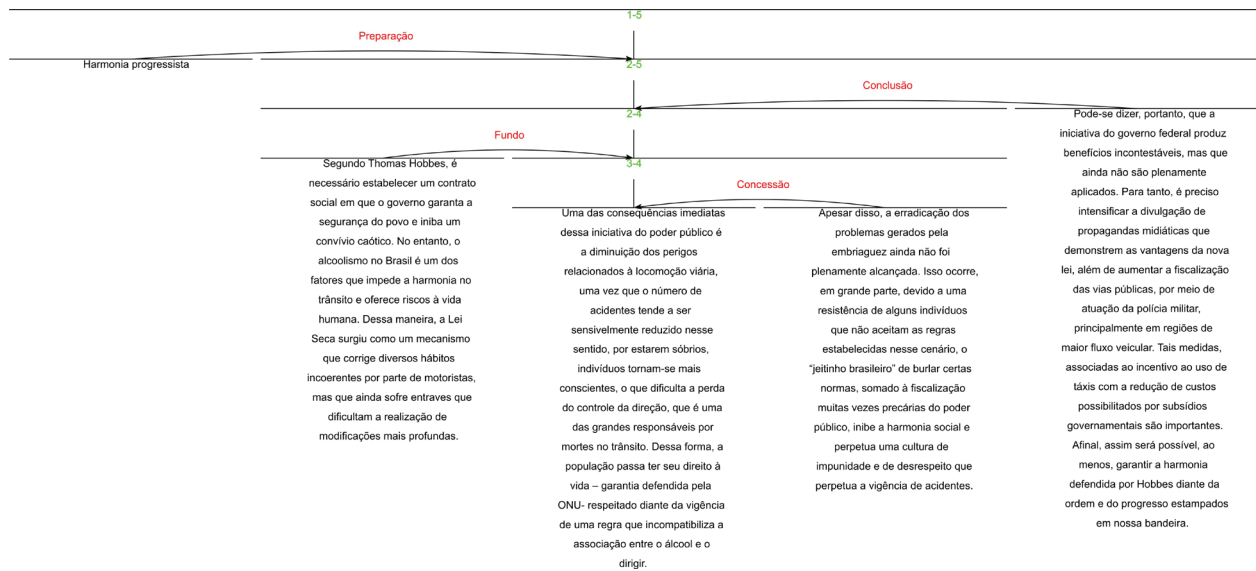


Figura 3. Macroestrutura de um texto do corpus.

Figure 3. Macrostructure of a text in the corpus.

Fonte: os autores.

atitude positiva³ do leitor em relação ao núcleo. Por fim, o *span*₅ estabelece relação de conclusão com o núcleo.

O texto dissertativo-argumentativo

A nomenclatura “dissertativo-argumentativo” é oriunda da junção de características da dissertação e da argumentação. De acordo com Garcia (1996)⁴, a dissertação tem como características a exposição, a explicação e a interpretação de ideias, ao passo que a argumentação se define pelo convencimento, pela persuasão na tentativa de se influenciar o destinatário do texto. Ainda segundo o autor, enquanto na dissertação o produtor do texto expressa sua opinião sobre determinado assunto, na argumentação, o produtor do texto objetiva “formar a opinião” (p. 370) do destinatário do texto, procurando demonstrar que está com a razão, que está de posse da verdade.

Segundo Garcia (1996), em geral, os manuais didáticos não distinguem dissertação de argumentação. Em geral, esses materiais consideram a argumentação um elemento constituinte da dissertação. Esse tratamento genérico dado pelos manuais didáticos a esse tipo de texto é o mesmo que se observa em avaliações de ingresso no ensino superior e também no Enem.

Ao tratar da estrutura de um texto dissertativo, Garcia (1996) apresenta a seguinte divisão: introdução (parte

do texto em que o produtor apresenta a ideia principal ou ideia-núcleo, tomando como base um fato circunstancial), desenvolvimento e conclusão.

Quando volta seu olhar para a argumentação, Garcia (1996) aponta os dois elementos principais para sua construção:

- (i) a consistência do raciocínio: para o autor, a legítima argumentação deve ser construtiva em sua finalidade, não deve ser vazia de sentido nem ausente de senso comum;
- (ii) a evidência das provas: conforme o autor, a evidência é “a certeza manifesta, a certeza a que se chega pelo raciocínio (*evidência de razão*) ou pela apresentação dos fatos (*evidência de fatos*)” (p. 371). Garcia apresenta cinco tipos de evidências, consideradas por ele as mais comuns: fatos, exemplos, ilustrações, dados estatísticos e testemunhos.

