

James Alves da Silva Junior

Um Avaliador Automático de Redações

Vitória, ES

2021

James Alves da Silva Junior

Um Avaliador Automático de Redações

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Informática da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do Grau de Mestre em Informática.

Universidade Federal do Espírito Santo – UFES

Centro Tecnológico

Programa de Pós-Graduação em Informática

Orientador: Claudine Badue

Coorientador: Elias Silva de Oliveira

Vitória, ES

2021

James Alves da Silva Junior

Um Avaliador Automático de Redações / James Alves da Silva Junior . –
Vitória, ES, 2021-

82 p. : il. (algumas color.) ; 30 cm.

Orientador: Claudine Badue

Coorientador: Elias Silva de Oliveira

Dissertação de Mestrado – Universidade Federal do Espírito Santo – UFES

Centro Tecnológico

Programa de Pós-Graduação em Informática, 2021.

1. Avaliação Automática de Redação. 2. Reconhecimento de Entidades
Nomeadas. 3. Processamento de linguagem natural (Computação). I. Badue,
Claudine. II. Oliveira, Elias Silva. III. Universidade Federal do Espírito Santo. IV.
Um Avaliador Automático de Redações

CDU 02:141:005.7

James Alves da Silva Junior

Um Avaliador Automático de Redações

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Informática da Universidade Federal do Espírito Santo, como requisito parcial para obtenção do Grau de Mestre em Informática.

Trabalho aprovado. Vitória, ES, 25 de setembro de 2014:

Claudine Badue
Orientador

Elias Silva de Oliveira
Co-orientador

Patrick Marques Ciarelli
Convidado 1

Thiago Oliveira dos Santos
Convidado 2

Vitória, ES
2021

*Aos meus pais, James e Dulcinéia.
E a minha amada esposa Solange.*

Agradecimentos

Primeiramente a Deus, porque d'Ele e por meio d'Ele todas as coisas são realizadas. A minha amada esposa Solange meu alicerce e motivação para continuar a alçar novos voos.

Aos meus pais James e Dulcineia que foram infinitamente pacientes, mesmo nas minhas ausências, durante toda a jornada. E aos meus queridos irmãos Fabiano e André pelo seu companheirismo.

Agradeço aos professores Claudine e Elias por todo apoio e paciência durante toda a pesquisa e por todo conhecimento compartilhado que trouxeram maturidade durante todo o processo.

Aos Membros da Banca Examinadora pela participação na defesa deste trabalho e pelas valiosas contribuições.

Em especial, agradeço aos colegas do LCAD, que muito me ajudaram no desenvolvimento deste trabalho.

Muito Obrigado!

*“Agora existe um estado de fluxo permanente
...o fato é que atualmente, todo ano é revolucionário
(Yuval Noah Harari)*

Resumo

No Brasil para o ingresso no ensino superior é necessário passar pelo Exame Nacional do Ensino Médio (ENEM) realizado pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). A nota final do exame interfere diretamente na vida do estudante, uma vez que a nota final do exame possibilita o ingresso em instituições de Ensino Superior. A nota também interfere na concessão de financiamento estudantil. Do ponto de vista do aluno uma nota baixa é uma perda de oportunidades.

A redação é parte importante da composição da nota do exame, uma vez que se a nota da redação for zerada o aluno será automaticamente eliminado do Sistema de Seleção Unificada (SISU) para ingresso em instituições públicas de ensino superior e também do processo seletivo de crédito estudantil.

Tendo em vista a importância da redação na composição da nota final exploramos nesse trabalho duas Competências que são avaliadas na redação do modelo ENEM/INEP. Buscamos criar um modelo de Avaliação Automática de Redações (AES) que tem como objetivo imitar avaliadores humanos ao darem notas em redações. Esta área existe a aproximadamente 50 anos e na língua inglesa apresenta avanços consistentes em avaliações em exames de segunda língua.

O sistema de correção automática de redação é pouco adotado na língua portuguesa no apoio ao aprendizado e até mesmo como sistema avaliativo. Tem crescido a pesquisa em torno do tema uma vez que existe a intenção de tornar a avaliação do ENEM completamente informatizada. Outra discussão sobre o tema de redação é sobre a dicotomia entre os avaliadores no momento de avaliar as redações e também como o modelo pode dar *feedbacks*.

Palavras-chaves: avaliação automática de redações, processamento de linguagem natural, ENEM.

Abstract

In Brazil, admission to higher education is necessary to pass the Exame Nacional do Ensino Médio (ENEM) conducted by the Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). A final grade of the exam directly interferes with student life, since the final grade of the exam makes it possible to enter higher education institutions. A note also interferes with student funding. The student's point of view a low grade is the loss of opportunities.

Writing is an important part of the composition of the exam score, since the writing score for students with zero will be automatically eliminated by the Sistema de Seleção Unificada (SISU) for admission to public institutions of higher education and also with the student credit process. .

In view of the importance of writing in the composition of the final grade, explore two Competencies that are evaluated in the writing of the ENEM/INEP model. This work seeks to create an Automatic Essay Evaluation model (AES) that aims to imitate human evaluators and give grades in essays. This area has been around for 50 years and the English language has made many consistent advances in estimates with TOEL, Cambrige.

The automatic writing correction system is not widely adopted in Portuguese without support for learning and even as an evaluation system. Research on the topic has grown since there is an intention to make the assessment of ENEM completely computerized. Other discussions on the subject of writing are about a dichotomy between the evaluators when evaluating the essays and also how the model can provide feedbacks to improve student performance when writing as essays.

Keywords: automation essay score. natural language processing. ENEM.

Lista de ilustrações

Figura 1 – Exemplo do resultado da execução do NER (PIROVANI, 2019)	32
Figura 2 – Arquitetura do sistema	41
Figura 3 – Representação da organização das características do conjunto de redações	42
Figura 4 – Exemplo da organização das características de duas redações.	43
Figura 5 – Características extraídas pelo AES	43
Figura 6 – Pipeline de treinamento e avaliação	47
Figura 7 – Exemplo de regressor linear para uma característica.	51
Figura 8 – Three simple graphs	52
Figura 9 – Proporção de notas pela quantidade de erros ortográficos no texto. . .	55
Figura 10 – Resultado dos experimentos propostos para Competência I	56
Figura 11 – Resultado dos experimentos propostos para Competência III	57

Lista de tabelas

Tabela 1 – Níveis de avaliação da Competência I	25
Tabela 2 – Níveis de avaliação da Competência III	26
Tabela 3 – Estrutura exigida da redação (INEP, 2018)	30
Tabela 4 – Resumo do estado da arte sobre AES	38
Tabela 5 – Resumo das Características de linguística e conteúdo implementadas no sistema	45
Tabela 6 – Resumo dos Experimentos Realizados	54
Tabela 7 – Resumo dos Experimentos Realizados	54
Tabela 8 – Propriedades do conjunto de redação para a Competência I	55
Tabela 9 – Propriedades do conjunto de redação para a Competência III	57
Tabela 10 – Comparação do Resultado com a métrica MAE para a Competência 1	58
Tabela 11 – Comparação do Resultado com a métrica QWK para a Competência 1	58
Tabela 12 – Comparação do Resultado com a métrica QWK para a Competência 3	58

Lista de abreviaturas e siglas

AES	<i>Automarion Essay Scoring</i>
ENEM	Exame Nacional do Ensino Médio
GA	<i>Genetic Algorithm</i>
INEP	Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira
NLP	<i>Natural Language Process</i>
NER	<i>Named Entity Recognition</i>

Sumário

1	INTRODUÇÃO	23
1.1	Contexto	23
1.2	Motivação	24
1.3	Objetivos	26
1.4	Contribuições	27
1.5	Metodologia	27
1.6	Estrutura do Trabalho	27
2	CONCEITOS BÁSICOS	29
2.1	Texto Dissertativo Argumentativo	29
2.2	Processamento de linguagem natural	30
2.3	Extração de Entidade Nomeadas (NER)	31
2.4	Métricas	32
2.4.1	Kappa Quadrático Ponderado (QWK)	32
2.4.2	<i>Mean Absolute Error</i> (MAE)	33
2.4.3	Avaliação de Juízes	33
3	TRABALHOS RELACIONADOS	35
3.1	Avaliação Automática de Redações (AES)	35
3.2	AES Aplicado a avaliação ENEM/INEP	36
3.3	Reconhecimento de Entidade Nomeadas (NER)	39
3.4	Avaliação de Juízes	39
4	ARQUITETURA DO SISTEMA	41
4.1	Representação Vetorial	42
4.2	Engenharia de Características	43
4.2.1	Características Linguísticas	44
4.2.2	Conteúdos da Redação	44
4.3	Avaliação de Redação	46
4.3.1	Seleção de características	49
4.3.2	Avaliação de modelo com <i>Cross-Validation</i>	50
4.4	Métodos de Aprendizado Supervisionado	50
4.4.1	Regressão Linear	51
4.4.2	K-Nearest Neighbor (KNN)	51
5	EXPERIMENTOS E RESULTADOS	53

5.1	Base de Dados	53
5.2	Experimentos Realizados	53
5.3	Experimentos para Avaliação da Competência I	54
5.4	Experimentos para Avaliação da Competência III	56
5.5	Comparação com Abordagens Apresentadas na Literatura	57
6	CONCLUSÕES E TRABALHOS FUTUROS	59
6.1	Conclusão	59
6.2	Trabalhos Futuros	59
	REFERÊNCIAS	61
	APÊNDICES	65
	APÊNDICE A – REGRAS GRAMATICAIIS DO <i>COGROO</i>	67

1 Introdução

Este capítulo proporciona ao leitor uma apresentação do contexto, motivação e objetivos deste trabalho, os métodos de pesquisa aplicados e a estrutura textual para dar ao leitor compreensão do processo de automatização para redução do esforço da correção da redação do Exame Nacional do Ensino Médio (ENEM), exame promovido pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), para a Competência I - *domínio da norma padrão da língua portuguesa*, e também para a Competência III - *selecionar relacionar, organizar e interpretar informações, fatos, opiniões e argumentos em defesa de um ponto de vista*.

1.1 Contexto

A redação é um dos itens avaliativos do ENEM/INEP e seu resultado é importante para o estudante pois com base em seu resultado final o aluno avaliado poderá ingressar no ensino superior em universidades públicas federais ou também em instituições privadas, além da obtenção de bolsas e programas de intercâmbios financiados por programas público de financiamento. A redação representa 20% da nota final do exame e ainda exige-se uma pontuação mínima de 450 pontos para ingresso em programas de financiamento estudantil.

Os critérios da correção da prova de redação são divididas em 5 Competências (INEP, 2018), e são elas:

- I demonstrar domínio da modalidade escrita formal da língua portuguesa;
- II compreender a proposta de redação e aplicar conceitos das várias áreas de conhecimento para desenvolver o tema, dentro dos limites estruturais do texto dissertativo-argumentativo em prosa;
- III selecionar, relacionar, organizar e interpretar informações, fatos, opiniões e argumentos em defesa de um ponto de vista;
- IV demonstrar conhecimento dos mecanismos linguísticos necessários para a construção da argumentação;
- V elaborar proposta de intervenção para o problema abordado, respeitando os direitos humanos.

Para cada Competência a nota atribuída será entre 0 e 200 pontos com intervalos de 40 pontos, totalizando 5 notas possíveis para cada Competência. O método de correção

da redação consiste na soma das médias entre dois avaliadores para cada Competência. Quando houver uma diferença de 80 pontos nas notas atribuídas entre qualquer uma das 5 Competências, ou uma diferença superior a 100 pontos da nota final a redação será avaliada novamente. A reavaliação é feita por um terceiro avaliador e a nota final será a média entre as notas dos avaliadores que mais se aproximarem. Persistindo a diferença de 80 pontos em alguma Competência ou 100 pontos na nota final entre os três avaliadores uma nova banca é convocada para avaliar a redação.

Ainda sobre o método de avaliação da redação de acordo com o guia do participante (INEP, 2018), o avaliador pode imputar nota zero a uma redação caso ela apresente as seguintes características: (i) fuga total ao tema, (ii) não obediência a estrutura do texto dissertativo-argumentativo; (iii) texto com até sete linhas, (iv) textos impróprios, desenhos e outras formas que são propositalmente desconexas da proposta do texto; (v) desrespeito aos direitos humanos e (vi) folha de resposta da redação em branco.

Por fim escolhemos duas Competências que serão foco desse trabalho, a começar pela Competência I que visa avaliar se o estudante domina a modalidade da escrita formal da língua portuguesa e também a fluidez da leitura do texto. A Tabela 1 mostra os níveis de desempenho que serão utilizados para avaliar a redação na Competência I segundo o manual do participante do ENEM/INEP (INEP, 2018).

A Tabela 2 apresenta os níveis de desempenho para a Competência III. Nessa Competência é esperado que o participante do ENEM/INEP apresente os seguintes elementos: informações, fatos, opiniões e argumentos. E com esses elementos defender o ponto de vista escolhido como tese do texto dissertativo-argumentativo. Avaliando como o estudante seleciona, relaciona, organiza e interpreta estes elementos no texto (INEP, 2018).

1.2 Motivação

A correção de redações se mostra um processo custoso financeiramente. Além disso, o elevado número de redações, crescente a cada edição do ENEM/INEP, também impacta no volume de trabalho dos avaliadores tornando o processo de correção exaustivo. Também há falta de padronização na concordância entre os avaliadores, que geram diferenças entre as notas atribuídas às redações pelos avaliadores.

No ano de 2016 foi noticiado o custo estimado da correção das redações por uma reportagem do Portal G1¹ para a avaliação do ano anterior. No ano de 2015 o custo estimado para a correção de cada redação foi R\$15.88 tendo sido corrigidas 6.54 milhões de redações, perfazendo um custo aproximado de R\$104 milhões.

