

# **WordSmith Tools: para uma análise da coesão sequencial em redações dissertativas argumentativas**

**Daniela Faria Grama**

Universidade Federal de Uberlândia (UFU), Uberlândia, Minas Gerais, Brasil  
daniela\_grama@hotmail.com

DOI: <http://dx.doi.org/10.21165/el.v45i2.598>

## **Resumo**

Este trabalho tem o objetivo de apresentar a metodologia, baseada em Linguística de *Corpus*, utilizada no estudo piloto da nossa dissertação de mestrado. O objetivo de tal estudo foi analisar quais elementos coesivos sequenciais foram usados de maneira inadequada em redações do tipo dissertativo-argumentativo estilo Enem, a fim de verificar se esse é um critério válido para escolhermos quais elementos coesivos serão alvo de estudo na nossa pesquisa de mestrado acadêmico que está em andamento. Nesse processo de experimentação, a abordagem da Linguística de *Corpus* e o programa de análise lexical *WordSmith Tools* mostraram-se valiosos, pois nos ajudaram sobremaneira a chegar aos resultados do estudo piloto.

**Palavras-chave:** Linguística de *Corpus*; coesão sequencial; redações dissertativas argumentativas.

## **WordSmith Tools: For an Analysis of Sequential Cohesive Links in dissertative-Argumentative texts**

### **Abstract**

This paper aims to present a methodology based on Corpus Linguistics, which was used in the pilot study of a master research our dissertation. The purpose of this study was to analyze what sequential cohesive links were used improperly in dissertative-argumentative texts that present the same styles as Enem does, in order to verify if this is a valid criterion for choosing what cohesive elements will be studied in the ongoing our master research. In this experimentation process, the approach of Corpus Linguistics with the lexical analysis program WordSmith Tools proved valuable, because they greatly helped us to achieve get the pilot study results.

**Keywords:** Corpus Linguistics; sequential cohesion; dissertative-argumentative texts.

## **Introdução**

Neste artigo, daremos enfoque à metodologia – baseada em Linguística de *Corpus* – adotada na realização de um estudo piloto para nossa dissertação de mestrado. O objetivo desse estudo é identificar quais elementos coesivos sequenciais existem em um *corpus* eletrônico de 237 redações dissertativas argumentativas estilo Exame Nacional do Ensino Médio (Enem) e, principalmente, quais são usados de maneira inadequada, uma vez que esse foi o critério escolhido para selecionar quais elementos coesivos sequenciais devem ser alvo de (re)definição lexicográfica, que é basicamente a proposta da nossa pesquisa em nível de mestrado acadêmico.

Vale ressaltar que a motivação para iniciar esta pesquisa surgiu após o desenvolvimento de uma Iniciação Científica (IC) denominada “Problemas de coesão na escrita dos gêneros discursivos da ordem do relatar: ‘notícia e relato’”, conforme Grama

(2013), e, principalmente, após experiências profissionais voltadas à área de redação. A atuação em monitorias e avaliações de redações, além da prática docente referente à disciplina de Redação, nos levou, por ocasião, a procurar em dicionários a definição de determinados elementos coesivos sequenciais, na tentativa de encontrar possíveis informações sobre seus empregos e sentidos, além de exemplos de uso. Como tal expectativa não foi correspondida, já que nos deparamos apenas com sinônimos que resultaram em círculo vicioso, pensamos que uma investigação a respeito do tratamento lexicográfico desse tipo de palavra poderia ser feita.

No que diz respeito ao estudo piloto, esclarecemos que ele é entendido como uma amostra do que uma pesquisa de mestrado ou doutorado será no seu fim; é também uma forma de o pesquisador testar os caminhos escolhidos por ele na elaboração de um projeto científico e uma oportunidade para conhecer as dificuldades e os problemas que seu trabalho poderá mostrar antes de ser tomado em proporção maior. Portanto, consideramos que esse procedimento é uma escolha acertada e prudente por parte do pesquisador. A seguir, damos início ao respaldo teórico utilizado em nossos estudos.

## 1. Fundamentação teórica

Nesta seção, apresentamos alguns conceitos e procedimentos que envolvem a metodologia/abordagem da Linguística de *Corpus*, tais como: a preparação do *corpus* e o uso de um programa de análise lexical (no nosso caso, o *WordSmith Tools*). Além disso, falamos também sobre a coesão sequencial, que diz respeito à função das palavras que estudamos.

### 1.1. Linguística de *Corpus*

Neste trabalho, lançamos mão da metodologia/abordagem da Linguística de *Corpus* (LC) para uma análise da linguagem verbal escrita. Em virtude disso, é pertinente situar o leitor sobre alguns conceitos e procedimentos que envolvem a referida análise. O primeiro deles diz respeito à preparação do *corpus*, que é um conjunto de textos compilados, de maneira especial, para se alcançar um objetivo.

A Linguística de *Corpus* ocupa-se da coleta e exploração de corpora, ou conjunto de dados linguísticos textuais coletados criteriosamente, com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à exploração da linguagem por meio de evidências empíricas, extraídas do computador (BERBER SARDINHA, 2004, p. 3).

Mi definición de corpus corresponde a un conjunto amplio de textos digitales de naturaleza específica y que cuenta con una organización predeterminada en torno a categorías identificables para la descripción y análisis de una variedad de lengua (PARODI, 2010, p. 25).

Quando falamos em LC, pressupomos a existência de um *corpus* que é delineado de modo específico. Tal especificidade justifica o “criteriosamente” presente na citação acima de Berber Sardinha. Conforme Fromm (2003) afirma, a elaboração de um *corpus* na LC precisa ser bem planejada, e isso depende das decisões que o pesquisador toma. Algumas questões podem nortear essa primeira fase do trabalho sob a perspectiva da LC, tais como: qual é a finalidade do *corpus*? O que será investigado?

Que tamanho o *corpus* terá e quais gêneros textuais de qual língua farão parte dele? Quais formas de compilação, organização e armazenamento serão utilizadas? É possível formar o *corpus* apenas com o auxílio da internet ou será preciso coletar os textos de outras maneiras que exijam autorização?