Apresentar fatos, segundo Garcia (1996), é uma estratégia argumentativa muito importante devido ao poder de convencimento e de persuasão dessa técnica. O autor exemplifica com o caso do sistema de saúde público brasileiro. Para ele, o fato de os contribuintes terem que recorrer a hospitais e planos particulares porque a saúde pública não tem condições de atendê-los satisfatoriamente é uma prova da deficiência desse sistema. O autor alerta,

⁴ Atitude positiva é um termo utilizado na RST para designar os diferentes objetivos do produtor do texto como, por exemplo, persuadir, criar crença, aprovação ou interesse. O produtor do texto age sobre seu interlocutor para levá-lo a concordar, acreditar ou agir de acordo com o conteúdo do núcleo (Mann e Thompson, 1988).

⁵ A primeira edição é de 1967.

no entanto, para o fato de que mudanças ao longo do tempo podem levar à relatividade de determinados fatos. Novas descobertas científicas podem, por exemplo, alterar concepções até então tidas como verdadeiras.

Os exemplos, por sua vez, são fatos típicos ou representativos de determinada situação. Garcia (1996) elucida essa técnica mencionando o caso de determinado professor, que tem que dar mais de dez aulas diárias, como exemplo típico dos sacrifícios a que se submetem os professores.

As ilustrações são um alongamento dos exemplos na forma de narrativa detalhada e entremeada de descrições. A ilustração pode ser hipotética ou real. No primeiro caso, deve ser verossímil e consistente, e deve ser introduzida no texto por meio de marcadores de hipoteticidade como “supondo que”, “imaginando que”. Embora não tenha valor como prova, tem grande valor didático e torna a tese mais aceitável. A ilustração real, por outro lado, é mais persuasiva que a hipotética porque serve como prova. Episódios históricos são listados por Garcia (1996) como exemplos de ilustração real.

Os dados estatísticos também são fatos e têm grande poder de convencimento, segundo Garcia (1996). No entanto, o autor faz uma ressalva: a interpretação dos dados estatísticos pode determinar o uso favorável ou contrário desses dados. Por exemplo, em 2014, o IBGE divulgou queda de 0,4% na taxa de analfabetismo de 2012 para 2013 entre maiores de 15 anos no Brasil (Agência Brasil, 2014). Porém, a mesma pesquisa revelou que o país ainda tinha, à época, 13 milhões de analfabetos. Esses dados podem ser utilizados tanto para valorizar ações governamentais contra o analfabetismo (queda de 0,4% de 2012 para 2013) quanto para criticar a educação no Brasil, uma vez que ainda há milhões de analfabetos no país.

Por fim, o testemunho é a apresentação de um fato por intermédio de terceiros. Deve-se, no entanto, atentar para o fato de que um mesmo fato presenciado por várias pessoas pode apresentar diferentes versões. Dessa forma, deve-se dar preferência, segundo Garcia (1996), ao testemunho autorizado. No Novo Testamento, por exemplo, a doutrina da ressurreição de Jesus Cristo se baseia nos testemunhos de pessoas que O viram ressurreto (Garcia, 1996).

Uma estrutura semelhante à de Garcia (1996) é apresentada por van Dijk (1980) para o texto argumentativo. O autor (van Dijk, 1992) chama essa estrutura de “superestrutura”, ou seja, “uma estrutura esquemática convencional” (p. 30), com formas relativamente fixas, composta por categorias vazias, que correspondem ao formato característico de um determinado tipo de texto. Nessa perspectiva, o texto argumentativo tem uma estrutura binária, composta pelas hipóteses (ou premissas) e pela conclusão.

A articulação dos argumentos se dá, segundo o autor em foco, nas hipóteses. Em geral, essa parte da

estrutura argumentativa, assim como as histórias, tem uma categoria inicial que funciona como fundo, na qual se especificam o tema dos argumentos, as noções e os objetos envolvidos, o problema e as intenções do falante.

O conceito de legitimidade, ainda segundo van Dijk (1985), diz respeito às bases semânticas que permitem que se derive uma conclusão. No exemplo “Pedro tirou quatro. Logo, não foi aprovado no exame” (tradução nossa), o argumento só terá legitimidade se for baseado em uma regra segundo a qual a nota quatro não é suficiente para aprovação em um exame. Esse argumento pode ser melhor explicado, explicitando-se a relação entre a insuficiência da nota quatro e a reprovação de Pedro. A essa categoria van Dijk (1985) chama reforço. O autor também apresenta a categoria marco. Nessa categoria, reconhece-se que a relação entre a insuficiência da nota e a reprovação só é importante em uma determinada situação: a de exames. A partir disso, pressupõe-se que Pedro realizou um exame final, dentre os quais um exame em particular tem papel mais importante.