¹ <https://g1.globo.com/educacao/enem/2016/noticia/corretores-de-redacao-do-enem-avaliam-em-media-74-redacoes-por-dia.ghml>

Tabela 1 – Níveis de avaliação da Competência I

Demonstra excelente domínio da modalidade escrita formal da língua portuguesa e de escolha de registro. Desvios gramaticais ou de convenções da escrita serão aceitos somente como excepcionalidade e quando não caracterizarem reincidência.	200 pontos
Demonstra bom domínio da modalidade escrita formal da língua portuguesa e de escolha de registro, com poucos desvios gramaticais e de convenções da escrita.	160 pontos
Demonstra domínio mediano da modalidade escrita formal da língua portuguesa e de escolha de registro, com alguns desvios gramaticais e de convenções da escrita.	120 pontos
Demonstra domínio insuficiente da modalidade escrita formal da língua portuguesa, com muitos desvios gramaticais, de escolha de registro e de convenções da escrita.	80 pontos
Demonstra domínio precário da modalidade escrita formal da língua portuguesa, de forma sistemática, com diversificados e frequentes desvios gramaticais, de escolha de registro e de convenções da escrita.	40 pontos
Demonstra desconhecimento da modalidade escrita formal da língua portuguesa.	0 pontos

A avaliação automática se tornará ao longo da próxima década uma ferramenta chave no aprendizado de escrita de redações uma vez visto que o Ministério da Educação (MEC) começa a implantar já no ENEM/INEP de 2020 a aplicação da prova totalmente eletrônica, com o atendimento a todo o território nacional a partir do ano de 2026².

Assim, motivado pelos problemas e custos levantados e também pelos desafios futuros da automatização da aplicação da avaliação de redações propomos uma arquitetura para automatizar o processo de avaliação reduzir os esforços dos avaliadores, e um modelo estatístico para avaliar as redações a partir de um conjunto de características extraídas do texto, mitigando falhas e omissão dos avaliadores humanos sobre as notas atribuídas às redações.

Além disso também abre ao estudante a possibilidade de praticar a escrita de redações sem a necessidade de submetê-las a um professor humano para avaliá-la. Estudantes com as maiores notas relataram a necessidade de treinar os fundamentos da gramática,

² http://portal.inep.gov.br/artigo/-/asset_publisher/B4AQV9zFY7Bv/content/exame-tera-aplicacao-digital-em-fase-piloto-em-2020-e-deixara-de-ter-versao-em-papel-em-2026/21206

Tabela 2 – Níveis de avaliação da Competência III

Apresenta informações, fatos e opiniões relacionados ao tema proposto, de forma consistente e organizada, configurando autoria, em defesa de um ponto de vista.	200 pontos
Apresenta informações, fatos e opiniões relacionados ao tema, de forma organizada, com indícios de autoria, em defesa de um ponto de vista.	160 pontos
Apresenta informações, fatos e opiniões relacionados ao tema, limitados aos argumentos dos textos motivadores e pouco organizados, em defesa de um ponto de vista.	120 pontos
Apresenta informações, fatos e opiniões relacionados ao tema, mas desorganizados ou contraditórios e limitados aos argumentos dos textos motivadores, em defesa de um ponto de vista.	80 pontos
Apresenta informações, fatos e opiniões pouco relacionados ao tema ou incoerentes e sem defesa de um ponto de vista.	40 pontos
Apresenta informações, fatos e opiniões não relacionados ao tema e sem defesa de um ponto de vista.	0 pontos

e também reforçam a necessidade de praticar diversos temas³, uma vez que o tema só é revelado no dia da prova.

1.3 Objetivos

O objetivo desse trabalho é analisar a aplicação de métodos de inteligência artificial aliados ao processamento de linguagem natural para automatização da avaliação de redações delimitado a Competência III - *Selecionar, relacionar, organizar e interpretar informações, fatos, opiniões e argumentos em defesa de um ponto de vista*, também melhorar os resultados obtidos por Júnior (2017) para a Competência I - *Demonstrar domínio da modalidade escrita formal da língua portuguesa*, critério avaliados segundo o ENEM/INEP.

Para além disso pretendemos:

- I. disponibilizar a base de dados utilizadas nesse trabalho contendo notas para as 5 Competências;

³ <https://www.agazeta.com.br/es/norte/capixaba-nota-1000-na-redacao-do-enem-da-dicas-para-fazer-bom-texto-0120>

- II. oferecer uma interface com avaliador humano sugerindo uma nota e mostrando os pontos de melhoria encontrados pelo modelo;
- III. disponibilizar um base de dados de redações anotadas com entidades nomeadas.

1.4 Contribuições

A principal contribuição deste trabalho é a análise automática de coesão textual em redações do ENEM/INEP, permitindo uma padronização do processo avaliativo e uma explicação ao avaliado do motivo da nota. Os experimentos realizados mostram que é possível a detecção dos seguintes casos nas redações:

- falta elementos textuais que demonstrem coesão no texto;
- evidenciamos a importância da citações de pessoas, organização e acontecimentos nas redações;
- a inclusão de um novo domínio nos Corpus para reconhecimento de entidades nomeadas.

1.5 Metodologia

Realizamos uma revisão na literatura, levantando os trabalhos relacionados com a pesquisa proposta nesse trabalho que é: avaliação automática de redações e a metrificação de coesão textual.

Após a realização da revisão propomos um modelo a partir da NLP de avaliar a redação a fim de identificar os erros e acertos concernentes a Competência III - selecionar, relacionar, organizar e interpretar informações, fatos, opiniões e argumentos em defesa de um ponto de vista. Aplicar esse modelo sobre um conjunto de redações já corrigidas por um especialista humano e extrair métricas sobre essas redações. Também propomos um modelo, com base no trabalho realizado por [Júnior \(2017\)](#), um sistema para correção da Competência I - apresentar domínio da norma padrão da língua portuguesa.

Sobre as métricas levantadas, aplicar modelos de aprendizagem de máquina e inteligência artificial para inferir uma nota dada a uma nova redação, a fim de comparar a nota aferida com avaliação de um especialista humano, e criticando quais métricas são mais importantes para o modelo a partir da análise por algoritmos genéticos.

1.6 Estrutura do Trabalho

Este trabalho está dividido nas seguintes Seções:

Capítulo 2 introdução ao conceitos fundamentais da correção automática de redações.

Capítulo 3 faz uma relação de trabalhos que se relacionam com o objetivo final do trabalho e traça uma comparação entre os modelos apresentados e o proposto.

Capítulo 4 formalização do método proposto de avaliação de coesão textual.

Capítulo 5 descrição dos experimentos e discussão dos resultados obtidos.

Capítulo 6 descrição da conclusão e dos trabalhos futuros.

2 Conceitos Básicos

Nesta seção veremos alguns conceitos básicos que serão usados no decorrer deste trabalho. Começamos pela apresentação da construção do texto da redação do ENEM/INEP, quais informações buscamos extrair e quais os fundamentos por trás desses elementos. Depois apresentamos alguns conceitos do processamento de linguagem natural e do reconhecimento de entidade nomeadas.

Por fim explicamos as métricas que usamos para avaliar a qualidade do modelo construído e comparar os resultados com outros trabalhos. Nosso objetivo não é fornecer uma cobertura exaustiva sobre os assuntos apresentados, apenas introduzir o leitor alguns argumentos e técnicas importantes utilizados nesse trabalho.

2.1 Texto Dissertativo Argumentativo

O texto dissertativo-argumentativo tem suas ideias fundamentadas por [Quintilianus e Halm \(1869\)](#) que lecionou gramática e retórica na Roma antiga. Ele postulou em 12 livros 5 cânones da retórica, que são: *Inuentio*, *Dispositio*, *Elocutio*, *Memoria* e *Pronuntitio*. Esses cânones são etapas para compor um texto dissertativo-argumentativo.

O primeiro cânone é a invenção (*Inuentio*) que baliza toda a ideia do texto e consiste na elaboração da estratégia argumentativa sobre o tema. O arranjo (*Dispositio*) é a organização lógica dos elementos da redação utilizada na argumentação, essa deverá ser a mais didática possível para tornar a tarefa de redigir muito mais fácil. O estilo (*Elocutio*) é considerado a estética literária associado a retórica e a estratégia de persuasão adotada pelo autor a fim de apoiar a argumentação do ponto de vista escolhido, para [Quintilianus e Halm \(1869\)](#) o estilo está diretamente relacionado aos diferentes gêneros literários já explorados na leitura do autor.

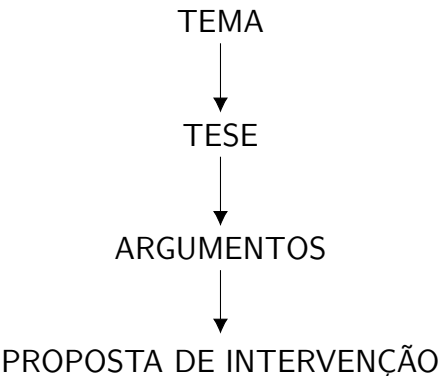
A memória (*Memoria*) é a capacidade de adquirir, armazenar e recuperar informações disponíveis. É necessário relacionar um repertório conhecimento prévio ao ponto de vista defendido em seu discurso, de acordo com o tema proposto na prova de redação do ENEM/INEP. Por fim a apresentação (*Pronuntitio*) o autor explicita sua tese com provas e argumentação para convencer e apresenta sua proposta de intervenção. Na redação esse cânone apresenta-se como uma boa caligrafia, em letras legíveis onde sua mensagem possa ser transmitida com clareza.

Daremos ênfase a presença de marcas textuais que demonstrem o uso da *Memoria*, uma vez que faz uma relação direta com a Competência III. E uma vez entendendo a constituição do texto dissertativo-argumentativo o manual de redação do ENEM/INEP

(INEP, 2018) determina a estrutura que a redação deva ser escrita.

Apresentamos na Tabela 3 a estrutura que é exigida pelo exame. A apresentação de *TEMA*, *TESE*, *ARGUMENTOS* e *PROPOSTA DE INTERVENÇÃO* tem uma forte relação com a apresentação com o cânone *Memoria* que vai para além do texto motivador, e requer do estudante a capacidade de relacionar e citar fatos correlatos que satisfaçam o que é pedido na proposta disposta na prova pela banca do ENEM/INEP. Por outro lado, o estudante também deve apresentar em seu texto uma boa ortografia, gramática e repertório vocabular para apresentar seus argumentos.

Tabela 3 – Estrutura exigida da redação (INEP, 2018)

<p>A prova de redação exigirá de você a produção de um texto em prosa, do tipo dissertativo-argumentativo, sobre um tema de ordem social, científica, cultural ou política. Os aspectos a serem avaliados relacionam-se às competências que devem ter sido desenvolvidas durante os anos de escolaridade. Nessa redação, você deverá defender uma tese – uma opinião a respeito do tema proposto –, apoiada em argumentos consistentes, estruturados com coerência e coesão, formando uma unidade textual. Seu texto deverá ser redigido de acordo com a modalidade escrita formal da língua portuguesa. Você também deverá elaborar uma proposta de intervenção social para o problema apresentado no desenvolvimento do texto que respeite os direitos humanos.</p>	 <pre> graph TD TEMA --> TESE TESE --> ARGUMENTOS ARGUMENTOS --> PROPOSTA_DE_INTERVENCAO[PROPOSTA DE INTERVENÇÃO] </pre>
---	--

Desde seu início a prova do ENEM/INEP aplica a avaliação de redação. No entanto a única modalidade textual adotada é a dissertativa-argumentativa. Essa exclusividade sistematizou e priorizou esse gênero textual, deixando de lado a prática de outros gêneros importantes para além da avaliação do ENEM/INEP.

2.2 Processamento de linguagem natural

Afim de extrair os elementos textuais esperados das redações usaremos o ferramental do processamento de linguagem natural (do inglês *Natural Language Process* - NLP) que é o ramo de pesquisa pelo qual procura-se fazer a inteligência artificial imitar o comportamento humano no processamento e geração de linguagem natural. Reconhecimento de regras, identificação de relações e funções das palavras são algumas observações realizadas por esses sistemas (PULMAN; SUKKARIEH, 2005). Devido a complexidade da compressão da linguagem natural os problemas são decompostos em pequenos sistemas tais como:

- Geração de Linguagem, como converter um dado processado para a linguagem humana;
- compreensão de linguagem, que é converter a linguagem natural em uma base de dados;
- tradução automática entre idiomas diferentes;
- análise de sentimento, determinando se um comentário é positivo ou negativo, por exemplo;
- verificação gramatical e ortográfica automática.

Cada um desses sistemas também tem um grau de complexidade considerável e em alguns casos compartilham tarefas semelhantes. Essas tarefas são decompostas em pequenas tarefas tidas como *de menor complexidade* tais como: *Tokenização*, detecção e segmentação de sentenças e etiquetagem, reconhecimento de entidades nomeadas, reconhecimento de resolução entre outros (SILVA, 2013).

As técnicas de NLP já são amplamente utilizadas na correção de redações vide o trabalho de Burstein et al. (2002) com o sistema *e-Rater* para compreender a adequação de textos escritos em testes de proficiência em língua estrangeira. Utilizado em provas como GMAT (*Graduate Management Admissions Test*) e o TWE (*Test of Written English*), onde o autor deve avaliar problemas situacionais geradas pela ferramenta.