É fato que a LC tem como forte característica dar a possibilidade de o pesquisador trabalhar com um número grande de textos, isto é, dezenas, centenas, milhares, enfim uma quantidade que seria praticamente impossível de ser analisada apenas sob o olhar humano. Por isso, a LC conta com o apoio de programas computacionais. Um deles, considerado por Berber Sardinha (2009) como merecedor de destaque e que é usado em nossa pesquisa, é o *WordSmith Tools* versão 5 (SCOTT, 2008). O programa de análise lexical conta com três ferramentas fundamentais: *WordList*, *Concord* e *Keywords*, que ocasionam variadas formas de trabalhar com o *corpus*.

A intenção do programa é servir como uma ferramenta que permita a consecução de tarefas relacionadas a análises de corpora. Isso significa que ele não foi concebido para efetuar por si só uma determinada análise para o usuário. Em outras palavras, o *WordSmith Tools* não foi feito para efetuar análises de projetos específicos; ele disponibiliza uma série de opções de ferramentas (daí o ‘tools’ em seu nome), algumas mais gerais, outras mais restritas, sem jamais supor que a análise termine com o processamento de dados que ele efetua (BERBER SARDINHA, 2009, p. 7).

Para aqueles que não conhecem a LC, é fundamental enfatizar que os programas não excluem a análise humana em relação ao *corpus*, eles apenas são facilitadores e trazem dados que seriam imperceptíveis e/ou incontáveis diante das limitações que o olhar humano possui. O trabalho que o linguista tem continua sendo indispensável, pois é o seu ponto de vista diante dos dados empíricos que dará direcionamento à sua pesquisa.

Além de ter como princípio a valorização de dados reais da língua para observação, a LC nos permite ver a língua como um sistema probabilístico: “a visão da linguagem como sistema probabilístico pressupõe que, embora muitos traços linguísticos sejam possíveis teoricamente, não ocorrem com a mesma frequência” (BERBER SARDINHA, 2004, p. 31). Um trabalho realizado pelo viés da LC considera a frequência e a posição de uma palavra no *corpus* como dois fatores que dizem muito sobre uma língua. Uma visão ainda pouco disseminada nos estudos linguísticos no Brasil, mas que é bastante promissora.

Dizer que a variação não é aleatória, na verdade, é afirmar que a linguagem é padronizada (*patterned*). A padronização se evidencia pela recorrência, isto é, uma colocação, coligação ou estrutura que se repete significativamente mostra sinais de ser, na verdade, um padrão lexical ou léxico-gramatical. A linguagem forma padrões que apresentam regularidade (estáveis em momentos distintos, isto é, têm frequência comparável em corpora distintos) e variação sistemática (correlacionam-se com variedades textuais, genéricas, dialetais etc.) (BERBER SARDINHA, 2004, p. 31).

Vale ressaltar que a frequência e a posição das palavras juntas dizem que a língua segue determinado padrão, que há preferências de uso e que os elementos lexicais e gramaticais não ocorrem de modo isolado, eles estão em companhia um do outro.

Ao contrário da visão tradicional de que a língua é composicional, ou seja, conhecendo o léxico e as regras gramaticais de uma língua, é possível ter absoluto domínio dela, na visão probabilística da língua, essa é formada por um grande número de unidades pré-fabricadas, que são aprendidas como um todo (TAGNIN, 2011, p. 284).

Essa visão em relação à língua dá a possibilidade de um aprendiz ou pesquisador conhecer melhor um idioma dentro de determinada realidade de uso dele. Por meio de um *corpus*, é possível visualizar a questão das unidades pré-fabricadas e várias outras, afinal o *corpus* “es sólo una colección finita de um universo infinito” (PARODI, 2010, p. 24).

De acordo com Shepherd (2012), os estudos com a metodologia/abordagem da LC ultimamente fazem interfaces com diversas áreas, não apenas com a Lexicografia, mas também com a Literatura, a Análise do Discurso, a Tradução, entre outras. No nosso caso, ou melhor, para a nossa pesquisa de mestrado, estamos relacionando a LC com a Linguística Textual e a Lexicografia Pedagógica; em específico, estamos nos valendo da abordagem para chegarmos à construção de um material que contenha definições para os elementos coesivos sequenciais. Diante disso, consideramos necessário falar sobre esses elementos de modo a esclarecer o que são e qual a sua importância.

## 1.2. Elementos coesivos sequenciais

Os elementos coesivos sequenciais são palavras, em sua maioria, gramaticais, que possuem a função de unir as frases e parágrafos de um texto, estabelecendo relações semânticas entre tais partes. São responsáveis pela progressão, pelo encadeamento e pela sequenciação de uma produção textual.

A coesão sequencial diz respeito aos procedimentos linguísticos por meio dos quais se estabelecem, entre segmentos do texto (enunciados, partes de enunciados, parágrafos e sequências textuais), diversos tipos de relações semânticas e/ou pragmáticas, à medida que se faz o texto progredir (KOCH, 2008, p.53).

Segundo Koch (2008), a coesão sequencial pode ser dividida em sequenciação parafrástica e frástica. A frástica apresenta dois mecanismos que são alvo de nossa pesquisa de mestrado: o de justaposição e o de conexão:

A justaposição com elementos sequenciadores estabelece um sequenciamento coesivo entre porções maiores ou menores da superfície textual. Tais elementos, denominados sinais de articulação operam, portanto, em diversos níveis hierarquizados: meta-nível ou nível metacomunicativo, em que funcionam como sinais demarcatórios e/ou sumarizadores de partes ou sequências textuais (ex.: por consequência, em virtude do exposto, dessa maneira, em resumo, essa posição, etc.) (KOCH, 2008, p. 66).

Outro tipo de sinais de articulação são os conectores interfrásticos, responsáveis pelo tipo de encadeamento a que se tem denominado conexão ou junção. (cf. a conjunção de Halliday e Hasan). Trata-se de conjunções, advérbios sentenciais (também chamados de advérbios de texto) e outras palavras (expressões) de ligação que estabelecem, entre orações, enunciados ou partes do texto, diversos tipos de relações semânticas e/ou pragmáticas (KOCH, 2008, p. 68).