Se uma explicação for necessária para a reprovação de Pedro, pode-se acrescentar mais um argumento, o de que Pedro não estudou o suficiente, que não se consegue uma aprovação em um exame sem esforço (justificativa). Uma exceção também poderia ser adicionada ao exemplo. Poder-se-ia afirmar que, apesar de ter obtido uma nota insuficiente em um exame, nos demais exames Pedro tirou notas boas. Ou seja, a estrutura argumentativa pode se tornar mais complexa pela recursividade da categoria argumento.

Por fim, a conclusão, segundo van Dijk (1980), contém as informações que foram inferidas das informações presentes nas hipóteses.

Fiorin e Savioli (1996) tratam tanto do texto dissertativo quanto das características da argumentação. Ao tratarem do texto dissertativo, definem-no como aquele que “analisa, interpreta, explica e avalia os dados da realidade” (p. 252). Por esse motivo, o texto dissertativo é abstrato e nele predominam interpretações genéricas. Quando situações particulares ou casos concretos ocorrem em uma dissertação, sua finalidade é ilustrar afirmações abstratas ou servir de argumento a favor ou contra essas afirmações. As relações que se estabelecem em um texto argumentativo são lógicas, tais como analogia, pertinência, causalidade, coexistência, correspondência, implicação. Em função dessas características, o tempo verbal predominante é o presente atemporal, mas outros tempos do sistema do presente também aparecem, tais como o presente com valor temporal, o pretérito perfeito e o futuro do presente. No que diz respeito à expressão da opinião, em um texto dissertativo, o ponto de vista do autor é apresentado explicitamente.

Por outro lado, ao tratarem da argumentação, Fiorin e Savioli (1996) consideram que, em sentindo amplo, todo texto é de alguma maneira argumentativo pelo fato de não

existir manifestação da linguagem que não seja persuasiva. Argumento, para os autores, é

todo procedimento lingüístico que visa a persuadir, a fazer o receptor aceitar o que lhe foi comunicado, a levá-lo a crer no que foi dito e a fazer o que foi proposto (p. 284).

No sentido restrito, há textos nos quais os mecanismos argumentativos estão explícitos, como no discurso publicitário e no discurso científico, ainda de acordo com o que defendem esses autores em estudo.

Alguns tipos de argumentos são apresentados por Fiorin e Savioli (1996): argumento de autoridade, argumento baseado no consenso, argumentos baseados em provas concretas, argumentos com base no raciocínio lógico, argumento da competência linguística.

O argumento de autoridade é a referência a autores de grande credibilidade e prestígio em determinada área de conhecimento. As referências a autores renomados tendem a aumentar a confiança do leitor ou do ouvinte nas ideias defendidas pelo produtor do texto. Pode-se evocar, aqui, a metáfora latina dos duendes apoiados sobre os ombros de gigantes (*nanos gigantum humeris insidentes*), que expressa a ideia de que novas descobertas são construídas sobre descobertas anteriores.

O argumento baseado no consenso consiste na utilização de proposições aceitas coletivamente como verdadeiras em determinado momento. Um exemplo que se destaca na atualidade é o do princípio da sustentabilidade. É consenso que os recursos naturais são finitos e que, portanto, as necessidades econômicas, sociais e ambientais devem ser conciliadas de forma que a satisfação das necessidades do presente não comprometa a satisfação das necessidades das gerações futuras.

Os argumentos baseados em provas concretas conferem credibilidade à tese defendida pelo autor do texto. Para Fiorin e Savioli (1996), podem ser provas concretas: cifras, estatísticas, dados históricos, fatos da experiência cotidiana, ilustrações (narração de um fato concreto, por exemplo) etc. Não é raro encontrarem-se, em redações de vestibular, afirmações que não podem ser comprovadas, como, por exemplo, que todo político é corrupto. Primeiramente, esse argumento seria derrubado pela referência a pelo menos um político honesto. Além disso, não seria possível encontrar provas para incriminar todos os políticos por crimes de corrupção.

No caso dos argumentos com base no raciocínio lógico, de posse de determinadas informações, o produtor do texto infere novas proposições por meio da dedução, da indução e da analogia.

Por fim, o argumento da competência linguística diz respeito ao uso da variedade linguística e do vocabulá-

rio adequados à situação comunicativa. No vestibular, por exemplo, espera-se que os alunos demonstrem domínio do português culto. Vale a pena ressaltar que dominar o português culto não significa “gotejar”, no texto, alguns vocábulos que não costumam fazer parte do léxico de jovens candidatos ao vestibular. Especialmente se o texto não foi produzido em adequação à norma culta, apresentando, por exemplo, problemas de regência, de concordância, de pontuação, de ortografia etc.