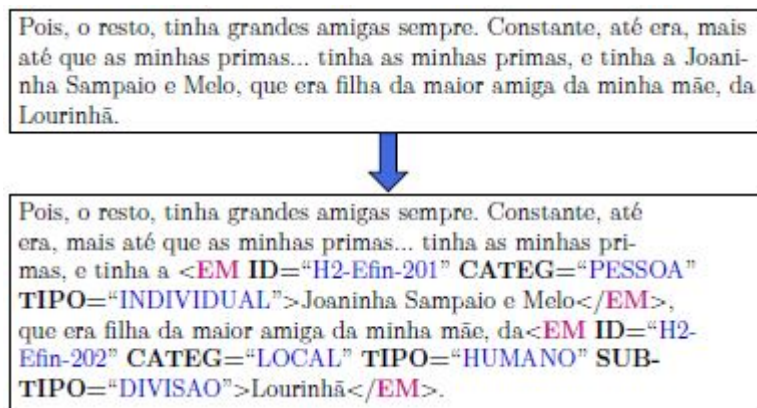
Com a NLP vamos extrair das redações uma série de características que serão repassadas aos algoritmos de aprendizado de máquina. Características tais como tipos de erros gramaticais, quantidade de palavras, frases e parágrafos, quantidade de pontuações (vírgulas, pontos, interrogações) presentes no texto.

2.3 Extração de Entidade Nomeadas (NER)

Dentre as tarefas da NLP damos destaque a tarefa de reconhecimento de entidades nomeadas que é definida como uma subárea de estudo no campo da extração de informação. É a tarefa de identificar e classificar automaticamente entidades nomeadas (NEs) em textos em forma livre (CHAN; ROTH, 2011). Entidades tais como *pessoa*, *lugar*, *organização*, *tempo* entre outros. Jiang (2012) defini a tarefa como encontrar uma sequência de palavras que designa uma entidade do mundo real. A exemplo do que é esperado, o sistema de NER produz textos estruturados como apresentado no exemplo da Figura 1 onde a entidade reconhecida é **Joaninha Pessoa** como uma entidade da categoria **PESSOA**.

NEs está presente na maioria dos textos. Um exemplo disso é que 10% dos textos informativos dos jornais apresentam nomes próprios (FRIBURGER; MAUREL, 2004).

Figura 1 – Exemplo do resultado da execução do NER (PIROVANI, 2019)



Para além da identificação de notícias o trabalho de Pirovani, Spalenza e Oliveira (2017) mostra a aplicação na criação automática que questões a partir do NER. Nesse trabalho usaremos esta técnica avaliar um texto a partir da perspectiva de que um texto deve apresentar dados e fatos para defesa do ponto de vista sobre o tema proposto na redação do ENEM/INEP como visto no trabalho de Alves et al. (2019).

2.4 Métricas

Para medir a eficiência tanto do sistema proposto nesse trabalho, quanto dos avaliadores humanos, adotamos medidas de eficácia. Também usaremos essas mesmas métricas para comparar nossos resultados com outros trabalhos. Nas próximas Seções descrevemos as métricas utilizadas para estes fins.

2.4.1 Kappa Quadrático Ponderado (QWK)

O QWK é usados para quantificar a concordância entre dois avaliadores quando realizam uma avaliação nominal ou ordinal de uma mesma amostra. Neste caso o QWK avaliará a concordância entre um avaliador humano e um AES.

A métrica é usado no desafio *Automated Student Assessment Prize* (ASAP) proposto no Kaggle (Hewlett Foundation, 2012) e em muitos outros trabalhos (ATTALI; BURSTEIN, 2006; CHEN; HE, 2013; ZESCH; WOJATZKI, 2015; AMORIM; VELOSO, 2017; FONSECA; MEDEIROS; KAMIKAWACHI, 2018) usam o QWK como métrica de avaliação da concordância entre a avaliação humana e a pontuação da máquina. Por essa ampla adoção adotamos essa métrica para referenciar a comparação entre os trabalhos da área.

A Equação (2.1) apresenta a fórmula do QWK, onde comparamos as notas do conjunto de redações w dadas pelos avaliadores i e j e normalizamos o quadrado da

diferença de cada nota pelo quadro da quantidade de elemento N menos 1.

$$\omega_{ij} = \frac{(i - j)^2}{(N - 1)^2} \quad (2.1)$$

É esperado que quão maior a concordância entre os avaliadores o valor de QWK seja próximo a 1, o que sinaliza que nosso modelo se aproxima de um avaliador humano ao avaliar uma redação. Caso contrario, se o valor de QWK se aproximar de 0, nosso modelo estará dando notas destoantes em relação ao humano, que avalia conforme os critérios do ENEM/INEP para as Competências que propomos avaliar nesse trabalho.

2.4.2 Mean Absolute Error (MAE)

Outra medida que usaremos nesse trabalho será o Erro Médio Absoluto (do inglês *Mean Absolute Error*), apresentado na Equação 2.2, é a média das diferenças entre os valores reais e os preditos. Dado que n é o número de amostra, y_i é a classe real e y^t é a classe predita pelo classificador.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y_i^t| \quad (2.2)$$

Logo o MAE será definido entre $0 > 0 > \infty$ e retornará a magnitude do erro e não sua direção. Quanto menor o erro melhor a avaliação do modelo. Essa métrica será usada para comparar os resultados obtidos com o trabalho proposto com a solução adotada por Júnior (2017).

2.4.3 Avaliação de Juízes

A necessidade de medir a concordância entre avaliadores se expressa diretamente no trabalho de (COHEN, 1960) que quantifica a correspondência entre avaliações psicológicas para o diagnóstico de doenças mentais. Nesse caso o diagnóstico é realizado por psicólogos através de observação comportamental. Essa abordagem permitiu aos profissionais da área alinharem critérios patológicos de doenças de transtorno mental.

Nossa proposta também visa estabelecer métrica de concordância tal qual o trabalho de Cohen (1960), porém para medir a concordância entre os avaliadores humanos para além das métricas impostas pelo ENEM/INEP. O trabalho de Alves e Oliveira (2019) propõe um método para comparar avaliadores humanos ao imputar notas a uma redação. O modelo consiste em obter a correlação de Pearson (BUSSAB; MORETTIN,), disposto na Equação 2.3, e quantifica a associação ρ entre as notas dadas pelos avaliadores n_i e n_j onde $i \neq j$. O valor da correlação será $-1 \leq \rho \leq 1$. As Competências são avaliadas uma

de cada vez, então para cada um das avaliadores haverá um ρ por Competência.

$$\rho = \frac{\text{cov}(n_i, n_j)}{\sqrt{\text{var}(n_i) \cdot \text{var}(n_j)}} \quad (2.3)$$

Na Seção 5 mostramos em maior detalhe como através desse ferramental observamos que mesmo entre avaliadores humanos é comum divergências avaliativas. Uma vez que nossa proposta é imitar apenas **um** avaliador é certo que haverá diferença na avaliação de modelos treinados a partir de múltiplos avaliadores, ou se tivermos modelos construídos a partir das notas dadas por um avaliador X e compararmos com as notas dadas por um outro avaliador qualquer diferente de X .

3 Trabalhos Relacionados

Neste capítulo será apresentado uma análise de trabalhos recentes que se relacionam e que valha de fundamentação para com o objetivo deste estudo.

3.1 Avaliação Automática de Redações (AES)

Segundo [Page \(1994\)](#) a avaliação automática de redações (do inglês *Automated Essay Scoring* - AES) consiste na tarefa de avaliar redações automaticamente imitando avaliadores humanos. Em seu trabalho propõe a descoberta de uma relação linear para inferir notas a partir de 30 características extraídas das redações em inglês com métodos de NLP. O método de relação linear utilizado para criar um modelo para avaliar as redações foi a regressão linear, algoritmo amplamente adotado em muitos trabalhos na área de AES ([PAGE, 1994](#); [LANDAUER](#); [LAHAM](#); [FOLTZ, 2003](#); [ATTALI](#); [BURSTEIN, 2006](#); [MILTSAKAKI](#); [KUKICH, 2004](#); [FAULKNER, 2014](#); [AMORIM](#); [VELOSO, 2017](#)).

À medida que a computação apresenta novas ferramentas as pesquisas na área de AES vem crescendo e novas técnicas de extração de características, predição de notas e a inclusão de novos idiomas à tarefa. Na década de 90 com o advento da internet, os avanços no campo da NLP e dos métodos estatísticos os AES entra como sistema de apoio combinado com o avaliador humano. Então o AES passa a ser usado em diferentes tipos de avaliações tais como *Graduate Record Examination* (GRE), *Test of English as a Foreign Language* (TOEFL), *Graduate Management Admissions Test* (GMAT), *Scholastic Aptitude Test* (SAT), *American College Testing* (ACT), *Test of English for International Communication* (TOEIC), *Analytic Writing Assessment* (AWA), *No Child Left Behind* (NCLB) e *Pearson Test of English* (PTE).

A utilização da AES uma avaliação de testes de proficiência de língua estrangeira é observado na proposta por ([BURSTEIN et al., 2002](#)), já comentada na sessão 2. Essa prática foi adotada mais tarde também para outros idiomas tais como mandarim([WANG](#); [CHANG](#); [LI, 2008](#)), japonês([ISHIOKA](#); [KAMEDA, 2004](#)) e espanhol([ALFONSECA](#); [PÉREZ, 2004](#)).

A área de AES ganha ainda mais atenção a partir do corpus liberado como parte de uma competição ASAP no *Kaggle* em 2012. O corpus disponível tornou-se popular na avaliação do sistemas de AES na língua inglesa. O conjunto de redações apresenta um grande volume de texto tendo aproximadamente 3000 redações para cada tema. A abundância de dados disponíveis e abertos fizeram prosperar o número de pesquisas de AES para a língua inglesa.

[Zupanc \(2018\)](#) apresenta em sua pesquisa abordagens para aferir coerência textual

aplicando métodos de análise de distância espacial para avaliar a coerência através de métricas de distancias, utilizando-se das sentenças do texto como um ponto no espaço. Também faz uso de ontologias baseado na proposta de [Gutierrez et al. \(2013\)](#) para avaliação da consistência lógica do texto. Além disso tem como métrica a similaridade entre a redação e o texto motivador.

Recentemente [Ke \(2019\)](#) apresentou a avaliação do texto argumentativo-persuasivo também para a língua inglesa e a utilização de uma rede neural para reconhecer no texto atributos que quantifiquem a qualidade dos argumentos a partir de *corpus* anotados para treinamento do modelo. Como consequência das predições obtidas o AES oferece ao autor da redação um *feedback* para melhorias no texto.

3.2 AES Aplicado a avaliação ENEM/INEP

Os trabalhos para correção automática para a língua portuguesa tem crescido, porém ainda é tímido se comparado com os avanços na língua inglesa. Em sua maioria as abordagens fazem uso de modelos baseados em regras gramáticas. O trabalho realizado por [Amorim e Veloso \(2017\)](#) utiliza-se da abordagem proposta por [Attali e Burstein \(2006\)](#). Para extrair as características das redações usou o *CoGrOO* ([KINOSHITA; SALVADOR; MENEZES, \(SILVA, 2013\)](#)). Afim de predizer as notas das cinco Competências com as *features* extraídas aplicou uma simples técnica de regressão linear tal qual apresentado por [Page \(1994\)](#). O método foi aplicado sobre um conjunto de 1840 redações divididos em 96 tópicos diferentes extraídas banco de redações do site *UOL*.

Em um novo trabalho sobre os dados coletados em [Amorim e Veloso \(2017\)](#) o trabalho de [Amorim, Cançado e Veloso \(2018\)](#) investiga a presença de vieses e apresenta *features* para qualificar avaliações enviesadas. O viés em geral pode ser definido como um indivíduo observado que foge a regra normal da população, nesse caso uma redação que apresente mesmas características de outras 10 porém recebeu uma nota diferente pode ser visto como caracterização do comportamento de viés. [Amorim, Cançado e Veloso \(2018\)](#) utiliza análise de palavras e expressões na redação para avaliar a presença de viés sobre as mesmas 1840 redações coletadas sobre as 5 Competências do ENEM/INEP.

Os sistemas de AES para a língua portuguesa também foi foco do trabalho de [Fonseca, Medeiros e Kamikawachi \(2018\)](#). Apresentando um conjunto de 56664 redações e técnicas que usam redes neurais. Para extrair as *features* foi utilizado o sistema *UNI-TEX*([MUNIZ, 2004](#)) para avaliação da ortografias das palavras. Além disso analisa a similaridade dos textos com o texto motivador do tema abordado. Por fim avalia se há no texto um conjunto de palavras e expressões mais frequentes em redações. Assim como no trabalho realizado por [Amorim e Veloso \(2017\)](#) o autor aborda a avaliação de todas as cinco Competências da redação do ENEM/INEP com uma arquitetura neural hierárquica

similar a utilizado por (DONG; ZHANG; YANG, 2017), com duas camadas de uma rede neural recorrentes que usam células bidirecional de memória de longo prazo (do inglês *bidirectional long short-term memory* - BiLSTM).

Em contra partida a abordagens que utilizam-se das mesmas *features* para avaliar todas as Competências, outras pesquisa criam sistemas especializados em atacar uma Competência por vez. O caso da abordagem proposta por Júnior (2017) trata apenas da Competência 1 - Domínio da norma culta padrão da língua portuguesa, utilizando o Support Vector Machine (SVM)(CORTES; VAPNIK, 1995) e o conjunto (*ensemble*) de classificador *Gradient Boosting*(GB)(FRIEDMAN, 2000) para criação de modelos que dêem uma nota para a redações. Além disso utiliza como *features* para avaliação da redação cada erro encontrado pelo corretor gramatical do CoGroo (SILVA, 2013) que estão além das que foram propostas no trabalho de (AMORIM; VELOSO, 2017), uma vez que ao invés de sumarizar os erros gramaticais apontados pelo CoGroo em uma única característica considera cada tipo de erro com uma característica do problema.