Conforme Koch (2011), os elementos coesivos são considerados pela gramática apenas como morfemas gramaticais, em oposição aos lexicais; o valor argumentativo e as possibilidades de sentido que imprimem, muitas vezes, são descartados no ensino regular. Nessa perspectiva, tanto Koch (2011) quanto Pauliukonis (2014) e Antunes (2005) ressaltam a importância do tratamento desses operadores:

Tanto nas gramáticas, como no ensino de língua materna, tem-se dado maior ênfase ao estudo dos morfemas lexicais e dos morfemas gramaticais flexionais e derivacionais, relegando-se a um plano totalmente secundário os elementos aqui abordados. Desse modo, eles passam praticamente despercebidos ao aprendiz, que – na melhor das hipóteses – limita-se a decorá-los, sem lhes dar maior atenção (KOCH, 2011, p. 107).

Para desenvolver a competência comunicativa dos alunos, que é um dos objetivos do ensino de gramática reflexiva, é necessária a aquisição de novas habilidades, que um enfoque discursivo do emprego dos conectores, por exemplo, pode ajudar a implementar. Ao lado desse ensino reflexivo e produtivo, o descritivo e o normativo também têm seu lugar, evidentemente, mas devem ser redimensionados, em comparação com as práticas que se têm observado tradicionalmente nas escolas, sobretudo aquelas relacionadas ao ensino de conectores e de outros elementos linguístico-discursivos na construção dos sentidos do texto (PAULIUKONIS, 2014, p. 40).

Nas gramáticas, em geral, a função atribuída aos conectores se resume àquela de unir termos de uma oração ou orações. Pouco ou nenhum destaque é dado à ligação entre períodos, entre parágrafos ou até mesmo entre blocos maiores do texto. As gramáticas costumam atribuir aos conectores, particularmente, às conjunções, um sentido a partir do qual se pode reconhecer o tipo de relação estabelecida (relação de causa, de tempo, de oposição, de adição, entre outras). Entretanto, a identificação desse sentido das conjunções e locuções tem servido, praticamente, somente para se chegar a uma classificação dessas conjunções e das respectivas orações em que aparecem (ANTUNES, 2005, p. 141-142).

Por meio desses excertos, é visível que os conectores sequenciais precisam ser estudados de maneira mais cautelosa, levando-se em consideração a sua função em determinado contexto linguístico. E, para que possamos analisar o uso desses elementos em grande proporção e para que elaboremos uma proposta lexicográfica para essas palavras, a Linguística de *Corpus* será a rota escolhida, uma grande aliada durante nossa jornada.

## **2. Metodologia e procedimentos**

Nesta parte, discorreremos sobre a coleta e preparação do nosso *corpus* de estudo; definiremos a sua tipologia e destacamos dois procedimentos: o reconhecimento de palavras candidatas a elementos coesivos sequenciais e a verificação dos usos inadequados dos elementos coesivos sequenciais.

## 2.1. A coleta e a preparação do corpus de redações

Para o estudo piloto, foram coletadas 237 redações do tipo dissertativo-argumentativo estilo Enem, referentes ao ano de 2014 e oriundas do *site* UOL Educação, que disponibiliza, desde 2007, propostas de redações para que os internautas possam praticar a escrita. O *site* conta com professores para corrigir as redações enviadas e mensalmente publica esses textos, autorizando que sejam reproduzidos para fins de estudo.

Inicialmente, cada redação foi copiada do *site* e colada em um arquivo *Word*, conforme ilustram as Figuras 1 e 2 a seguir:

UOL Educação Web | Dicionários | Imagens

Atualidades Banco de Redações Bullying Enem por escola Nova ortografia Planos de aula Trote universitário Notícias Fotos Vídeos Infográficos Dicionários Onde estudar País e professores Pesquisa escolar Testes e simulados

**BANCO DE REDAÇÕES**

REDAÇÃO

Aluno: \*\*\*  
Idade: \*\*\*  
Colégio: \*\*\*

NOTA 5,0

**As duas vias da ciclofaixa**

Um dos temas mais discutido [discutidos] hoje no Brasil é a mobilidade urbana, devido aos problemas gerados pela super população [superpopulação] de veículos nos grandes centro urbanos. As prefeituras tem [têm] tentado criar alternativas viárias para minimizar o fluxo de veículos em avenidas e ruas da cidade, uma delas são as ciclofaixas. Mas os problemas não acabam por aí. [aí, a] falta de estrutura nos grandes centros são [é] um dos obstáculos enfrentado pelo poder público, além é claro [além, é claro,] das opiniões controversa [controversas] sobre qualquer projeto urbano que mexa com a vida das pessoas. [No caso das ciclofaixas [ciclofaixas,] parte da população é a favor da criação desses caminhos alternativos, desde que não afetem o seu cotidiano, pois com a criação da ciclofaixa e [é] inevitável a

Figura 1. Imagem parcial de uma redação avaliada pelo UOL Educação

Fonte: <<http://educacao.uol.com.br/bancodereacoes>>

Acesso em: 7 jan 2015.