As características da dissertação e da argumentação apresentadas acima podem ser observadas claramente em materiais didáticos voltados para alunos do ensino médio. Tome-se como exemplo Sarmento e Tufano (2009). Os autores afirmam que, “na dissertação, o autor defende seu ponto de vista com uma série de argumentações, com o propósito de convencer o leitor” (p. 378), ou seja, a argumentação está contida na definição de dissertação desses autores.

No que diz respeito à estrutura, os autores, assim como Garcia (1996), apresentam a divisão em três partes:

- (i) introdução: de acordo com os autores, na introdução, o produtor do texto deve apresentar a tese, que é a tomada de posição diante da ideia-núcleo do texto;
- (ii) desenvolvimento: nessa parte, o produtor do texto deve apresentar argumentos para defender e justificar a tese proposta na introdução;
- (iii) conclusão: nessa parte, o produtor do texto deve retomar a ideia-núcleo e dar um fechamento aos argumentos apresentados ao longo do texto.

Ao tratarem do texto dissertativo-argumentativo, os autores apresentam quatro tipos de argumentos: argumentos com base em citação, argumentos com base no senso comum, argumentos com base em evidência, argumentos com base no raciocínio lógico. Esses quatro tipos de argumentos, como se pode observar, se assemelham a alguns dos tipos de argumentos mencionados anteriormente neste trabalho com base em Garcia (1996) e em Fiorin e Savioli (1996).

Na próxima seção deste trabalho, apresentam-se os procedimentos metodológicos adotados para a realização da pesquisa e, em seguida, os pressupostos teóricos aqui apresentados são utilizados para a análise do corpus.

Metodologia

A pesquisa foi realizada com um corpus constituído de 22 textos dissertativo-argumentativos produzidos por candidatos do Enem 2013 que foram disponibilizados por grandes portais de notícias da Internet. Esses textos foram digitalizados e segmentados manualmente em EDUs⁶.

⁶ *Elementary Discourse Units* são “blocos mínimos de construção de uma árvore discursiva” (Carlson e Marcu, 2001, tradução nossa). Geralmente as EDUs correspondem a orações, com exceção de orações completivas e de orações restritivas.

No Enem 2013 (INEP, 2013), o comando da produção textual foi o seguinte:

A partir da leitura dos textos motivadores seguintes e com base nos conhecimentos construídos ao longo de sua formação, redija texto dissertativo-argumentativo na modalidade escrita formal da língua portuguesa sobre o tema “Efeitos da implantação da Lei Seca no Brasil”, apresentando proposta de intervenção, que respeite os direitos humanos. Selecione, organize e relacione, de forma coerente e coesa, argumentos e fatos para defesa de seu ponto de vista.

A identificação da UC no corpus seguiu os passos a seguir:

- (i) a UC de cada um dos textos do corpus foi identificada em conjunto por dois anotadores com experiência em RST;
- (ii) o padrão foi determinado utilizando-se os mesmos fatores de pesquisa anterior (Antonio, 2015) a respeito do gênero resposta argumentativa (com exceção da retomada da pergunta, que é uma característica relevante apenas para a RA; ademais, no comando do Enem para a dissertação não havia pergunta): posição da UC no texto, frequência de substantivos, adjetivos, verbos e advérbios;
- (iii) nas redações em que houve discrepância, um superanotador (Hovy, 2010) confirmou uma UC e descartou as demais.

A pesquisa apresenta tanto um lado quantitativo, ao apresentar frequência de ocorrência de determinadas palavras, quanto um lado qualitativo, ao analisar as relações estabelecidas pelas porções textuais.

Análise

A primeira característica investigada é a posição da UC. Como os textos dissertativo-argumentativos do corpus têm extensão diferente, verificou-se a posição relativa da UC em cada texto. Os resultados são apresentados no Quadro 1. Na primeira coluna, apresenta-se a posição relativa da UC e, na coluna da direita, a quantidade de UCs encontradas naquela posição. Tomando-se as três primeiras linhas de resultados do quadro como exemplo, observa-se que, no corpus, foram encontradas duas UCs nas porções de texto que constituem os primeiros 10% dos textos; cinco UCs nas porções localizadas entre 10% e 20% da extensão dos textos; seis UCs nas porções localizadas entre 20% e 30% da extensão dos textos.