Também temos a proposta de Filho et al. (2018) que busca avaliar apenas a Competência II - compreender a proposta de redação e aplicar conceitos das várias áreas de conhecimento para desenvolver o tema, dentro dos limites estruturais do texto dissertativo-argumentativo em prosa. Faz uso das *features* mais comuns já abordadas por Amorim e Veloso (2017) e Júnior (2017) associadas a extração de características de estruturas argumentativas obtidas através da contagem de palavras que foram retiradas do Dicionário analógico da Língua Portuguesa(AZEVEDO, 1999). Um dicionário analógico reúne palavras em grupos de acordo com sua afinidade de ideias, partindo de conceitos para indicar seus significantes linguísticos. Essas também faz uso do SVM para gerar um modelo baseado regressão para aferir notas. Filho et al. (2018) ressalta que a desproporção da quantidade de exemplos redações para cada nota pode ter efeitos negativos sobre o aprendizado do classificador. Para contornar esse problema faz uso do algoritmo *Synthetic Minority Oversampling Technique*(SMOTE) para balancear sinteticamente o balanceamento das amostras.

Tabela 4 – Resumo do estado da arte sobre AES

AES	Tipos Atributos	Metodologia	Modelo Predicação	Dataset	Quantidade Redações
Geral					
Zupanc (2018)	NLP	NLP	Regressão Linear, <i>random forest</i> , <i>feed-forward neural networks</i> , <i>extremely randomized trees</i> e <i>regression trees</i>	Kaggle	17195
Português					
Júnior (2017)	NLP	NLP	SVM e GB	UOL Educação Brasil Escola	953 4902
Amorim e Veloso (2017)	NLP	NLP	Regressão Linear	UOL Educação	953
Fonseca, Medeiros e Kamikawachi (2018)	NLP	NLP	BiLSTM	Formulado pelo autor	56664
Filho et al. (2018)	NLP	NLP	SVM	UOL Educação	1982

3.3 Reconhecimento de Entidade Nomeadas (NER)

Este trabalho estabelece um ponto além das regras gramaticas na correção de redação para a língua portuguesa e também dos trabalhos de AES para o ENEM/INEP. Adotamos um modelo de reconhecimento de entidades para avaliar se o aluno utiliza-se de dados e fatos na elaboração de sua redação, ou se a estrutura do texto apresenta com clareza tais elementos ao ponto que um sistema de NER treinado em outro domínio de texto possa reconhecer tais estruturas.

Escolhemos um sistema de abordagem híbrida para língua Portuguesa defendido por (PIROVANI, 2019) que combina a técnica de Aprendizagem de Máquina *Conditional random fields* (CRF)(LAFFERTY; MCCALLUM; PEREIRA, 2001) e uma estratégia linguística conhecida como gramática local (LG) que é uma forma de representar as regras da abordagem linguística, formalismo introduzido por Gross (1999). "Gramáticas Locais são construídas manualmente e são uma forma de agrupar ou capturar expressões que possuem características comuns, sejam elas sintáticas ou semânticas, que têm sido usadas em tarefas como NER e análise de sentimentos"(PIROVANI, 2019).

Pirovani et al. (2019) mostra a aplicação da técnica de NER em diversos contextos, em texto de múltiplos domínios. Estendendo essa proposta ao domínio de redação NER vai para além da contagem de palavras e identificação de classes gramaticais e do uso de dicionários abrindo espaço para uso dos padrões comuns da língua portuguesa.

Fazal, Dillon e Chang (2011) já propunha a utilização da técnica de NER em sua abordagem de AES denominada *OzEgrader*. A proposta considera diferentes aspectos do conteúdo e do estilo: audiência, estrutura do texto, caráter e cenário, parágrafos, vocabulário, estrutura da frase, pontuação, ortografia, coesão e ideias. Para modelar relações lineares ou não lineares entre atributos além do sistema de NER o *OzEgrader* também emprega métodos como redes neurais artificiais e regressão nebulosa para avaliar redações.

A abordagem utilizada nesse trabalho foi proposta em Alves et al. (2019) em que foram utilizadas o reconhecimento de 5 categorias de NEs para avaliar as redações. Os resultado obtidos para o reconhecimento de entidades e a predição de notas para a Competência 3 não utilizou nenhum texto do domínio de redações e obteve-se resultado para prever pelo menos as notas da Competência III.

3.4 Avaliação de Juízes

Além de avaliar a redação é necessário pensar também nos avaliadores responsáveis por corrigir as redações. Como visto no Capítulo 1 o custo da correção da redação é elevado e cada vez que os professores divergem em uma ou mais Competências, tendo uma

diferença de 80 pontos entre eles, esse custo aumentará, porque um novo avaliador será convocado a corrigir a redação, e se prevalecer a diferença uma nova banca de avaliadores será formada.

Pensando também na figura do professor especialista que "ensinará" o modelo a avaliar as redações é de máxima importância a qualidade dos avaliadores humanos, pois influenciará diretamente nos modelos de avaliadores criados pelos algoritmos de aprendizado de máquina. É importante que avaliadores apresentem alguma uniformidade com outros avaliadores ao avaliar as redações.

O trabalho de [Alves e Oliveira \(2019\)](#) aborda um modelo estatístico para perfilação de avaliadores de redação em um ambiente virtual de aprendizado. A abordagem discute uma forma de quantificar o quão bem treinado um avaliador de redações é monitorando a cada nova redação corrigida pelo professor como ele se assemelha a avaliadores experientes.

A exemplo do trabalho de [Matos \(2014\)](#) que analisa avaliadores de maneira manual em uma comparação intraclass e explicita o descompasso significativo na concordância entre avaliadores de redação, mesmo utilizando critérios objetivos para a correção de redações do ENEM/INEP. Quando não há critérios definidos na avaliação de atividades há déficit na aprendizagem do educando ([ARTER; CHAPPUIS, 2014](#)).

4 Arquitetura do Sistema

Propomos a utilização de uma arquitetura modularizada e escalável para a extração de características afim de que o sistema acompanhe o volume de redações a ser avaliado como é possível observar na Figura 2. Utilizando um *software* de contêiner Docker¹ para virtualização do sistema operacional e embarcando cada sistema desse ambiente.

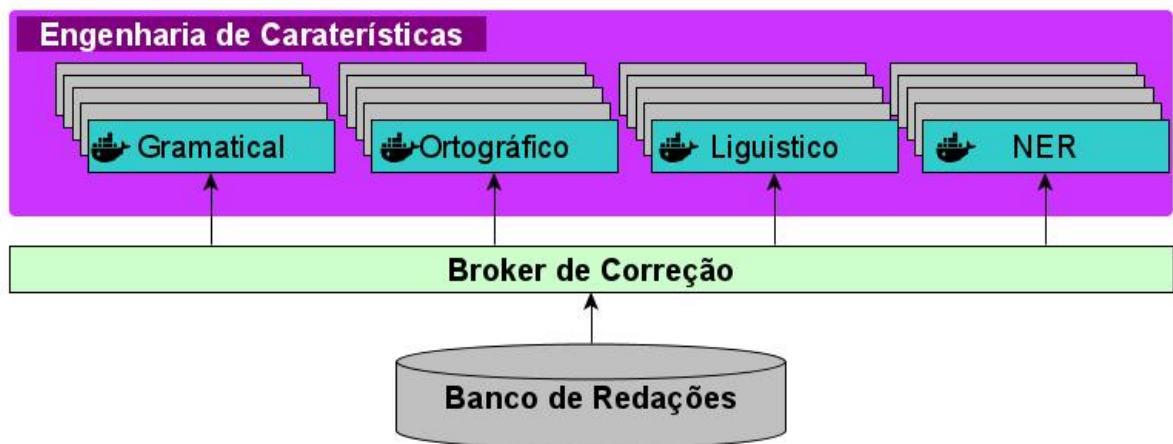


Figura 2 – Arquitetura do sistema

O sistema de *Broker* de Correção(BC) conhece a função de cada nó e é responsável por fazer a leitura do banco de redações e disparar a cada nó de sistema a operação que deve ser realizada. O BC também é responsável por gerar os modelos de aprendizado de máquina que será usado para corrigir as redações. Por fim também avaliará novas redações a partir do modelo gerado anteriormente.

Afim de que o sistema seja aperfeiçoado e que os módulo desenvolvidos estejam livremente disponíveis no seguinte repositório: <https://gitlab.com/Jamesasj/dissertacao_aes>. O módulo de NER é parte do trabalho de (PIROVANI, 2019) e seu código fonte não será disponibilizado neste projeto.

Nas próximas sessões apresentaremos primeiramente o formato em que as redações são representadas dentro do sistema de AES após a execução da extração de características. Depois apresentamos cada sistema de extração de características suas técnicas e arquiteturas. Também discutiremos sobre a arquitetura do BC e as operações de aprendizado de máquina incorporado a ele. Por fim apresentamos as técnicas utilizadas para aferir uma nota a partir do conjunto dos dados.

¹ <https://www.docker.com/>

4.1 Representação Vetorial

Para representar as redações adotamos o modelo vetorial proposto por [Salton, Wong e Yang \(1975\)](#). O modelo vetorial é uma forma de interpretação computacional de documentos. Esse modelo se baseia na indexação da frequência de termos, do inglês *Term Frequency - TF*, obtidos a partir processamento do *bag of words*, conjunto de palavras obtidas a partir dos documentos.

Adaptamos o modelo proposto para representar as características apresentadas na sessão anterior. No nosso modelo cada redação r compõe um conjunto de documentos $R = \{r_1, r_2, r_3, \dots, r_{|R|}\}$. Cada redação r é composto por *features* f extraídas a partir do módulo de engenharia de características, que será apresentado na Seção 4.2, assim representamos uma redação como um vetor $r_i = \{f_1, f_2, f_3, \dots, f_k\}$.

A redação passa a ser compreendida no espaço vetorial, pois o vetor representa o texto em formato numérico. Com os valores f para cada característica k . Essa transformação é realizada para a entrada das redações nos algoritmos de aprendizagem de máquinas, apresentados na Seção 4.3. A Figura 3 representa conceitualmente como a representação das redações após a extração das características para a formação do *dataset* para treinamento e teste dos modelos. As redações analisadas são incluídas na tabela de dados de acordo com o que se deseja fazer.

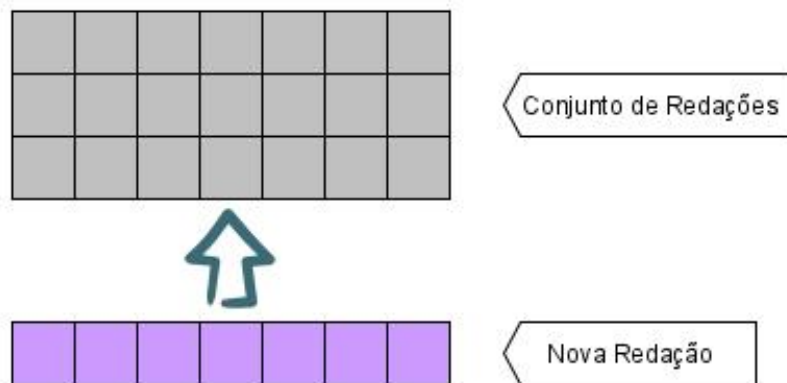


Figura 3 – Representação da organização das características do conjunto de redações

A Figura 4 mostra como são organizadas as *features* para duas redações. Cada uma delas é separada por vírgulas e cada um dos valores representa uma característica coletada pelos módulos de extração de características. Adiante, nas próximas sessões, são apresentadas as caraterísticas.

```

tema_100_avaliao_11_resposta ,20 ,28 ,28 ,9 ,2 ,0 ,70 ,0 ,0 ,17
,8 ,5 ,9 ,29 ,29 ,31 ,5 ,10 ,7 ,9 ,27 ,0 ,0 ,0 ,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,1,0
,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0
,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0
,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,1,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0
,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0
,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0
,0,0 ,1,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,1,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0
tema_100_avaliao_13_resposta ,33 ,23 ,27 ,14 ,2 ,1 ,66 ,0 ,0 ,10
,3 ,5 ,13 ,21 ,44 ,27 ,0 ,3 ,7 ,17 ,45 ,0 ,0 ,0 ,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0
,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0
,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0
,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0
,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0
,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0
,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,1,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0
,0,0 ,1,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0 ,0,0

```

Figura 4 – Exemplo da organização das características de duas redações.

4.2 Engenharia de Características

Desenvolvemos um AES baseado nas características disponíveis na literatura. Usaremos duas classes de características como está disposto na Figura 5 as características de linguística e a de conteúdo. A Figura também mostra as sub-divisões destas duas classes clássicas do sistemas de AES ([ATTALI; BURSTEIN, 2006](#)).

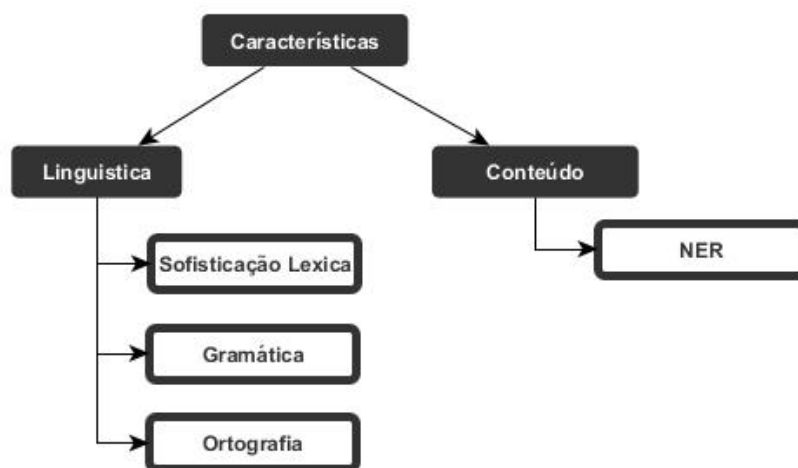


Figura 5 – Características extraídas pelo AES

Estas classes de características são usadas nesse trabalho para descrever a qualidade da escrita do aluno e através da extração automática e da seleção de um modelo de aprendizado que avaliará as redações. Atributos como sofisticação léxica, erros gramaticais,

pontuação, diversidade léxica são as subclasses das *features* de linguística.