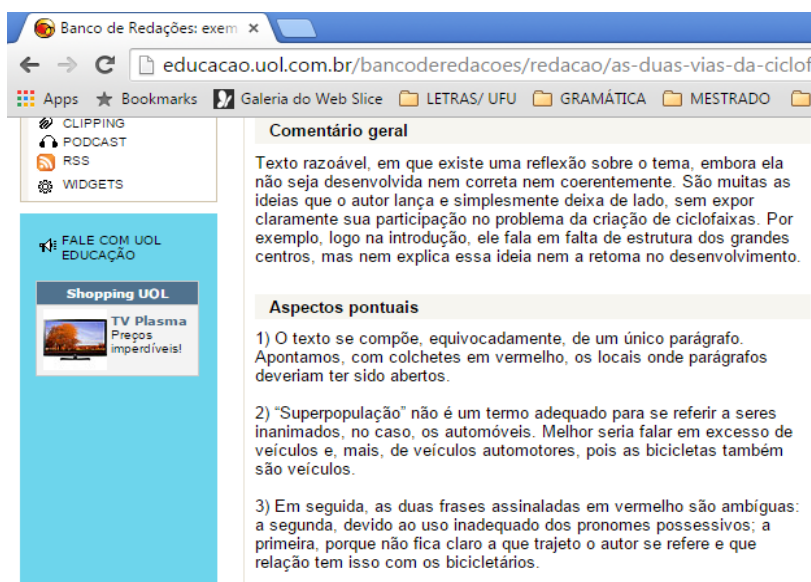
### As duas vias da ciclofaixa

Um dos temas mais discutido [discutidos] hoje no Brasil é a mobilidade urbana, devido aos problemas gerados pela super população [superpopulação] de veículos nos grandes centro urbanos. As prefeituras tem [têm] tentado criar alternativas viárias para minimizar o fluxo de veículos em avenidas e ruas da cidade, uma delas são as ciclofaixas. Mas os problemas não acabam por aí. [aí, a] falta de estrutura nos grandes centros são [é] um dos obstáculos enfrentado pelo poder público, além é claro [além, é claro,] das opiniões controversa [controversas] sobre qualquer projeto urbano que mexa com a vida das pessoas. [No caso das ciclofaixas [ciclofaixas,] parte da população é a favor da criação desses caminhos alternativos, desde que não afetem o seu cotidiano, pois com a criação da ciclofaixa e [é] inevitável a alteração no trânsito [trânsito] local. Por exemplo [exemplo,] em Belo Horizonte foi criado [Horizonte, foram criados] os bicicletários, mantido [mantidos] por um grande banco privado, a [privado. A] ideia é até boa mas, verificando os caminhos para percorrer com a bike [bike,] as faixas permitidas acabam antes mesmo do término [término] do seu trajeto. [Para que o projeto de ciclofaixas funcione de verdade [verdade,] é preciso uma ajuda mútua [mútua] entre a população e o poder público, desde suas discussões até a sua implantação, infelizmente [implantação, infelizmente,] isso é algo que não se vê, pois a ideia do governante é meramente fazer meia dúzias [dúzia] de metros de ciclofaixas para mostrar a sua atenção ao tema. E a população também [também,] por sua vez [vez,] tem que fiscalizar e ajudar a preservar o bem público para que a ciclofaixa funcione mesmo como ciclofaixa e não como mão preferencial [preferencial] para motociclistas e [ou] estacionamento de veículos, algo que é muito comum nas poucas ciclofaixas existentes.

Figura 2. Cópia da redação em arquivo *Word*

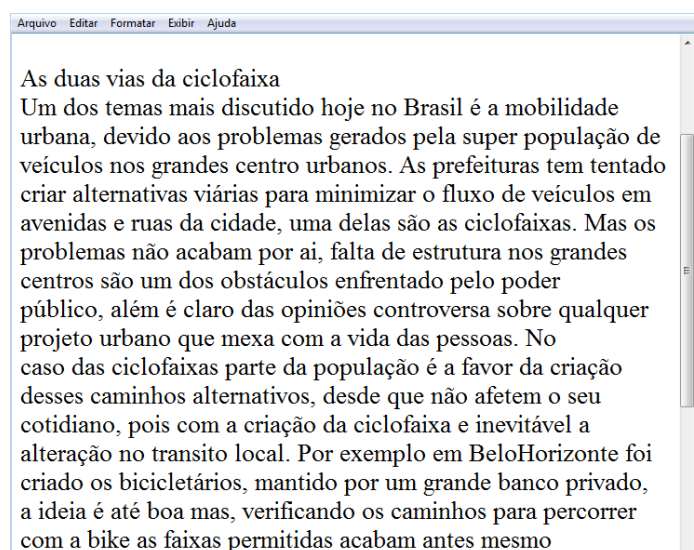
Fonte: Elaboração própria

No arquivo *Word*, foi possível iniciar o processo de limpeza dos textos que, conforme a figura mostra, são publicados com as correções dos professores realizadas na cor verde e dentro de colchetes. Os destaques feitos na cor vermelha são para chamar a atenção do escritor e, por vezes, são comentados à parte nas seções abaixo da redação, como podemos ver na Figura 3:



**Figura 3. Comentários sobre uma redação avaliada**  
Fonte: <<http://educacao.uol.com.br/bancoderedacoes>>  
Acesso em: 7 jan 2015.

Após limpar todas as intervenções referentes à revisão textual e separar as palavras que estavam unidas devido ao ato de copiar e colar, os textos foram salvos em arquivos TXT (texto sem formatação) e em codificação *Unicode*, para que pudessem, enfim, ser lidos pelo programa de análise lexical *WordSmith Tools*. A Figura 4 exemplifica uma redação pronta para ser trabalhada no *WordSmith Tools*:



**Figura 4. Imagem parcial de uma redação limpa em arquivo TXT**  
Fonte: Elaboração própria

## 2.2. Tipologia do corpus

Após o processo de compilação, o *corpus* para o estudo piloto apresentou-se com as seguintes características tipológicas:

**Quadro 1. Tipologia do corpus de estudo<sup>1</sup>**

<b>Língua</b>	Monolíngue (português)
<b>Modo</b>	Escrito
<b>Tempo</b>	Sincrônico/ Contemporâneo
<b>Seleção</b>	Estático
<b>Conteúdo</b>	Especializado (redações dissertativas argumentativas estilo Enem)
<b>Autoria</b>	Falantes nativos e não nativos/ individual
<b>Finalidade</b>	De estudo (análise dos usos inadequados dos elementos coesivos sequenciais)
<b>Tamanho</b>	Pequeno (58.169 <i>tokens</i> )
<b>Nível de codificação</b>	Com cabeçalho (local e data), com nomeação e sem etiquetas

Fonte: Elaboração própria

## 2.3. Corpus de redações: palavras candidatas a elementos coesivos sequenciais

Com o *corpus* de redações pronto, o próximo passo foi usar a ferramenta *WordList* do programa *WordSmith Tools*. Ela exibe uma lista de palavras do *corpus* e dá a possibilidade de visualizá-la em ordem alfabética e em ordem de frequência. A Figura 5 apresenta a nossa *WordList* por ordem das palavras mais frequentes:

N	Word	Freq	%	Texts	%_lemma	Se
1	A	2.321	3,99	237	100,00	
2	DE	2.311	3,97	237	100,00	
3	QUE	1.877	3,23	236	99,58	
4	O	1.824	3,14	237	100,00	
5	E	1.710	2,94	235	99,16	
6	É	896	1,54	219	92,41	
7	PARA	844	1,45	220	92,83	
8	NÃO	729	1,25	209	88,19	
9	DA	720	1,24	209	88,19	
10	DO	693	1,19	212	89,45	
11	EM	685	1,18	225	94,94	
12	UMA	645	1,11	208	87,76	
13	UM	641	1,10	213	89,87	
14	OS	635	1,09	204	86,08	
15	COM	569	0,98	204	86,08	
16	SE	528	0,91	188	79,32	
17	COMO	439	0,75	195	82,28	
18	POR	430	0,74	189	79,75	
19	MAIS	422	0,73	174	73,42	
20	AS	407	0,70	175	73,84	
21	NO	398	0,68	170	71,73	
22	NA	341	0,59	168	70,89	
23	DOS	311	0,53	160	67,51	

**Figura 5. Parte da WordList do corpus de redações em ordem de frequência**

Fonte: Elaboração própria

A partir da *WordList*, fizemos uma leitura de todas as palavras listadas e anotamos em um quadro num documento em formato *Word* aquelas candidatas a elementos coesivos sequenciais, juntamente com a sua frequência. É importante ressaltar que, embora o programa facilite bastante a análise do *corpus* de pesquisa, em alguns momentos essa só pode ser feita por meio do olhar humano e do esforço do pesquisador. No nosso caso, a anotação das palavras candidatas a elementos coesivos sequenciais foi uma tarefa árdua e que não teve a pretensão de ser levada à exaustão.

<sup>1</sup> Seguimos as nomenclaturas mencionadas por Berber Sardinha (2004) para caracterizar nosso *corpus*.



Posteriormente, todas as palavras candidatas a elementos coesivos sequenciais tiveram de ser verificadas em seu contexto linguístico no *corpus*, para que pudéssemos considerá-las como elementos coesivos sequenciais de fato. Para isso, utilizamos a ferramenta *Concord* do mesmo programa de análise lexical. Para ilustrar tal atividade, segue abaixo a Figura 6, que mostra uma palavra candidata a elemento coesivo sequencial “*seja*” que, de fato, cumpriu essa função ao verificarmos as linhas de concordância em que ela era usada na forma “*ou seja*”. Na sequência, a Figura 7 exemplifica uma palavra candidata a elemento coesivo sequencial “*melhor*” que não cumpriu a função de auxiliar na sequenciação e, por isso, foi descartada no estudo piloto.

N	Concordance
1	de um de um candidato que não <b>seja</b> o daquele visto em panfletos
2	que entretenha, ou que ao menos <b>seja</b> bem feita e infelizmente não é
3	eles para tomar uma decisão que lhes <b>seja</b> satisfatória. Aparentemente não há
4	toda a parte visual da campanha, ou <b>seja</b> , tem mais chance de atrair a
5	vivem na pobreza extrema, ou <b>seja</b> , tem rendimento médio domiciliar
6	participar da vida política, ou <b>seja</b> , tomar conhecimentos das leis que
7	o horário político fazendo com que ele <b>seja</b> levado a serio e surta o efeito que
8	do parágrafo anterior, talvez o objetivo <b>seja</b> alcançado a partir do momento
9	portadoras de alguma dificuldade, <b>seja</b> ela física ou Mental (este um bom
10	eleitorais, já estão sendo transmitidas, <b>seja</b> nos rádios ou nas televisões. Com

Figura 6. Parte das linhas de concordância da palavra “*seja*”

Fonte: Elaboração própria

N	Concordance	Set
38	das manifestações é necessário um <b>melhor</b> treinamento da corporação de	
39	que tem uma situação financeira <b>melhor</b> possua uma qualidade de vida	
40	tomar as atitudes necessárias para o <b>melhor</b> desenvolvimento da cidade,	
41	e o eleito, pode-se fazer um governo <b>melhor</b> tornando o país muito mais	
42	mais pobre desejará apenas uma vida <b>melhor</b> , em quanto a “burguesia”, por	
43	cidadão se conscientize em fazer o <b>melhor</b> , sem tirar vantagens, pois assim	
44	seu preço são únicos, a maioria elege o <b>melhor</b> que acha, o perdedor é a minoria	
45	, confrontamento com a policia etc. A <b>melhor</b> forma de protesto é nas urnas e	
46	da população nas redes sociais, o <b>melhor</b> momento de demonstrar toda	
47	suas crenças e de optar por aquilo que <b>melhor</b> lhe aprover. A descrença na	
48	a escolha do candidato que <b>melhor</b> condiz com as expectativas do	

Figura 7. Parte das linhas de concordância da palavra “*melhor*”

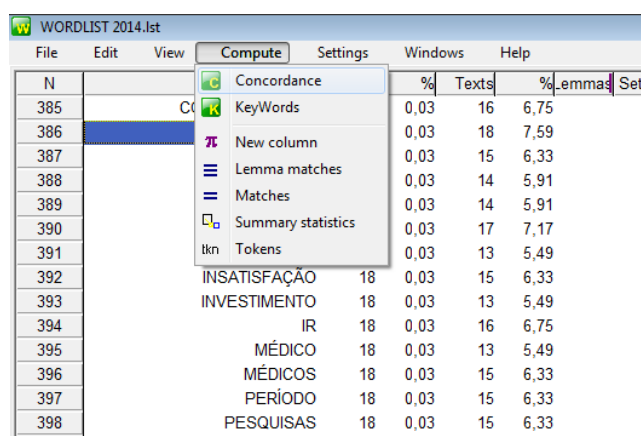
Fonte: Elaboração própria

Até o momento, constatamos que o *corpus* possui 227 elementos coesivos sequenciais diferentes. Por enquanto, o *corpus* ainda está em fase de análise, e, em virtude disso, ainda consideramos a possibilidade de o número de elementos coesivos sofrer alterações, pois, cada vez que interagimos com o *corpus*, podemos descobrir e observar palavras que passaram despercebidas até então.