Ao contrário do gênero RA, em que a UC ocorre em posição inicial em 90,7% dos textos (Antonio, 2015)⁷, no texto dissertativo-argumentativo em apenas uma redação a primeira cláusula foi identificada como UC; em

Quadro 1. Posição relativa da UC nos textos dissertativo-argumentativos.

Chart 1. Relative position of the central unit in the argumentative texts.

Posição relativa	N
0%-10%	2
10%-20%	5
20%-30%	6
30%-40%	4
40%-50%	1
50%-60%	3
60%-90%	0
90%-100%	1

Fonte: os autores.

uma outra redação, a segunda cláusula foi identificada como UC. Por outro lado, na porção final dos textos, só se observou uma ocorrência da UC. A posição mais comum da UC nos textos do corpus é a porção que vai de 20% a 30% da extensão do texto, como pode ser observado no exemplo a seguir.

A cada ano, milhares de pessoas perdem a vida no trânsito do Brasil. A quantidade de vítimas fatais nas estradas, ruas e avenidas brasileiras pode ser comparada aos registros de mortos em conflitos armados em diversos países. Este quadro evidencia uma “guerra” cujas “munições” são chope e cerveja: metade dos acidentes fatais está relacionada ao consumo de bebidas alcoólicas por motoristas, segundo o Ministério da Saúde. *Em vigor desde 2008, a Lei Seca tem colhido resultados positivos, mas ainda são necessárias outras ações, de governos e da sociedade em geral, para que as vias brasileiras tornem-se ainda mais seguras.*

No exemplo, a UC (destacada por nós em *itálico*) encerra o primeiro parágrafo, apresentando a temática que será elaborada nos parágrafos seguintes, a saber, ações do governo e da população que tornem as vias mais seguras.

Essa diferença organizacional entre o gênero RA e o texto dissertativo-argumentativo pode ser explicada a partir de diferenças nas condições de produção. De um lado, no gênero RA, o produtor deve iniciar seu texto respondendo a pergunta que o motivou. Por outro lado, no texto dissertativo-argumentativo é necessária a criação de um pano de fundo com informações que situem o destinatário do texto com relação à temática, o que se pode observar no início

⁷ Como já se afirmou anteriormente, neste trabalho não se leva em conta a retomada da pergunta, que é característica específica do gênero RA. Ademais, o comando da prova de produção textual do Enem não continha pergunta.

do exemplo anterior. Tal observação é corroborada pela superestrutura de van Dijk (1980) do texto argumentativo já apresentada neste trabalho, segundo a qual a parte inicial do texto funciona como fundo, especificando o tema dos argumentos, as noções e os objetos envolvidos, o problema e as intenções do falante. Também é ratificada pela estrutura da dissertação de Garcia (1996) e amplamente divulgada em manuais didáticos como o de Sarmento e Tufano (2009).

Outro traço importante na busca de um padrão é a frequência de determinadas palavras utilizadas nas UCs. Esse procedimento demonstrou ser relevante na identificação das UCs no gênero RA⁸ (Antonio, 2015). Nos quadros 2 e 3, apresentam-se, respectivamente, os substantivos e os adjetivos com mais de três ocorrências nas UCs dos textos dissertativo-argumentativos do corpus.

No quadro acima, foram destacadas em *itálico* as palavras que compõem o comando da proposta de redação, uma vez que os candidatos tendem a utilizá-las como forma de garantir o atendimento do tema solicitado. O vocábulo “lei” é o que apresenta maior número de ocorrências. Quando acompanhado do adjetivo “seca”, formando o substantivo composto “lei seca”, apresenta treze ocorrências. Desacompanhado, o substantivo “lei” apre-

Quadro 2. Substantivos com mais de três ocorrências no corpus.

Chart 2. Nouns with more than three occurrences in the corpus.

Substantivos	N
<i>Lei</i>	16
Número(s)/Estatísticas/Dados	10
Acidentes	9
Diminuição/Queda/Redução	7
<i>Efeitos/Resultados/Impacto</i>	7
<i>Implantação</i>	6
<i>Brasil/País</i>	5
Motorista(s)	4
Trânsito	4
Álcool	3
Fiscalização	3

Fonte: os autores.

sentou outras três ocorrências. O substantivo “efeito(s)” (presente no comando) e os substantivos “resultados” e “impactos”, tratados pelos candidatos como seus sinônimos, tiveram sete ocorrências, e em cinco delas estavam acompanhados do adjetivo “positivo(s)”. Por sua vez, o substantivo “implantação”, também presente no texto de apoio, teve seis ocorrências. Por fim, o substantivo “Brasil” (presente no texto de apoio) e o substantivo “país”, tratado como seu sinônimo, totalizaram cinco ocorrências.