No segundo grupo de características trabalharemos o reconhecimento de conteúdo do texto utilizando o NER para identificar elementos no texto que sejam entidades do mundo real. Existem outras classes e subclasses que ainda são passíveis ou de implementação, ou de adaptação para a língua portuguesa. Podemos encontrar uma relação extensa com mais classes de características no trabalho de Zupanc (2018). Cada característica que foi usada nesse trabalho com sua respectiva subclasse é apresentada na Tabela 5.

4.2.1 Características Linguísticas

Os atributos da classe da linguística buscam descrever os aspectos de sofisticação léxica, identificar os erros de gramática e os erros ortográficos dispostos na redação.

A extração de **sofisticação léxica** obtém características gerais que demonstram, por exemplo variabilidade do vocabulário do estudante, quantidade de parágrafos que foi produzida, quantidade de palavras no texto. Para isso utilizamos a análise a partir do *OpenNLP*.

Para avaliar a **Gramática** utilizamos o programa *Cogroo* (SILVA, 2013). A análise da gramática consiste em avaliar se a redação submetida obedece ao conjunto de regras que regem o uso da língua portuguesa. Cada regra se torna uma característica que será usada no modelo e caso a regra seja ofendida na redação será contabilizada no vetor de características. O *CoGroo* reconhece 124 regras gramaticais da língua portuguesa.

A análise da **Ortografia** consiste em avaliar a quantidade de palavras não reconhecidas por um dicionário dentro do texto. Utilizamos nesse trabalho o dicionário *Hunspell*. Por fim essas quantidades, de informação do texto e dos erros apontados pelo *Cogroo* e *Hunspell* são adicionados ao vetor da redação apresentado na Figura 4.

4.2.2 Conteúdos da Redação

Em nosso trabalho avaliamos a quantidade de entidades que podem ser encontradas no texto a partir do sistema proposto por Pirovani (2019) denominado *CRF+LG*. Quantificamos cada tipo de entidade reconhecido pelo sistema. Em uma visão geral a maioria dos trabalhos para a língua portuguesa as abordagens utilizando *software* de NLP tais como *coGroo* ou *UNITEX* identificam apenas a presença de um pronome no texto. Nossa abordagem diferencia-se das demais a partir do momento que é à abordagem do *CRF+LG* é capaz de qualificar as NEs.

Além disso, como apresentado na Tabela 5, podemos avaliar as características pelas suas proporções apresentadas no texto. Avaliando se um aluno que apresenta um volume muito superior de informações e fatos terá mais sucesso do que um aluno que não apresente nenhuma entidade.

Sofisticação Léxica	PoS tag
1. #palavras,	17. adj
2. #palavras únicas,	18. adv
3. #sentenças,	19. art
4. #parágrafos,	20. conj-c
5. sentenças > 70 palavras,	21. conj-s
6. #palavras/#sentenças,	22. ec
7. #sentenças/#parágrafos,	23. n
Entidades Nomeadas	24. num
8. #entidades PESSOA,	25. pp
9. #entidades LUGAR,	26. pron-det
10. #entidades VALOR,	27. pron-indp
11. #entidades TEMPO,	28. pron-pers
12. #entidades ORGANIZAÇÃO,	29. prop
13. #NEs,	30. prp
14. #NE/#palavras,	31. punc
15. #NEs/#sentenças,	32. v-fin
16. #NEs/#parágrafos,	33. v-ger
	34. v-inf
	35. vp
	36. v-pcp
Ortografia	Gramática
37. #erros ortográficos	39. #erros gramaticais,
38. #erros ortográficos/#palavras	40. #erros gramaticais/#palavras
	41-164 Regras Gramaticais do CoGroo (SILVA, 2013), disponíveis no Apêndice A,

= "quantidade de"

Tabela 5 – Resumo das Características de linguística e conteúdo implementadas no sistema

4.3 Avaliação de Redação

Após a extração de características todos os vetores de redação são submetidos ao BC. Nesse módulo será gerado o modelo responsável por avaliar as redações. A Figura 6 mostra o *pipeline* que será executado para treinamento de modelos e avaliação de novas redações. Quando uma nova redação for submetida ao BC será ele o responsável por avaliar e devolver ao emissor a nota da redação e as respectivas inconsistências anotadas pelos módulos de extração de características.

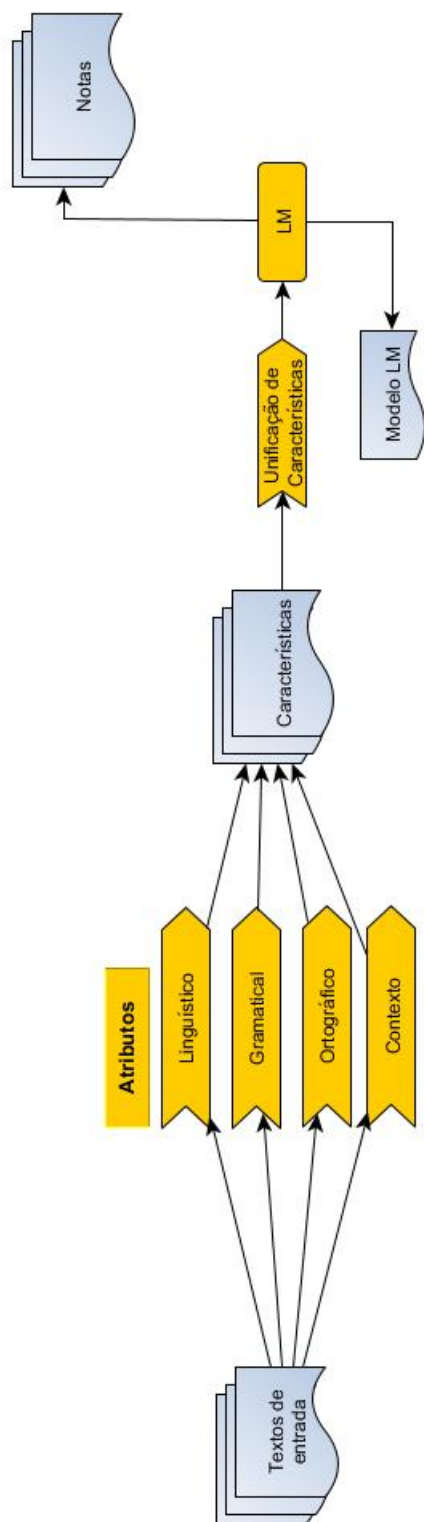


Figura 6 – Pipeline de treinamento e avaliação

A extração de características foi abordada na seção anterior, agora explanaremos as etapas do processo de avaliação da redação. O problema de avaliação de redação pode ser descrito como um problema de classificação de documentos, onde desejamos avaliar a qualidade da redação submetida a partir das características extraídas do texto. Existe uma vastidão de algoritmos desenvolvidos que buscam qualificar documentos automaticamente e muitos foram desenvolvidos na área de pesquisa de aprendizado de máquina.

Como vimos na Seção 2 nossa abordagem faz uso de algoritmos supervisionados para avaliação de redações. Algoritmos supervisionados consistem basicamente em treinar modelos a partir de redações pré-classificadas por avaliadores humanos. Com essas redações pré-classificadas, já em formato vetorial, podemos treinar o avaliador automático como detalhado a seguir:

- Separação das redações em conjuntos de treinamento e de teste;
- seleção das características;
- treinamento do avaliador;
- avaliação do modelo treinado;
- Avaliar novas redações com modelo treinado.

Podemos definir nosso problema formalmente como dado uma coleção R de redações e um conjunto $C = c_1, c_2, \dots, c_L$ de L notas possíveis com suas respectivas avaliações, um classificador de redações é uma função binária $F : R \times C \rightarrow 0, 1$, uma função que atribui um valor 0 ou 1 para cada par $[r_j, c_p]$, tal que $r_j \in R$ e $c_p \in C$. Se o valor atribuído for 1, dizemos que a redação r_j foi avaliada com a nota c_p . Se o valor for 0 dizemos que a redação r_j não foi avaliada com a nota c_p .

Treinamos um modelo para cada Competência da redação do ENEM/INEP que abordamos nesse trabalho. Logo temos a restrição de que cada redação r_j pode ser avaliado apenas com uma única nota c_p .

A etapa de separação é basicamente reserva de uma parte do conjunto R das redações que foram previamente avaliadas por avaliadores humanos para avaliação do modelo gerado denominamos esse sub-conjunto de R_{ts} . Uma vez que as notas C para esse sub-conjunto de R é conhecido podemos comparar a nota do avaliador humano com a nota dada pelo AES treinado uma vez que o modelo gerado não conhece nenhuma das redações previamente reservadas do conjunto inicial. Denominamos o conjunto remanescente de R_{tr} e este sub-conjunto será usado para treinamento do AES. A seguir nas próximas seções descrevemos os detalhes das demais etapas de treinamento do AES usando estes sub-conjuntos gerados.

4.3.1 Seleção de características

Essa etapa consiste em identificar quais características f_k das redações r_j que melhor representem redações com a nota c_p . Segundo (BAEZA-YATES; RIBEIRO-NETO, 2011) um espaço de características muito grande pode tornar os modelos pouco práticos pois a avaliação de novas redações se tornaria em um processo muito lento.

A descoberta de padrões nas características extraídas do texto é um passo importante não só para reduzir o número de *features*, mas também para aumentar a eficiência e a acurácia da avaliação de redações (RAYMER et al., 2000). Gerando assim novo conjunto de redações que tem apenas as *features* relevantes, que explicam cada nota.

Modelamos o processo de seleção de características como um problema de otimização das métricas de classificação. Para otimizar os resultado adotamos a técnica de Algoritmo Genético (*Genetic Algorithm* - GA) (GOLDBERG, 1989). A técnica é inspirada no processo de seleção natural focada para os processos de busca e otimização, aplicando técnicas de hereditariedade, mutação e seleção. Pelo processo observador por Charles Darwin na natureza onde por meio de muitas gerações, populações de uma mesma espécie evoluem de acordo com os princípios da seleção natural (GOLDSCHMIDT; PASSOS; BEZERRA, 2015).

Nossa abordagem consiste em inicialmente gerar aleatoriamente uma população V de vetores binários v_z com o tamanho k do número de características. Quando a posição k do vetor v_z correspondente a características for igual a 1 a característica será mantida no conjunto R e de seus sub-conjuntos. Quando v_z na posição k for 0 a característica na posição correspondente será removido do conjunto R e de seus sub-conjuntos Rts e Rtr .

Uma vez executado o treinamento e avaliação de toda a população de V teremos o resultado da métrica de avaliação do modelo, exposta na Seção 2, para todas as combinações feitas aleatoriamente nessa primeira população, chamaremos aqui essa métrica obtida de aptidão. A métrica indica o quão bem o cromossomo v_z da população consegue avaliar uma redação. Quão maior a aptidão do cromossomo melhor. A etapa de avaliação é detalhada na próxima Seção. Sobre essa população é realizado processo de escolha de cromossomos para uma nova iteração de seleção de características e avaliação de modelos.

A nova população é gerada a partir de três operação básicas de um GA. A primeira é chamada de seleção, serão escolhidos alguns cromossomos da população atual e serão incluídos da nova população. Essa escolha pode se dar por avaliar a porção mais apta de cromossomos, de maneira estocástica, elitismo, entre outros meios. A segunda operação é a combinação entre pares de cromossomos, onde ocorre troca de um ou vários genes dando origem a um novo vetor. E a terceira operação é a mutação que consiste em alterar probabilisticamente um ou vários genes de um cromossomo.

Após a geração dessa nova população teremos uma nova iteração de treinamento e

avaliação da aptidão dos cromossomos. Esse processo ocorre até que se alcance o número de iterações definido na parametrização inicial. Vale ressaltar que quanto maior for a aptidão de um cromossomo maior será o custo computacional, uma vez que a aptidão tende a aumentar com o aumento do número de iterações.

4.3.2 Avaliação de modelo com *Cross-Validation*

Após escolhermos as características que serão usadas no treinamento no modelo do AES e efetivamente executar os algoritmos de aprendizado de máquina sobre o conjunto de redações devemos lançar mão de técnicas que avaliem o modelo e forneça uma medida de confiabilidade sobre este avaliador. Também conhecida como rotação de estimação a validação cruzada (em inglês *Cross-Validation*) é adotado com método padrão para garantir a validação estatística dos resultados de classificação dos modelos de aprendizado de máquina. Kohavi (2001) define formalmente o processo de *k-fold cross validation* como a sub-divisão aleatória do conjunto Rtr em k sub-conjuntos mutuamente exclusivos $Rtr_1, Rtr_2, \dots, Rtr_k$ com um tamanho aproximadamente igual.

Então começamos um processo de treinamento e teste k vezes; cada rodada $\theta \in 1, 2, \dots, k$ são treinados com Rtr Rtr_θ e testados com Rtr_θ . O *cross-validation* avaliará o modelo sobre os resultados de todas as rodadas. Vale ressaltar que nesse processo o conjunto de teste Rts não será utilizado. Só são usados os dados de treinamento.

