

Uma Proposta de Revisor para Identificação de Desvios de Linguagem em Textos Descritivos Formais

Jonathan Nau¹, Aluizio Haendchen Filho¹, Guilherme Passero^{1,2}

¹Núcleo de Inteligência Artificial e Sistemas Inteligentes (NIASI)
Centro Universitário de Brusque (UNIFEBE) – Brusque, SC – Brasil

²Laboratório de Inteligência Aplicada (LIA)
Universidade do Vale do Itajaí (UNIVALI) – Itajaí, SC – Brasil

jonathan_nau@live.com, aluizioh@terra.com.br,
guilherme.passero@gmail.com

***Abstract.** The revision of formal texts is a complex task and occurs in several areas. The objective of this work is to create a tool to support the revision of texts and promote studies in automatic correction of descriptive texts. We propose a reviewer for automatic identification of language deviations in formal descriptive texts, using natural language processing techniques to achieve these goals.*

1. Introdução

Ao contrário das figuras de linguagem, que representam realce e beleza às mensagens emitidas, os desvios de linguagem são palavras ou construções que ferem as normas gramaticais [Leal 2012, Pliger 2009]. Os desvios de linguagem costumam ocorrer por descuido, ou ainda por desconhecimento das regras por parte do emissor. De acordo com Rino et al. (2014), o público-alvo dos revisores gramaticais tem se revelado insatisfeito com a restrição de intervenção aos problemas ortográfico-sintáticos, exigindo a consideração de problemas relacionados ao estilo, muito mais vinculados à eficácia comunicativa do que a simples adequação à norma gramatical.

Nesse contexto, este trabalho apresenta uma proposta de um revisor para identificação de desvios de linguagem em textos descritivos formais. Uma ferramenta com tal finalidade pode ter várias aplicações, tanto no âmbito acadêmico quanto empresarial, desde sistemas de correção automática de redações até revisão de artigos e documentos. Em pesquisas na literatura, não identificamos ferramentas similares que abrangem todos os tipos de desvio considerados nesta pesquisa.

2. Vícios e Desvios de Linguagem

Dentre os problemas relacionados ao desvio de linguagem, podem ser citados [Pinheiro 2007]: o uso de clichês/chavões, o emprego de marcas de oralidade, a repetição exagerada de palavras e os vícios de linguagem. Os vícios de linguagem citados na literatura [Leal 2012, Pliger 2009] abrangem os seguintes tipos: (i) ambiguidade: uso de palavras com duplo sentido; (ii) arcaísmo: abrange expressões que caíram em desuso; (iii) barbarismo: emprego desnecessário de palavras estrangeiras; (iv) cacófato: abrange junção de palavras que resultam em som desagradável ou obsceno; (v) colisão: aproximação de sons consonantais idênticos ou semelhantes; (vi) eco: repetição

desagradável de terminações iguais; (vii) hiato: aproximação de vogais idênticas; (viii) plebeísmo: abrange qualquer desvio que caracteriza a falta de instrução (p. ex. gíria); (ix) preciosismo: uso excessivo de palavras para exprimir ideias simples; e (x) solecismo: desvio em relação à sintaxe.

3. Proposta de Solução

A Tabela 1 mostra o escopo da proposta de solução e as técnicas empregadas para cada tipo de desvio de linguagem. Conforme a tabela, desvios como ambiguidade e preciosismo não serão abrangidos pela proposta, visto que exigem a consideração de problemas de natureza semântico-pragmática, relacionados a níveis maiores do que a oração [Rino 2014]. O solecismo é um tipo de desvio já resolvido por alguns corretores gramaticais, p. ex. o ReGra [Rino 2014] utilizado pelo Microsoft Word. Os procedimentos adotados para a coleta, análise e desenvolvimento da solução estão divididos em duas partes: (i) construção do corpus e (ii) aplicação das técnicas.

Tabela 1. Proposta para Identificação de Desvios de Linguagem

Tipo	Proposta de solução	Corpus		Técnica		
		Redação	Catálogo	Stopword	Lematização	N-gramas
Vícios de linguagem	Ambiguidade	Não	-	-	-	-
	Arcaísmos	Sim	x	x	x	x
	Barbarismo	Sim	x	x	-	x
	Cacófato	Sim	x	x	-	x
	Colisão	Sim	x	-	-	x
	Eco	Sim	x	-	-	x
	Hiato	Sim	x	-	-	x
	Plebeísmo	Sim	x	x	-	x
	Pleonasmo	Sim	x	x	-	x
	Preciosismo	Não	-	-	-	-
Solecismo	Não	-	-	-	-	
Clichês/chavões	Sim	x	x	x	x	x
Marcas de oralidade	Sim	x	x	-	-	x
Repetição exagerada de palavras	Sim	x	-	x	-	x

Para a construção do corpus de redações, foram obtidas 2.000 redações do Banco de Redações UOL [UOL Educação 2016]. Os catálogos são listas de palavras e expressões necessárias para detecção dos tipos de desvio relacionados que serão elaborados com auxílio de especialistas da língua portuguesa. As técnicas são aplicadas em duas fases: no pré-processamento e no processamento. No pré-processamento, o algoritmo transforma nos corpora todas as palavras para minúscula e realiza a remoção dos acentos. Na etapa de processamento são aplicadas técnicas de *stopwords* e lematização para a análise de alguns tipos de desvio e n-gramas para todos, e um conjunto de procedimentos específicos para cada tipo.

3. Considerações Finais

Neste artigo foi proposta a construção de um revisor para identificação de desvios de linguagem em textos descritivos formais. Por meio da delimitação dos tipos de desvios, das técnicas de pré-processamento e processamento, foi projetada uma solução que preenche uma lacuna existente no mercado.

Referências

Leal, E. S. (2012) Vícios de Linguagem e Idiotismos: A Fala Como Unidade de Estudos nas Gramáticas Normativas Brasileiras em Língua Portuguesa entre 1881-1890. Páginas 255-265. Revista Uniletras. 2012.

Pinheiro, G.M. (2007) “Redações do ENEM: estudo dos desvios da norma padrão sob a perspectiva de corpus”. Dissertação apresentada à Faculdade de Filosofia, Letras e Ciências Humanas da Universidade de São Paulo. São Paulo, 2007.

Pliger, D. B. (2009) “Vícios de Linguagem e a Norma Culta”. Trabalho de Conclusão de Curso apresentado à Faculdade de Educação São Luís. São Paulo, 2009.

Rino, L.H.M. et al. (2002) Aspectos da Construção de um Revisor Gramatical Automático para o Português. Revista Estudos Linguísticos. 2002.

UOL Educação (2016) “Banco de Redações”.

<https://educacao.uol.com.br/bancoderedacoes/>, Dezembro, 2016.