Alguns outros vocábulos tratados como sinônimos pelos produtores do texto foram agrupados para efeito de tabulação dos dados, como é o caso de “diminuição/queda/redução”, com sete ocorrências. Em geral, esses substantivos são utilizados na UC, com frequência, pelos candidatos, pelo fato de indicarem um efeito da implantação da lei seca, a saber, a diminuição do número de acidentes. Este último substantivo (“acidentes”) também foi muito utilizado, apresentando nove ocorrências. Por meio dessas construções, os candidatos atenderam o comando da proposta de redação.

Alguns outros substantivos como “motorista”, “trânsito”, “álcool” e “fiscalização” também aparecem no quadro entre as palavras com maior número de ocorrências pelo fato de ajudarem a compor o modelo cognitivo da implantação da lei seca.

O exemplo a seguir representa uma UC com as características descritas nos parágrafos anteriores. Nesse exemplo, destacam-se, em *itálico*, os vocábulos encontrados nos Quadros 2 e 3.

A Lei Seca, desde sua implantação, gerou resultados positivos tanto nas estatísticas quanto no comportamento dos motoristas.

Não foram encontrados verbos nocionais com mais de três ocorrências nas UCs. O único verbo com mais de três ocorrências é o verbo “ser”, mas, pelo fato de exercer apenas função de cópula, não será objeto de análise deste

Quadro 3. Adjetivos com mais de três ocorrências no corpus.

Chart 3. Adjectives with more than three occurrences in the corpus.

Adjetivos	N
<i>Seca</i>	13
Positivos	5

Fonte: os autores.

⁸ O tema do texto determina a frequência de ocorrência de substantivos e adjetivos. Como a prova de produção textual do Enem 2013 e a prova de redação do Vestibular de Verão 2013 da UEM trataram de temas diferentes, os quadros com os substantivos e adjetivos mais frequentes no gênero RA não serão apresentados neste trabalho.

trabalho. Nos resultados do gênero RA, o verbo “ser” também foi o que apresentou ocorrência mais alta. No entanto, o caráter mais subjetivo que a resposta pode assumir no gênero RA favoreceu a emergência de verbos evidenciais que expressam atitude proposicional. Exemplos disso são os verbos “acreditar”, “crer” e “pensar”, que atuaram como matriz da sentença na qual a resposta era dada. Um exemplo desse uso pode ser observado em “Acredito que o segredo do vestibular é uma mistura desses três fatores” (Antonio, 2015, p. 5). Por outro lado, devido ao caráter mais objetivo da expressão do ponto de vista nas dissertações, não se encontrou nenhum verbo desse tipo.

Alguns advérbios merecem destaque na identificação das UCs. O advérbio “ainda” apresentou cinco ocorrências e foi utilizado para indicar que a implantação da lei seca não solucionou totalmente o problema de acidentes de trânsito causados por motoristas alcoolizados, ou seja, ainda ocorrem acidentes, apesar da implantação da lei. É o que se pode observar no exemplo a seguir.

Apesar da queda dos números, *ainda* muitas tragédias insistem em acontecer, ocasionadas pela desastrosa união álcool e direção.

O advérbio “consideravelmente” e o adjetivo “considerável” são utilizados para modular os efeitos apresentados pela implantação da lei seca. Há resultados positivos consideráveis, mas não uma solução completa do problema dos acidentes causados por motoristas embriagados, como pode ser observado no exemplo a seguir.

Nesse aspecto, a Lei Seca implantada no Brasil reduziu *consideravelmente* o número de mortes por acidentes de trânsito. Entretanto, o individualismo da sociedade e o sistema de transporte dificultam um efeito definitivo.

Os dados de outras classes, como os conectivos, por exemplo, não foram tabulados pelo fato de essas palavras se repetirem ao longo do texto. No caso de se criar um sistema de identificação automática das UCs, a máquina poderia identificar uma EDU equivocadamente como UC ao encontrar um conectivo que estivesse presente tanto na UC quanto em outras partes do texto. O conectivo “mas”, que apresenta a maior frequência de ocorrência nas UCs (duas ocorrências), por exemplo, apresenta 44 ocorrências em todo o corpus. Dessa forma, utilizar esse conectivo como fator de identificação das UCs não seria boa ideia, pois há mais ocorrências dele em EDUs que não são UCs do que em EDUs que são UCs.