4.4 Métodos de Aprendizado Supervisionado

Dentro da arquitetura apresentada na Figura 2 temos o BC que centraliza as *features* dos textos e também nos fornece o *pipeline* de treinamento de um avaliador automático de redação apresentado na Figura 6. Dentro do BC também ficam os métodos de inteligência artificial que geraram os modelos de AES. Nessa seção apresentamos os modelos que foram incluídos dentro do BC.

Podemos classificar os algoritmos de inteligência artificial em três classificações quanto ao seu modelo de aprendizado (BAEZA-YATES; RIBEIRO-NETO, 2011). A primeira classificação são os métodos baseados em aprendizado supervisionado que requer uma função de aprendizado dos dados de treinamento. Nos sistemas supervisionados o especialista humano treina o algoritmo com uma quantidade de dados para atuação dessa tarefa. O segundo são os modelos baseados em aprendizado não supervisionado tidos como independente de exemplos classificados por especialistas, nenhum dado de treinamento é fornecido para aprendizado. Por fim temos o aprendizado semi-supervisionado que combina uma quantidade pequena de dados rotulados com uma porção maior de dados não rotulados para melhorar a predição.

A essência do desafio de avaliar redações está em imitar avaliadores humanos ao dar notas para cada Competência, assim utilizaremos na nossa proposta modelos que se furtam do aprendizado supervisionado. Aproveitamos o conhecimento que será adquirido a partir de redações corrigidas por avaliadores humanos. A seguir descrevemos alguns métodos utilizados em nosso experimentos.

4.4.1 Regressão Linear

O primeiro algoritmo aplicado no nosso modelo foi a regressão linear pela sua simplicidade e sua ampla aplicação em diversas outras abordagens (AMORIM; CANÇADO; VELOSO, 2018) e também outros idiomas (PAGE, 1994; LANDAUER; LAHAM; FOLTZ, 2003; ATTALI; BURSTEIN, 2006; MILTSAKAKI; KUKICH, 2004; FAULKNER, 2014). Definimos como exemplo a Figura 7 onde gostaríamos de prever uma nota Y dado um quantidade de erros ortográficos X .

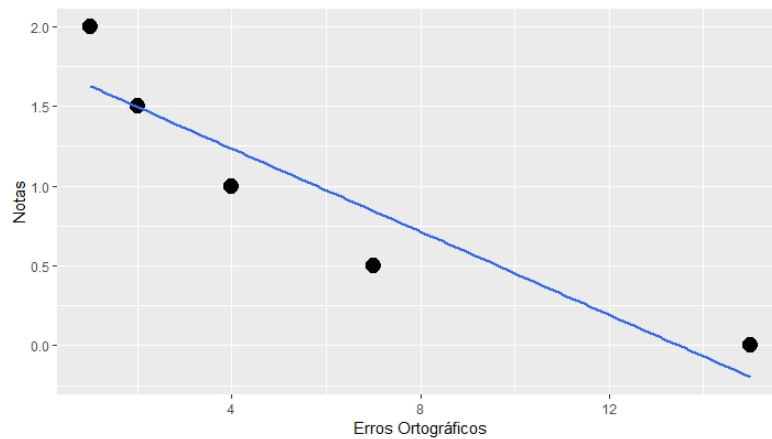


Figura 7 – Exemplo de regressor linear para uma característica.

Observando a Figura 7 logo percebemos que y , onde $y \in Y$, vemos que as notas vão reduzindo à medida que os erros ortográficos aumentam. Um modelo razoável para $E(Y|x)$ pode ser:

$$E(Y|x) = \mu(x) = \alpha + \beta x \quad (4.1)$$

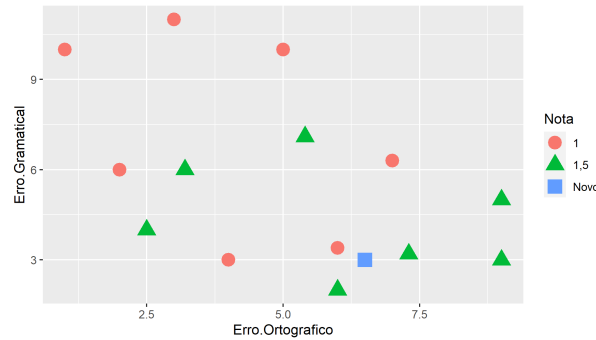
Ou seja as notas são uma função linear dos erros ortográficos. Podemos deduzir uma regressão linear pressupõe um α e um β que explique a relação linear entre X e Y . Sabemos que α representa o intercepto, o ponto onde a reta exposta na Figura 7 corta o eixo das ordenadas, e β o coeficiente angular, que é o quanto a média das notas Y varia a medida que os erros ortográficos aumentam X .

4.4.2 K-Nearest Neighbor (KNN)

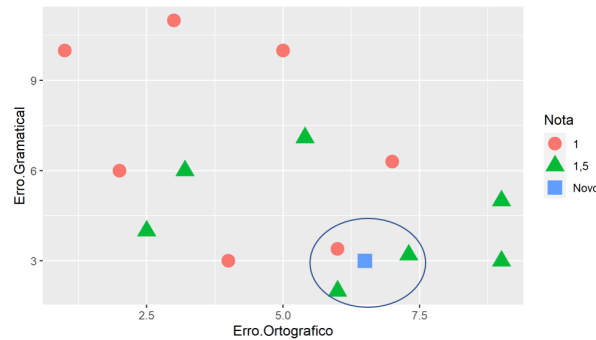
Também utilizamos nesse trabalho o método K -NN para aferir notas de redação. No método do K -NN o conjunto de dados é armazenado, quando houver uma nova redação

ela será comparada com todas as redações do conjunto de treinamento para identificar os k vizinhos mais próximos, mais semelhantes de acordo com alguma métrica. As métricas mais utilizadas são a *Euclidiana* e de *Manhattan*.

Considere o exemplo a baixo onde um conjunto de redações fictícias representadas na Figura 8a. Para fins de exemplo temos duas notas possíveis neste conjunto de redações e uma nova redação para ser avaliada pelo modelo.



(a) Conjunto de redação



(b) Delimitação dos k vizinhos mais próximos

Figura 8 – Three simple graphs

Para avaliar um conjunto o método K -NN consiste nos seguintes passos:

1. Cálculo da distância do novo registro a cada um dos registros existentes no conjunto de redações.
2. Identificação das k redações do conjunto que apresentem menor distância em relação a nova redação.
3. Apuração da nota mais frequente entre os K registros identificados na etapa anterior.

Supondo que queremos encontrar os K vizinhos mais próximos sendo $K = 3$, mostramos na Figura 8b os três elementos mais próximos ao novo registro delimitado pelo algoritmo K -NN. Ao avaliar a classe com maior ocorrência, delimitada pelos 3 vizinhos mais próximos, diremos que a nova redação tem nota 1.5. Durante a pesquisa utilizamos K -NN apenas no conjunto de redações para avaliar a Competência I.

5 Experimentos e Resultados

Neste capítulo detalhamos os experimentos realizados com o modelo proposto no Capítulo 4. Detalharemos o conjunto de redações utilizadas nesse trabalho. Logo depois discutimos os resultados obtidos e comparamos com outros trabalhos que avaliaram as mesmas competências.

5.1 Base de Dados

A base de dados desse trabalho é composta por redações extraídas do banco de redações do *UOL*. As redações coletadas do portal ([UOL, 2019](#)) também foram utilizados nos trabalhos de [Júnior \(2017\)](#) e [Amorim, Cançado e Veloso \(2018\)](#), [Amorim e Veloso \(2017\)](#).

Nas redações do ENEM/INEP as notas das competências são dadas entre 0-200 com interpolações de 50, as redações que usamos nos experimentos realizados nesse trabalho as notas estão entre 0-2 com interpolação de 0.5.

Obtemos um total de 2211 redações cada redação teve em média 307.33 palavras e 11.26 sentenças. Os alunos também escrevem em média 5.52 parágrafos. No conjunto de redações 100% dos textos apresentam erro ortográficos ou erros gramaticais.

Na execução dos experimentos seguimos os métodos adotados por [Fonseca, Medeiros e Kamikawachi \(2018\)](#), [Amorim, Cançado e Veloso \(2018\)](#), [Júnior \(2017\)](#) e removemos as redações que foram zeradas, para mitigar ruído causado por elas no processo de aprendizagem de máquina. Nas regras do ENEM/INEP são determinados 5 motivos pelos quais uma redação deva ser zerada. Assim nós não usaremos as redações nota zero do nosso conjunto de redações pois podem introduzir ruído nos modelos de AES.

5.2 Experimentos Realizados

Para cada Competência realizamos experimentos e comparamos os resultados obtidos entre os experimentos. Depois comparamos os resultados obtidos com outros modelos descritos na literatura disponível. Nossa base de dados extraída do portal *UOL* é comparada com os resultados obtidos por [Júnior \(2017\)](#) para a Competência I e para os trabalhos de [Fonseca, Medeiros e Kamikawachi \(2018\)](#) e [Amorim e Veloso \(2017\)](#) comparamos os resultados tanto para a Competência I quanto para a Competência III.

A Tabela 6 apresenta os experimentos realizados para a Competência I. Combinamos as características e métodos propostos na Sessão 4 para criar um modelo de AES para as

Competência I. Para encontrar os erros ortográficos das redações utilizamos o dicionário *Hunspell*. E a extração gramatical utilizamos o revisor *Cogroo*. Avaliamos quais as classes são de maior relevância para cada uma das Competências avaliadas através da combinação de *features* utilizando GA.

Tabela 6 – Resumo dos Experimentos Realizados

SCGA	Gramática	Ortografia	Léxico	Redações Zeradas
Não	Sim	Sim	Sim	Sim
Sim	Sim	Sim	Sim	Sim
Não	Sim	Sim	Sim	Não
Sim	Sim	Sim	Sim	Não
SCGA: Seleção de Características GA				

Para a Competência III realizamos os experimentos na Tabela 7. Os Experimentos foram utilizados afim de relacionar as *features* obtidas pelo sistema de NER. Também removemos as redações zeradas para remover as inconsistência causadas por estes tipos de redações.

Tabela 7 – Resumo dos Experimentos Realizados

SCGA	NER	Gramática	Redações Zeradas
Não	Não	Sim	Sim
Sim	Não	Sim	Sim
Não	Sim	Sim	Sim
Sim	Sim	Não	Sim
Não	Sim	Não	Não
Sim	Sim	Sim	Não
SCGA: Seleção de Características GA			

5.3 Experimentos para Avaliação da Competência I

Em nossa primeira sessão de experimentos queremos avaliar o modelo proposto para a avaliação da Competência I. Não utilizaremos as características de conteúdo pois a Competência I visa avaliar o conhecimento das regras que regem a língua portuguesa. Nossa hipótese é que existe uma correlação entre os atributos e as notas dadas a cada redação, e com a seleção de atributos mais importantes, igualar ou melhorar os resultados obtidos por Amorim e Veloso (2017), Fonseca, Medeiros e Kamikawachi (2018) e Júnior (2017) para a Competência I.

A análise exploratória das características da Tabela 8 mostra a relevância de algumas

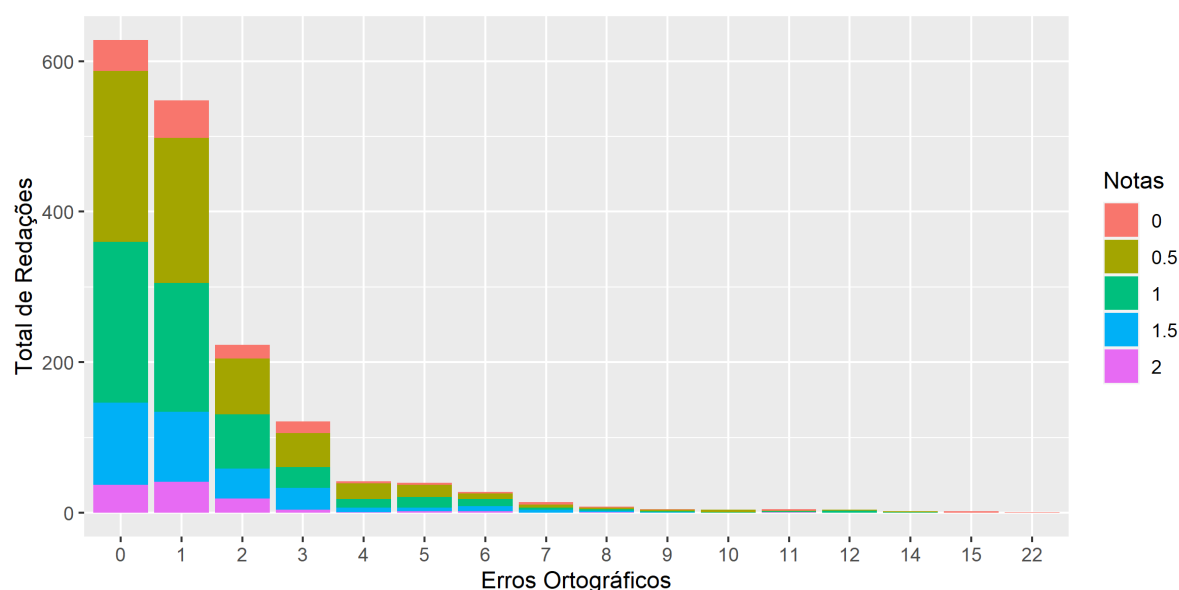
features extraídas para Competência I. Observamos que a média erros ortográficos tem uma relação inversa a nota da redação.