#### 2.4. Procedimentos para verificação dos usos inadequados dos elementos coesivos sequenciais

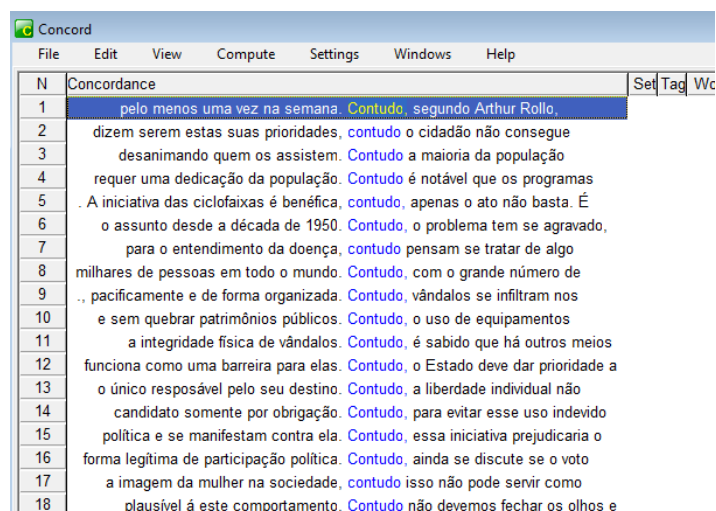
Para a identificação dos elementos coesivos sequenciais usados de maneira inadequada pelos internautas que enviaram sua redação ao site UOL Educação, foi necessário verificar os contextos linguísticos dos 227 elementos. Essa também foi uma atividade extremamente trabalhosa e que exigiu tempo, uma vez que o elemento coesivo sequencial de maior frequência “e” teve 1709 ocorrências, por exemplo. No entanto, verificar tudo isso, manualmente, com certeza se tornaria mais cansativo e com maior probabilidade de falhas do que com um programa de análise lexical específico para esse tipo de trabalho.

Diante disso, após localizar na *WordList* o elemento coesivo sequencial, clicamos na aba “Compute” e, em seguida, em *Concordance*, como mostra a Figura 8:



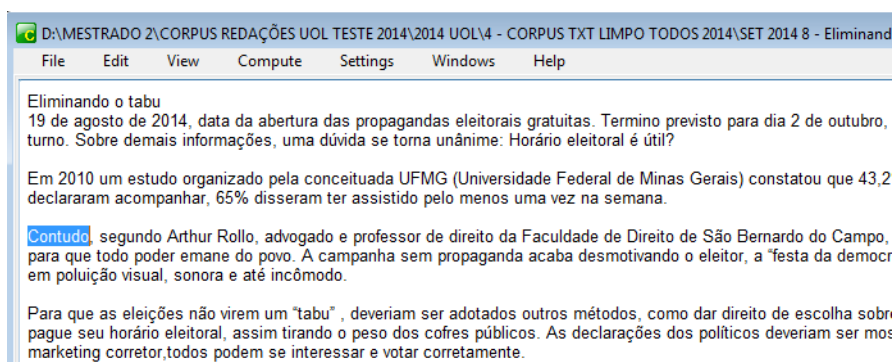
**Figura 8. Um caminho para as linhas de concordância**  
**Fonte: Elaboração própria**

A ferramenta *Concord* nos mostra as linhas de concordância (*Concordance*) da palavra em questão. Isso significa que ela nos permite visualizar especificamente todos as vezes que uma palavra foi usada, conforme ilustra a Figura 9:



**Figura 9. Imagem parcial das linhas de concordância do “Contudo”**  
**Fonte: Elaboração própria**

E, com um clique duplo em cima de cada linha de concordância, conseguimos ver e analisar o contexto linguístico em que a palavra foi usada, como no exemplo da Figura 10 a seguir:



**Figura 10. Contexto linguístico do “Contudo”**

**Fonte: Elaboração própria**

Com esse procedimento, identificamos que, dos 227 elementos coesivos sequenciais presentes no *corpus* de estudo, 33 apresentaram usos inadequados em diferentes proporções. Tais resultados foram colocados em um quadro em documento *Word* que contém as seguintes informações: elemento coesivo sequencial; frequência no *corpus* de estudo; número de inadequações e porcentagem de inadequação. A seguir, mostraremos a análise dos usos inadequados de alguns desses elementos.

### 3. Análises

O critério semântico conduziu a análise dos elementos coesivos sequenciais usados de maneira inadequada. Consideramos como inadequado os usos que de fato trouxeram problema para a compreensão de determinada parte do texto. Apresentamos, a seguir, alguns excertos do *corpus* de redações. Todos foram copiados na íntegra, sem correções gramaticais.

#### 3.1. Conhecendo usos inadequados do elemento coesivo sequencial “contudo”

(01) As propagandas eleitorais tem como intenção informar e fazer os eleitores se interessarem pelos candidatos, porém é necessário que haja dedicação da população.

O intuito dos partidos em formular as propagandas pode até ser de atrair a atenção populacional, fazendo-as com base em pesquisas e estatísticas para atingir a região em que seus candidatos tem mais chances de voto, e é esse tipo de pensamento que faz uma parte da população recusar a assistir tais programações.

Entretanto mesmo que essa seja a realidade é possível a utilização do meio de comunicação sendo essa a televisão ou o rádio para uma escolha mais certa do candidato, já que as propostas de mudança e realizações são apresentadas nos programas é de interesse do eleitor se certificar se realmente é possível que tais propostas aconteçam e qual o histórico do candidato no meio político, e isso requer uma dedicação da população.

**Contudo** é notável que os programas políticos são de certa maneira entediante e uma perda de tempo, porém qual perda será maior a de minutos, horas ou anos?