Conclusão

Neste trabalho, objetivou-se investigar critérios para a identificação da unidade central (UC) de textos dissertativo-argumentativos. Para isso, aplicaram-se os mesmos critérios de trabalho anterior sobre textos do gêne-

ro resposta argumentativa a um corpus formado por textos dissertativo-argumentativos produzidos por candidatos do Exame Nacional do Ensino Médio de 2013. Pelo fato de se utilizarem os mesmos critérios, retomaram-se, para fins de comparação, os resultados dessa pesquisa anterior. O aparato teórico-metodológico utilizado foi a RST.

Ao contrário do gênero RA, em que a UC ocorre em posição inicial na maioria dos textos investigados por Antonio (2015), no texto dissertativo-argumentativo a posição mais comum da UC é na porção que vai de 20% a 30% da extensão do texto. Em apenas dois textos dissertativo-argumentativos a UC apareceu no início do texto. Nessa posição, a função da UC é apresentar a temática que será elaborada nos parágrafos seguintes. Essa diferença organizacional entre o gênero RA e o texto dissertativo-argumentativo pode ser explicada com base nas diferenças das condições de produção de um e de outro texto. De um lado, no gênero RA, o produtor deve iniciar seu texto respondendo a pergunta que o motivou. Por outro lado, no texto dissertativo-argumentativo é necessária a criação de um pano de fundo com informações que situem o destinatário do texto com relação à temática.

Outro traço importante na busca de um padrão é a frequência de determinadas palavras utilizadas nas UCs. Substantivos e adjetivos que compõem o comando da proposta de redação apresentaram alta frequência, uma vez que os candidatos tendem a utilizar essas palavras como forma de garantir o atendimento ao tema solicitado.

Alguns outros substantivos como “motorista”, “trânsito”, “álcool” e “fiscalização” também aparecem no quadro entre as palavras com maior número de ocorrências pelo fato de ajudarem a compor o modelo cognitivo da implantação da lei seca.

No que diz respeito aos verbos, não foram encontrados verbos nocionais com mais de três ocorrências nas UCs. O único verbo com mais de três ocorrências é o verbo “ser”, mas, pelo fato de exercer apenas função de cópula, não foi objeto de análise deste trabalho. Por outro lado, o caráter mais subjetivo que a resposta pode assumir no gênero RA favoreceu a emergência de verbos evidenciais que expressam atitude proposicional, como os verbos “acreditar”, “crer” e “pensar”. Esses verbos atuaram como matriz da sentença na qual a resposta era dada. Já nos textos dissertativo-argumentativos, devido ao caráter mais objetivo da expressão do ponto de vista, não se encontrou nenhum verbo desse tipo.

Alguns advérbios merecem destaque na identificação das UCs. O advérbio “ainda” apresentou cinco ocorrências e foi utilizado para indicar que a implantação da lei seca não solucionou totalmente o problema de acidentes de trânsito causados por motoristas alcoolizados, ou seja, ainda ocorrem acidentes, apesar da implantação da lei. O advérbio “consideravelmente” e o adjetivo “considerável” são utilizados para modular os efeitos apresentados pela implantação da lei seca. Há resultados positivos con-

sideráveis, mas não uma solução completa do problema dos acidentes causados por motoristas embriagados.

Espera-se que os resultados desta pesquisa possam subsidiar a criação de aplicações de avaliação automática de textos, bem como a preparação de materiais para treinamento das bancas de avaliação e de manuais de produção textual para alunos do ensino médio. Em termos de trabalhos futuros, pretende-se trabalhar em conjunto com pesquisadores de Processamento de Línguas Naturais para se disponibilizar ferramenta on-line para identificação de UCs de textos argumentativos.