Tabela 8 – Propriedades do conjunto de redação para a Competência I

Notas	0	0.5	1	1.5	2
Redações	67	324	673	458	152
Média Palavras	220.55	236.68	264.52	290.30	298.66
Média Parágrafos	5.25	5.31	5.38	5.46	5.59
Média Erros Gramaticais	1.72	2.11	1.81	1.75	1.53
Média Erros Ortográficos	2.76	1.73	1.33	1.19	0.94

Ao aprofundar a análise sob a ótica dos erros ortográficos na Figura 9. Apesar das relações encontradas na Tabela 8 mesmo as redações que foram zeradas não tiveram nenhum erro ortográfico encontrado pelo Hunspell.

Figura 9 – Proporção de notas pela quantidade de erros ortográficos no texto.

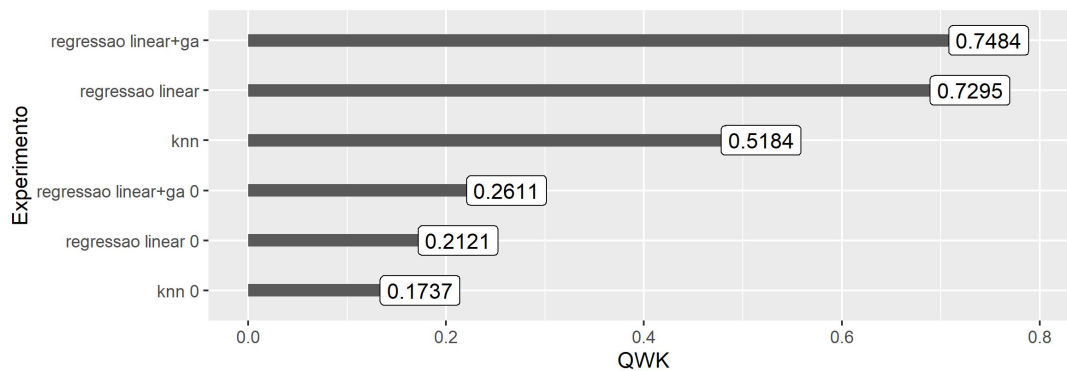


Essa pequena parte inconsistente é provocada principalmente pelo dicionário que não reconhece algumas palavras, ou que não reconhece palavras fora de contextos. Ainda há as redações que estão escritas corretamente, entretanto fogem ao tema proposto no enunciado, ou recaem em qualquer outra regra para zerar uma redação disposta na Sessão 1. Como consequências da utilização dessas redações na criação de um modelo, o modelo torna-se inconsistente ao dar notas as novas redações.

A Figura 10, a seguir, mostrar os resultados do experimento para a Competência I. Podemos observar a diferença entre modelos que utilizam as redações zeradas e sem essas redações no modelo.

Em um primeiro olhar sobre os resultados vemos que o Regressor tem um melhor resultado que as estratégias que utilizaram KNN, mesmo quando as redações zeradas

Figura 10 – Resultado dos experimentos propostos para Competência I



são removidas do *dataset*. Comparando as estratégias de Regressão Linear com seleção de característica com GA temos melhor resultado para a concordância com QWK com 0.7484.

5.4 Experimentos para Avaliação da Competência III

Passamos agora aos resultados obtidos com os resultado obtidos com a proposta de abordagem para a correção da Competência III. Nossa hipótese é que uma boa redação conter citações de pessoas, locais, fatos, dados ou acontecimentos para fundamentação da argumentação.

Para identificarmos essas entidades nomeadas utilizamos o modelo proposto por Pirovani (2019), que reconhece até 10 entidades nomeadas. Na nossa abordagens adotamos apenas 5 entidade nomeadas utilizadas no trabalho Pirovani et al. (2019), combinado ao modelo proposto no trabalho proposto por Alves et al. (2019).

Na Tabela 9 nós exibimos a média de NEs por cada classe de nota da redação modelo ENEM/INEP. Podemos notar uma sensível diferença entre a média de entidades entre as notas mais baixas e as mais altas.

Nossa premissa inicial é que para a Competência III tem pouca ou nenhuma relação com a ortografia ou com a gramática apresentada na redação, uma vez que sua temática é relacionada a apresentar o uso de dados e informações para tomada de decisão da situação problema apresentada.

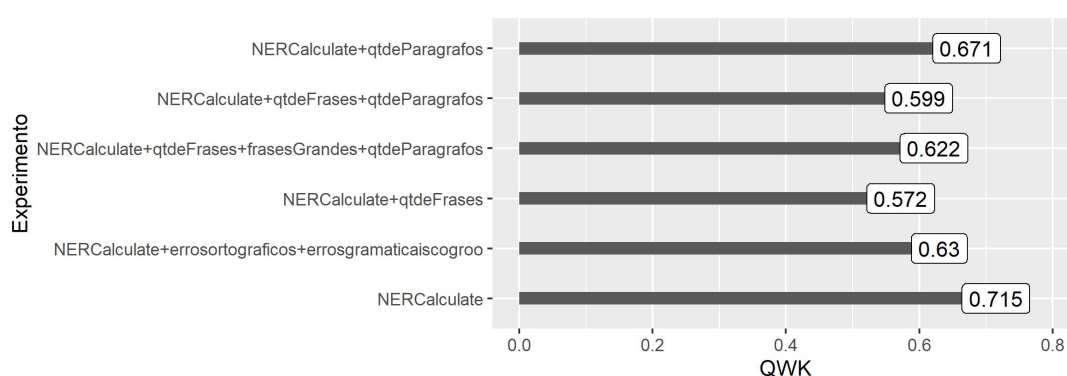
Durante a execução dos experimentos criamos algumas combinações de *features* na criação dos modelos. Na Figura 11 mostramos os resultados da execução dos experimentos da Competência III. Ao executar o experimento **NERCalculate**, que tem uma combinação entre as *features* gramaticas, ortográficas e o conteúdo da redação, com um $QWK = 0.715$.

A regra de avaliação da Competência III não demonstrar ter relação direta com a ortografia e com a gramática, entretanto ao executar os experimentos podemos notar que

Tabela 9 – Propriedades do conjunto de redação para a Competência III

Notas	0	0.5	1	1.5	2
Redações	142	598	531	296	108
Média Palavras	222.15	244.83	278.65	293.71	325.56
Média Parágrafos	5.30	5.29	5.45	5.55	5.57
Média Erros Gramaticais	1.89	1.88	1.79	1.69	1.95
Média Erros Ortográficos	1.99	1.37	1.29	1.36	1.26
Média Citação Pessoas	0.77	0.51	0.51	0.62	0.61
Média Citação de Tempo	0.52	0.49	0.52	0.61	0.74
Média Citação de Locais	0.80	0.93	0.94	1.17	1.23
Média Citação de Valores	0.37	0.38	0.44	0.46	0.48
Média Citação de Organizações	0.27	0.47	0.47	0.68	0.85

Figura 11 – Resultado dos experimentos propostos para Competência III



os resultados obtidos usando essas características, são os melhores resultados dentro os experimentos realizados.

5.5 Comparação com Abordagens Apresentadas na Literatura

Os resultados da abordagem do RL e do RL+GA foram comparados com outros três sistemas de AES para língua portuguesa. Ressaltamos que nenhum dos resultados apresentados nesse trabalho não levam em consideração o a distancia de 0.5 entre a nota do sistema e a nota do especialista apresentado no trabalho de Júnior (2017).

A primeira proposta a ser comparada é a abordagem de Júnior (2017). A abordagem usa a métrica MAE para mensurar a qualidade do AES. Como é possível observar na tabela 10 que os resultados obtidos com RL+GA e RL superaram os resultados obtidos pela abordagem Júnior (2017) com a utilização de PSO e SVM e usa como base de dados as redações UOL.

Também comparamos com os trabalho de Fonseca, Medeiros e Kamikawachi (2018) e Amorim, Cançado e Veloso (2018), entretanto a métrica de comparação que usaremos será o QWK. Na Tabela 11 comparamos os resultados obtidos e em se tratando de concordância

Tabela 10 – Comparação do Resultado com a métrica MAE para a Competência 1

Modelo	MAE
RL + GA	0.3042
RL	0.3089
Júnior (2017)	0.3107
KNN	0.3479

o modelo RL+GA os resultado se mostram superiores as abordagens da literatura.

Ressaltamos que o trabalho de Amorim, Cançado e Veloso (2018) utiliza a mesma fonte de redações UOL. Já o trabalho de Fonseca, Medeiros e Kamikawachi (2018) utiliza um banco de redações privados, não disponível para a comunidade.

Tabela 11 – Comparação do Resultado com a métrica QWK para a Competência 1

Modelo	QWK
RL + GA	0.7484
RL	0.7017
Fonseca, Medeiros e Kamikawachi (2018)	0.676
KNN	0.5184
Amorim, Cançado e Veloso (2018)	0.315

Na Tabela 12 compara os resultado obtidos para a Competência III o resultado obtido com a utilização do método RL+GA supera os métodos utilizados em outra abordagens. Por hora não existem trabalhos focados unicamente na Competência III, comparamos nossa proposta com trabalhos que tem tentam exprimir um modelo que avalie todos as competências com as mesmas *features*.

Tabela 12 – Comparação do Resultado com a métrica QWK para a Competência 3

Modelo	QWK
RL + GA	0.619
Fonseca, Medeiros e Kamikawachi (2018)	0.508
Amorim, Cançado e Veloso (2018)	0.2679

Os resultados de Fonseca, Medeiros e Kamikawachi (2018) levam em consideração um *dataset* com 16 mil redações, um conjunto superior aos utilizados neste e em outros trabalhos apresentados.

6 Conclusões e Trabalhos Futuros

Nesta sessão apresentamos um resumo das principais contribuições listadas na Sessão 1.

6.1 Conclusão

Mais de 5 milhões de redações são avaliadas durante o ENEM. O evento é visto por muitos estudantes como a oportunidade de adentrar no ensino superior. Com essa visão da importância da prova para os alunos é importante que os critérios de avaliação sejam os mais claros possíveis e que a correção seja imparcial e haja uma coerência mínima entre os avaliadores por conta das consequências financeiras e acadêmicas para os alunos caso a nota dada seja destoante ou insuficiente.

Primeiramente apresentamos uma proposta de modelo que busca dar continuidade ao trabalho de [Júnior \(2017\)](#). Como resultado deste novo modelo melhoramos o avaliador automático pelo modelo e também superamos resultados comparado a [Fonseca, Medeiros e Kamikawachi \(2018\)](#) que tem um *dataset* superior aos outros trabalhos que se propoem a automatizar o processo de correção de redações para a Competência I.

Por fim contribuimos com a avaliação da Competência III com a proposta inicial de avaliar se as redações submetidas contém elementos textos que denotem a presença de entidades nomeadas como *pessoas*, *organização*, *obras*, *acontecimentos*, entre outros. Para isso foram anotadas por humanos 110 redação afim de incluir um novo domínios textual utilizando as técnica proposta no trabalho de [Pirovani et al. \(2019\)](#).

A abordagem para correção da Competência III mostra resultados superiores aos obtidos em outros trabalhos ([AMORIM; CANÇADO; VELOSO, 2018](#); [FONSECA; MEDEIROS; KAMIKAWACHI, 2018](#)), obtendo como resultado $qwk = 0.6$ removendo do treinamento avaliações cujo a nota dada foi 0, e também um $qwk = 0.43$ levando em conta as redações zeradas.

6.2 Trabalhos Futuros

Por fim para melhorar os resultados apresentados algumas possibilidades para futuras pesquisas afim de ampliar o conhecimento sobre AES para a língua portuguesa:

- I. Avaliar massivamente os avaliadores que dão notas as redações, uma vez que suas discordâncias são evidenciadas por modelos estatísticos. A valiar quais avaliadores

colaboram com uma uniformização de avaliação melhora o conjunto de treinamento do modelo, uma vez que o papel de um avaliador automático é imitar um humano ao dar notas;

- II. aperfeiçoar a detecção de redações que fujam ao tema proposto, ou uma redação que apresente elementos que possa zera-la. Um modelo capaz extrair tais características ou inferir essa situação poderia servir de peso para o modelo atual;
- III. melhorar as ferramentas e extração de características tais como reconhecimento de entidades nomeadas em redações ampliando o conjunto de dados de treinamento e ampliando a biblioteca de gramatica local para esse domínio. Aprimorar ainda mais o reconhecimento de erros de concordâncias, gramaticais e ortográficos;

Referências

- ALFONSECA, E.; PÉREZ, D. Automatic assessment of open ended questions with a bleu-inspired algorithm and shallow nlp. In: SPRINGER. *International Conference on Natural Language Processing (in Spain)*. Berlin, Heidelberg, 2004. p. 25–35. Citado na página 35.
- ALVES, J.; OLIVEIRA, E. Avaliação de juízes: Um modelo estatístico para perfilação de avaliadores. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*. Distrito Federal, Brasília: [s.n.], 2019. v. 30, n. 1, p. 1653. Citado 2 vezes nas páginas 33 e 40.
- ALVES, J. et al. Attali, y., burstein, j. (2006). automated essay scoring with e-rater® v.2. journal of technology, learning, and assessment, 4(3). *Journal of Technology, Learning, and Assessment*, v. 4, 02 2019. Citado 3 vezes nas páginas 32, 39 e 56.
- AMORIM, E.; CANÇADO, M.; VELOSO, A. Automated Essay Scoring in the Presence of Biased Ratings. *Naacl*, p. 229–237, 2018. Citado 6 vezes nas páginas 36, 51, 53, 57, 58 e 59.
- AMORIM, E. C. F.; VELOSO, A. A Multi-aspect Analysis of Automatic Essay Scoring for Brazilian Portuguese. *Proceedings of the Student Research Workshop at the 15th Conference of the European Chapter of the Association for Computational Linguistics*, p. 94–102, 2017. Citado 7 vezes nas páginas 32, 35, 36, 37, 38, 53 e 54.
- ARTER, J. A.; CHAPPUIS, J. *Creating & Recognizing Quality Rubrics*. Nee York, USA: Pearson Education. Citado na página 40.
- ATTALI, Y.; BURSTEIN, J. Attali, y., burstein, j. (2006). automated essay scoring with e-rater® v.2. journal of technology, learning, and assessment, 4(3). *Journal of Technology, Learning, and Assessment*, v. 4, 02 2006. Citado 5 vezes nas páginas 32, 35, 36, 43 e 51.
- AZEVEDO, S. *Dicionário analógico da Língua Portuguesa*. 8. ed. São Paulo: Lexikon Editora, 1999. Citado na página 37.
- BAEZA-YATES, R.; RIBEIRO-NETO, B. *Modern Information Retrieval*. 2. ed. Harlow, England: Pearson Addison Wesley, 2011. Citado 2 vezes nas páginas 49 e 50.
- BURSTEIN, J. et al. Automated scoring using a hybrid feature identification technique. Vol. 1, 05 2002. Citado 2 vezes nas páginas 31 e 35.
- BUSSAB, W. O.; MORETTIN, P. A. *Estatística Básica*. São Paulo: Saraiva. Citado na página 33.
- CHAN, Y. S.; ROTH, D. Exploiting Syntactico-Semantic Structures for Relation Extraction. In: ASSOCIATION FOR COMPUTATIONAL LINGUISTICS. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*. Portland, Oregon, 2011. p. 551–560. Citado na página 31.