Colocamos o texto na íntegra para evidenciar que se trata de uma produção sem conclusão ou com uma conclusão malfeita. Na primeira hipótese, texto sem conclusão, o “contudo” estaria inadequado, porque o leitor não poderia contrariar (como fez) a ideia de que as propagandas têm seu valor, já que ele constrói seu texto a favor delas.

Nesse caso, ele deveria modalizar sua escrita e substituir o “contudo” pelo “embora”: “Embora possamos reconhecer que os programas políticos são de certa maneira entediante, eles devem ser valorizados por nós, para que não tenhamos arrependimentos que irão durar anos”.

Na segunda hipótese, conclusão malfeita, o autor teria de introduzir um elemento conclusivo, “portanto”, “diante disso”, “assim”, por exemplo, e não deveria acrescentar ideias novas “entediante e perda de tempo” que, além de desfavorecerem a argumentação construída do texto em questão, são indevidas por se tratar do fechamento do texto, que deve retomar as principais ideias já mencionadas e, no caso do Enem, propor soluções.

(02) É inaceitável o modo como age a policia, além do mais a capacidade da segurança social vem surpreendendo e demonstrando que os objetivos para manutenção da ordem pública não são tão eficientes e a ação dos profissionais mediante as situações críticas é na maioria das vezes resolvidas de forma inadequada, na qual civis são retaliados covardemente por homens armados, sendo que seus deveres também giram em torno de proteger a integridade física de vândalos.

**Contudo**, é sabido que há outros meios de parar a massa delituosa, formas de contenção que não irá agredir a integridade física de culpados ou inocentes, pois, a partir, do momento em que uma bomba de efeito moral ou spray de pimenta são lançadas, seus efeitos podem atingir outrem e até mesmo quem não tem nada a ver com a situação e acabam ficando asfixiados com a entrada de gás lacrimogêneo em suas casas. esses profissionais incapacitados adequadamente para esse tipo de situação.

O parágrafo que se inicia com “Contudo” acrescenta argumentos que vão na mesma direção dos argumentos usados no parágrafo anterior, por isso o “contudo” deveria ser substituído por um elemento de coesão de adição: ademais, também, ainda, entre outros.

### 3.2. Conhecendo usos inadequados do elemento coesivo sequencial “para tal efeito”

(01) Desde o período de busca por conquista de territórios, povos antigos chegavam às novas terras preparados para o ataque. Além da escravidão conferida aos povos conquistados, e do acúmulo de riquezas, os novos senhores de terra mantinham relações sexuais constantes e forçadas com as mulheres nativas dessas regiões. Prática esta que não era vista como crime, nem muito menos, penalizada. **Para tal efeito**, a cultura machista veio sendo proferida e até os dias de hoje ainda há dúvida sobre: Se o estupro é um crime, quem é o culpado?

O elemento coesivo “para tal efeito” deve ser utilizado quando anteriormente foi dito uma situação ou condição que se espera alcançar, que é idealizada, almejada. Com a introdução desse elemento, espera-se que o escritor coloque, em seguida, o que deve ser feito para se chegar ao que foi estabelecido como objetivo ou meta a ser alcançada. “Para tal efeito” é um elemento de coesão que nenhum estudioso lido por nós trabalhou; e o entendemos nos seguintes sentidos: “para isso”; “para que isso aconteça” e “para que se alcance tais resultados, é preciso”. No trecho acima, o elemento foi usado de maneira inadequada, porque anteriormente o autor criticou o abuso que as mulheres negras e escravas sofriam na época da escravidão e, em seguida, o autor fala que o resultado disso foi a repercussão do machismo. Se entendemos como resultado,

consequência, não cabe o elemento “para tal efeito” entre os períodos, e sim, por exemplo: “**uma das consequências disso**”, “**em virtude disso e de outros fatores**” etc.

(02) Portanto, a partir do que foi explanado acima, pode-se perceber que a "cultura do estupro", onde tal crime é visto como normal e as suas vítimas são vistas como culpadas, não é apenas presenciado na sociedade brasileira, pois, a educação machista que é incorporada pela maioria das famílias desde os tempos mais remotos, tira a culpa do sexo masculino na sua maioria dos casos. **Para tal efeito** é necessário que sejam criadas novas leis, cada vez mais rígidas, de fiscalização e penalização para os praticantes dessa violência. Campanhas na mídia também poderiam ser de grande valia, para esclarecer para a população que a causa do estupro não é a vítima e sim, o próprio estuprador.

O autor usa “para tal efeito” como se fosse descrever, em seguida, o que deveria ser feito para alcançar o que foi dito antes, no entanto, a sua opinião anterior é uma crítica em relação ao fato de que educação machista ainda existe. Portanto, ele deveria usar, no lugar de “para tal efeito”, “para que o machismo seja amenizado”, “para que uma cultura da igualdade entre os sexos seja propagada e o machismo extinto”.

### 3.3. Conhecendo usos inadequados do elemento coesivo sequencial “ou seja”

(01) O primeiro ambiente que a criança convive é sua casa, **ou seja**, família, vizinhos, igreja.

O elemento “ou seja” pode ser usado para esclarecer o que foi dito anteriormente. Notamos que na frase acima ele é usado com sentido de explicar melhor o que seria o ambiente “casa”. No entanto, isso não ocorre, porque não consideramos que igreja e vizinhos sejam referências que ampliem a noção de casa; elas apenas complementam o círculo social de convivência que uma criança pode ter. Por isso, sugerimos a reelaboração do trecho: “Os primeiros ambientes nos quais uma criança convive são sua casa, ou seja, seu lar, com sua família, a casa dos vizinhos e a igreja”.

(02) Enfim, o aumento dos ciclistas poderiam ser maior, se houvesse mais investimentos de segurança, deve haver mais umas guardas, **ou seja**, colocar as placas de alertas.

Na mesma esteira da análise anterior, o “ou seja” nesse contexto linguístico também está inadequado, porque “colocar as placas de alertas” não esclarece ao leitor o trecho anterior “deve haver mais umas guardas”. A relação de sentido não se constitui ao usar tal elemento de ligação entre as orações. Seguimos com uma sugestão de reelaboração que introduz o elemento de exemplificação: “Enfim, o aumento dos ciclistas poderia ser maior, se houvesse mais investimentos de segurança, **tais como**: mais guardas e placas de alertas nas ruas”.