Referências

- AGÊNCIA BRASIL. 2014. Analfabetismo cai 0,4 pontos percentuais mas ainda atinge 13 milhões. Disponível em: <http://agenciabrasil.ebc.com.br/educacao/noticia/2014-09/Analfabetismo-cai-0,4-pontos-percentuais-mas-ainda-atinge-13-milhoes>. Acesso em: 10/07/2015.
- ANTONIO, J.D. 2015. Detecting central units in argumentative answer genre: signals that influence annotators' agreement. In: Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural, XXXI, Alicante, 2015. *Actas...* Disponível em: http://www.academia.edu/27165050/Detecting_central_units_in_argumentative_answer_genre_signals_that_influence_annotators_agreement. Acesso em: 10/07/2015.
- BURSTEIN, J.; MARCU, D.; ANDREYEV, S.; CHODOROW, M. 2001. Towards automatic classification of discourse elements in essays. In: Annual Meeting on Association for Computational Linguistics, 39, Toulouse, 2001. *Proceedings...* Toulouse, ACL, p. 98–105.
- CARLSON, L.; MARCU, D. 2001. *Discourse Tagging Reference Manual*. Los Angeles, University of Southern California, 87 p.
- EGG, M.; REDEKER, G. 2010. How complex is discourse structure. In: International Conference on Language Resources and Evaluation, 7, Valetta, 2010. *Proceedings...* Valetta, LREC, p. 1619–1623.
- FIORIN, J.L.; SAVIOLI, F.P. 1996. *Lições de texto: leitura e redação*. São Paulo, Ática, 416 p.
- GARCIA, O.M. 1996. *Comunicação em prosa moderna*. 21ª ed., Rio de Janeiro, FGV, 522 p.
- HOVY, E. 2010. Annotation: A tutorial. Disponível em: <http://acl2010.org/tutorials.html>. Acesso em: 10/07/2015.
- INEP. 2013. Caderno do Enem 2013. Disponível em: http://download.inep.gov.br/educacao_basica/enem/provas/2013/caderno_enem2013_dom_rosa.pdf. Acesso em: 20/05/2016.
- IRUSKIETA, M.; ILARRAZA, A.D.; LABAKA, G.; LERSUNDI, M. 2015. The Detection of Central Units in Basque scientific abstracts. In: Congreso de la Sociedad Española del Procesamiento del Lenguaje Natural, XXXI, Alicante, 2015. *Actas...* Alicante, SEPLN.
- IRUSKIETA, M.; ILARRAZA, A.D.; LERSUNDI, M. 2014. The annotation of the Central Unit in Rhetorical Structure Trees: A Key Step in Annotating Rhetorical Relations. In: International Conference on Computational Linguistics, 25, Dublin, 2014. *Proceedings...* Dublin, COLING, p. 466–475.
- MANN, W.C.; TABOADA, M. 2010. RST Web Site. Disponível em: <http://www.sfu.ca/rst>. Acesso em: 10/07/2015.
- MANN, W.C.; THOMPSON, S.A. 1983. *Relational Propositions in Discourse*. Los Angeles, Information Sciences Institute, 38 p.
- MANN, W.C.; THOMPSON, S.A. 1988. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8(3):243–281. <https://doi.org/10.1515/text.1.1988.8.3.243>
- MATTHIESSEN, C. 2005. Remembering Bill Mann. *Computational Linguistics*, 31(2):161–172. <https://doi.org/10.1162/0891201054224002>
- O'DONNELL, M. 2000. RSTTool 2.4: A markup tool for Rhetorical Structure Theory. In: International Conference on Natural Language Generation, 1, Mitzpe Ramon, 2000. *Proceedings...* Mitzpe Ramon, INLG, p. 253–256.
- PARDO, T.A.S.; RINO, L.H.M.; NUNES, M.G.V. 2003. GistSumm: A summarization tool based on a new extractive method. In: N.J. MAMEDE; J. BAPTISTA; I. TRANCOSO; M.G.V. NUNES (eds.), *Computational Processing of the Portuguese Language*. Berlin, Springer-Verlag, p. 210–218. https://doi.org/10.1007/3-540-45011-4_34
- SARMENTO, L.L.; TUFANO, D. 2009. *Português: Literatura, Gramática, Produção de Texto*. São Paulo, Moderna, 439 p.
- STEDE, M. 2008. RST revisited: Disentangling nuclearity. In: C. FABRICIUS-HANSEN; W. RAMM (eds.), *"Subordination" versus "coordination" in sentence and text*. Amsterdam, John Benjamins, p. 33–57. <https://doi.org/10.1075/slcs.98.03ste>
- TABOADA, M. 2009. Implicit and explicit coherence relations. In: J. RENKEMA (ed.), *Discourse, of course*. Amsterdam, John Benjamins, p. 125–138. <https://doi.org/10.1075/z.148.13tab>
- TABOADA, M.; GÓMEZ-GONZÁLEZ, M.L.Á. 2013. Discourse markers and coherence relations: Comparison across markers, languages and modalities. In: M. TABOADA; S.D. SUÁREZ; E.G. ÁLVAREZ (eds.), *Contrastive Discourse Analysis: Functional and Corpus Perspectives*. Sheffield, Equinox, p. 17–41.
- VAN DIJK, T.A. 1980. *Macrostructures: An interdisciplinary study of global structures in discourse, interaction and cognition*. Hillsdale, Lawrence Erlbaum, 317 p.
- VAN DIJK, T.A. 1985. *La ciencia del texto*. 3ª ed., Barcelona, Paidós, 309 p.
- VAN DIJK, T.A. 1992. *Cognição, discurso e interação*. São Paulo, Contexto, 207 p.

Submetido: 21/10/2016

Aceito: 25/05/2017