### 3.4. Elementos coesivos: a análise da frequência x inadequações

Identificamos a porcentagem de erros em relação à frequência, cálculo imprescindível para dizer qual elemento coesivo foi usado mais vezes de maneira inadequada no *corpus*.

Os elementos coesivos mais frequentes não são aqueles que apresentam maior porcentagem de usos inadequados. Talvez, isso seja previsível, mas confirmamos que os escritores internautas usam mais aquelas palavras que conhecem, ou seja, que sabem usar no contexto linguístico adequado. Por exemplo: "*porém*" (frequência 67; uso inadequado: 3; porcentagem de inadequação: 4,4%) e "*contudo*" (frequência 18; uso inadequado 5; porcentagem de inadequação: 27,7%).

Os elementos que apresentaram maior tendência de serem usados de maneira inadequada foram os que tiveram menor ocorrência e maior número de inadequação. Abaixo mostraremos os 10 primeiros:

Tabela 1: Porcentagem de inadequação dos elementos

<i>ELEMENTOS</i>	<i>FREQ.</i>	<i>Nº DE USOS INADEQUADOS</i>	<i>%</i>
1. Em conseqüente disso	1	1	100,0
2. Para tal efeito	2	2	100,0
3. Sobre demais informações	1	1	100,0
4. Apesar disso	2	1	50,0
5. Por conseguinte	3	1	33,3
6. Contudo	18	5	27,7
7. Porque	39	8	20,5
8. Ou seja	21	4	19,0
9. Por sua vez	6	1	16,6
10. Conseqüentemente 12/ Conseqüente 1	13	2	15,3

Fonte: Elaboração própria

### Comentários Finais

Podemos dizer que a realização desse estudo piloto pelo viés da Linguística de *Corpus* foi satisfatória, na medida em que conseguimos atingir o objetivo de identificar quais elementos coesivos foram usados de maneira inadequada nas redações dissertativas argumentativas estilo Enem, num recorte de 237 redações do nosso *corpus* de estudo. A metodologia baseada na Linguística de *Corpus* facilitou sobremaneira um trabalho que, se fosse feito sem um *corpus* eletrônico e sem o auxílio de um programa de análise lexical como o *WordSmith Tools*, com certeza teria um *corpus* de estudo reduzido e haveria uma margem de erro muito maior ao contabilizarmos manualmente a frequência de uso de uma palavra.

Destacamos também que o estudo piloto demonstrado neste trabalho foi de grande importância, visto que nos permitiu visualizar uma amostra do que será a pesquisa de mestrado em proporções maiores. Assim, temos uma previsão das possibilidades de um elemento coesivo ser recorrente ou não em um *corpus* maior e das probabilidades de ocorrer inadequações em relação ao uso deles. E essas descobertas, certamente, nos dizem como devemos direcionar nossa pesquisa.

Uma questão fundamental que o estudo piloto nos mostrou foi que identificar os elementos coesivos sequenciais existentes e verificar o uso inadequado deles no *corpus* de redações demandou muito tempo. Em virtude disso, não é viável que esse procedimento seja refeito em um *corpus* com uma quantidade de textos aproximadamente 5 vezes maior, conforme tínhamos planejado inicialmente. Portanto, acreditamos que devemos trabalhar daqui em diante apenas com os 33 elementos coesivos que foram usados de forma inadequada no estudo piloto, para que possamos cumprir os objetivos propostos na nossa dissertação de mestrado.

## REFERÊNCIAS

- ANTUNES, I. *Lutar com palavras – coesão e coerência*. São Paulo: Parábola, 2005. 199 p.
- BERBER SARDINHA, T. *Linguística de Corpus*. Barueri, São Paulo: Manole, 2004. 410 p.
- \_\_\_\_\_. *Pesquisa em Linguística de Corpus com WordSmith Tools*. Campinas: Mercado das Letras, 2009. 299 p.
- FROMM, G. O uso de *corpora* na análise linguística. *Factus*, São Paulo, v.1, n.1, p. 69-76, 2003.
- GRAMA, D. F. Problemas de coesão na escrita dos gêneros discursivos da ordem do relatar: ‘notícia e relato’. *Horizonte Científico*, Uberlândia, v.7, n.1, p. 1-30, set. 2013. Disponível em: <<http://www.seer.ufu.br/index.php/horizontecientifico/article/view/17790/12804>>. Acesso em: 24 mar. 2016.
- KOCH, I. G. V. *A coesão textual*. 21. ed. São Paulo: Contexto, 2008. 84 p.
- \_\_\_\_\_. *Argumentação e linguagem*. 13. ed. São Paulo: Cortez, 2011. 239 p.
- PARODI, G. *Linguística de Corpus: de la teoría a la empiria*. Madrid: Iberoamericana, 2010. p. 13-35.
- PAULIUKONIS, M. A. L. Conectores de oposição: reflexões e propostas para o ensino. *Gragoatá*, Niterói, v.19, n.36, p. 28-42, 1 sem. 2014. Disponível em: <<http://www.uff.br/revistagragoata/ojs/index.php/gragoata/article/view/27>>. Acesso em: 17 set. 2014.
- SCOTT, M. *WordSmith Tools version 5*. Liverpool: Lexical Analysis Software, 2008.
- SHEPHERD, T. M. G. Panorama da Linguística de Corpus. In: SHEPHERD, T. M. G. et al. (orgs). *Caminhos da Linguística de Corpus*. Campinas: Mercado das Letras, 2012. p. 15-30.
- TAGNIN, S. E. O. Linguística de *Corpus* e Fraseologia: Uma feita para a outra. In: ORTIZ, M. L. A.; UNTERNBAUMEN, E. H. (orgs.). *Uma (re)visão da teoria e da pesquisa fraseológicas*. Campinas: Pontes, 2011. p. 227-302.

**Recebido em:** 11/09/2015

**Aprovado em:** 02/04/2016