- CHEN, H.; HE, B. Automated essay scoring by maximizing human-machine agreement. *EMNLP 2013 - 2013 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, p. 1741–1752, 01 2013. Citado na página 32.
- COHEN, J. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, v. 20, n. 1, p. 37–46, 1960. Citado na página 33.
- CORTES, C.; VAPNIK, V. Support-vector networks. *Machine Learning*, v. 20, n. 3, p. 273–297, Sep 1995. Citado na página 37.
- DONG, F.; ZHANG, Y.; YANG, J. Attention-based recurrent convolutional neural network for automatic essay scoring. In: *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*. Vancouver, Canada: Association for Computational Linguistics, 2017. p. 153–162. Citado na página 37.
- FAULKNER, A. Automated classification of stance in student essays: An approach using stance target information and the wikipedia link-based measure. *Proceedings of the 27th International Florida Artificial Intelligence Research Society Conference, FLAIRS 2014*, p. 174–179, 2014. Citado 2 vezes nas páginas 35 e 51.
- FAZAL, A.; DILLON, T.; CHANG, E. Noise reduction in essay datasets for automated essay grading. In: SPRINGER. *OTM Confederated International Conferences "On the Move to Meaningful Internet Systems"*. Berlin, Heidelberg, 2011. p. 484–493. Citado na página 39.
- FILHO, A. H. et al. An approach to evaluate adherence to the theme and the argumentative structure of essays. *Procedia Computer Science*, Elsevier B.V., v. 126, p. 788–797, 2018. Citado 2 vezes nas páginas 37 e 38.
- FONSECA, E.; MEDEIROS, I.; KAMIKAWACHI, D. Automatically Grading Brazilian Student Essays. Springer International Publishing, v. 7243, p. 170–179, 2018. Citado 8 vezes nas páginas 32, 36, 38, 53, 54, 57, 58 e 59.
- FRIBURGER, N.; MAUREL, D. Finite-state transducer cascades to extract named entities in texts. *Theor. Comput. Sci.*, v. 313, p. 93–104, 2004. Citado na página 31.
- FRIEDMAN, J. H. Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, v. 29, p. 1189–1232, 2000. Citado na página 37.
- GOLDBERG, D. E. *Genetic Algorithms in Search, Optimization and Machine Learning*. 1st. ed. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1989. ISBN 0201157675. Citado na página 49.
- GOLDSCHMIDT, R.; PASSOS, E.; BEZERRA, E. *Data Mining: Conceitos, técnicas, algoritmos, orientações e aplicações*. 2st. ed. Rio de Janeiro, Rio de Janeiro: Elsevier, 2015. Citado na página 49.
- GROSS, M. A Bootstrap Method for Constructing Local Grammars. In: BOKAN, N. (Ed.). *Proceedings of the Symposium on Contemporary Mathematics*. Belgrado, Sérvia: University of Belgrad, 1999. p. 229–250. Citado na página 39.
- GUTIERREZ, F. et al. Hybrid ontology-based information extraction for automated text grading. In: IEEE. *Proceedings - 2013 12th International Conference on Machine Learning and Applications, ICMLA 2013*. Miami, Florida, 2013. p. 359–364. Citado na página 36.

- Hewlett Foundation. . 2012. Disponível em: <<https://www.kaggle.com/c/asap-aes>>. Acesso em: 05/02/2020. Citado na página 32.
- INEP. *Redação no ENEM 2018*. Brasília, 2018. Disponível em: <http://download.inep.gov.br/educacao_basica/enem/guia_participante/2018/manual_de_redacao_do_enem_2018.pdf>. Acesso em: 04/02/2019. Citado 4 vezes nas páginas 17, 23, 24 e 30.
- ISHIOKA, T.; KAMEDA, M. Automated japanese essay scoring system: jess. In: IEEE. *Proceedings. 15th International Workshop on Database and Expert Systems Applications, 2004*. Zaragoza, Spain, 2004. p. 4–8. Citado na página 35.
- JIANG, J. Information extraction from text. In: _____. *Mining Text Data*. Boston, MA: Springer US, 2012. p. 11–41. Citado na página 31.
- JÚNIOR, C. R. *Proposta de um Sistema Automático de Avaliação de Redações do Enem, Foco na Competência 1: Demonstrar Domínio da Modalidade Escrita Formal da Língua Portuguesa*. Dissertação (Mestrado) — Universidade Federal de Espírito Santo, 2017. Citado 10 vezes nas páginas 26, 27, 33, 37, 38, 53, 54, 57, 58 e 59.
- KE, Z. *Automated Essay Scoring: Argument Persuasiveness*. Tese (Doutorado), 2019. Citado na página 36.
- KINOSHITA, J.; SALVADOR, N.; MENEZES, C. E. D. CoGrOO – Um Corretor Gramatical acopl á vel ao OpenOffice. v. 0, p. 1737–1740. Citado na página 36.
- KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. v. 14, 2001. Citado na página 50.
- LAFFERTY, J.; MCCALLUM, A.; PEREIRA, F. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. In: *Proceedings of the Eighteenth International Conference on Machine Learning, ICML 2001*. San Francisco, CA, USA: ACM, 2001. v. 1, p. 282–289. Citado na página 39.
- LANDAUER, T.; LAHAM, D.; FOLTZ, P. Automated scoring and annotation of essays with the intelligent essay assessor. *Automated essay scoring: A cross-disciplinary perspective*, p. 87–112, 2003. Citado 2 vezes nas páginas 35 e 51.
- MATOS, D. A. S. Confiabilidade e concordância entre juízes : aplicações na área educacional. *Estudos em Avaliação Educacional*, v. 25, n. 59, p. 298–324, 2014. Citado na página 40.
- MILTSAKAKI, E.; KUKICH, k. Evaluation of text coherence for electronic essay scoring systems. *Natural Language Engineering*, v. 10, p. 25 – 55, 2004. Citado 2 vezes nas páginas 35 e 51.
- MUNIZ, M. A construção de recursos linguístico-computacionais para o português do brasil: o projeto unitex-pb. 2004. Citado na página 36.
- PAGE, E. B. Computer grading of student prose, using modern concepts and software. *The Journal of experimental education*, Taylor & Francis, v. 62, n. 2, p. 127–142, 1994. Citado 3 vezes nas páginas 35, 36 e 51.
- PIROVANI, J. et al. Adapting ner (crf+ lg) for many textual genres. 2019. Citado 3 vezes nas páginas 39, 56 e 59.

- PIROVANI, J.; SPALENZA, M.; OLIVEIRA, E. Geração automática de questões a partir do reconhecimento de entidades nomeadas em textos didáticos. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*. [S.l.: s.n.], 2017. v. 28, n. 1, p. 1147. Citado na página 32.
- PIROVANI, J. P. C. *CRF+ LG: Uma Abordagem Híbrida para o Reconhecimento de Entidades Nomeadas em Português*. Tese (Doutorado) — Universidade Federal de Espírito Santo, 2019. Citado 6 vezes nas páginas 15, 32, 39, 41, 44 e 56.
- PULMAN, S.; SUKKARIEH, J. Automatic short answer marking. 01 2005. Citado na página 30.
- QUINTILIANUS, M. F.; HALM, K. *Institutio oratoria*. Harlow, England: Teubner, 1869. v. 2. Citado na página 29.
- RAYMER, M. L. et al. Dimensionality reduction using genetic algorithms. *IEEE Transactions on Evolutionary Computation*, v. 4, n. 2, p. 164–171, 2000. Citado na página 49.
- SALTON, G.; WONG, A.; YANG, C.-S. A vector space model for automatic indexing. *Commun. ACM*, v. 18, p. 613–620, 1975. Citado na página 42.
- SILVA, W. D. C. de M. *Aprimorando o Corretor Gramatical CoGrOO*. Dissertação (Mestrado) — Universidade de São Paulo, 2013. Citado 5 vezes nas páginas 31, 36, 37, 44 e 45.
- UOL. __. 2019. Disponível em: <<https://educacao.uol.com.br/bancoderedacoes/>>. Acesso em: 20/07/2019. Citado na página 53.
- WANG, H.-C.; CHANG, C.-Y.; LI, T.-Y. Assessing creative problem-solving with automated text grading. *Comput. Educ.*, Elsevier Science Ltd., GBR, v. 51, n. 4, p. 1450–1466, 2008. Citado na página 35.
- ZESCH, T.; WOJATZKI, M. Task-independent features for automated essay grading. In: . [S.l.: s.n.], 2015. Citado na página 32.
- ZUPANC, K. *Semantics-based automated essay evaluation*. Tese (Doutorado) — University of Ljubljana, 2018. Citado 3 vezes nas páginas 35, 38 e 44.

Apêndices

APÊNDICE A – Regras Gramaticais do *CoGroo*

id	Categoria	Grupo	Mensagem
xml:1	Crase	à + substantivo masculino	Não ocorre crase antes de palavras masculinas.
xml:2	Crase	à + substantivo masculino	Não ocorre crase antes de palavras masculinas.
xml:3	Crase	à + substantivo feminino plural	Não há crase neste caso, somente no plural ("às").
xml:4	Crase	à + verbo	Não acontece crase antes de verbo.
xml:5	Crase	a + indicador de horas	Ocorre crase em expressões indicativas de horas.
xml:6	Crase	a + indicador de horas	Ocorre crase em expressões indicativas de horas.
xml:7	Crase	a + lado de	Ocorre crase em expressões como esta.
xml:8	Crase	a + lado de	Ocorre crase em expressões como esta.
xml:9	Crase	em relação a + substantivo feminino	"Em relação" rege a preposição "a", logo há crase aqui.
xml:10	Crase	em relação a + substantivo feminino	"Em relação" rege a preposição "a", logo há crase aqui.
xml:11	Crase	com relação a	"Com relação" rege a preposição "a", logo há crase aqui.

xml:12	Crise	com relação a	"Com relação"rege a preposição "a", logo há crase aqui.
xml:13	Crise	devido a	"Devido"rege a preposição "a", logo há crase aqui.
xml:14	Crise	devido a	"Devido"rege a preposição "a", logo há crase aqui.
xml:15	Crise	à + pronomes de tratamento	Pronomes de tratamento não admitem artigo, logo não há crase aqui.
xml:16	Crise	à + pronomes de tratamento	Pronomes de tratamento não admitem artigo, logo não há crase aqui.
xml:17	Concordância adjetivo-substantivo	substantivo + adjetivo	O adjetivo concorda com o substantivo a que se refere.
xml:18	Concordância adjetivo-substantivo	substantivo + adjetivo	O adjetivo concorda com o substantivo a que se refere.
xml:19	Concordância adjetivo-substantivo	substantivo + adjetivo	O adjetivo concorda com o substantivo a que se refere.
xml:20	Concordância adjetivo-substantivo	substantivo + adjetivo	O adjetivo concorda com o substantivo a que se refere.
xml:21	Concordância adjetivo-substantivo	adjetivo + substantivo	O adjetivo concorda com o substantivo a que se refere.
xml:22	Concordância adjetivo-substantivo	adjetivo + substantivo	O adjetivo concorda com o substantivo a que se refere.

xml:23	Concordância adjetivo-substantivo	adjetivo + substantivo	O adjetivo concorda com o substantivo a que se refere.
xml:24	Concordância adjetivo-substantivo	adjetivo + substantivo	O adjetivo concorda com o substantivo a que se refere.
xml:25	Em anexo	em anexo	A expressão "em anexo" é invariável.
xml:26	Em anexo	em anexo	A expressão "em anexo" é invariável.
xml:27	Em anexo	em anexo	A expressão "em anexo" é invariável.
xml:28	Em anexo	anexo como adjetivo	O adjetivo "anexo" concorda com o substantivo a que se refere.
xml:29	Em anexo	anexo como adjetivo	O adjetivo "anexo" concorda com o substantivo a que se refere.
xml:30	Em anexo	anexo como adjetivo	.
xml:31	Em anexo	anexo como adjetivo	.
xml:32	Em anexo	anexo como adjetivo	O adjetivo "anexo" concorda com o substantivo a que se refere.
xml:33	Em anexo	anexo como adjetivo	O adjetivo "anexo" concorda com o substantivo a que se refere.

xml:34	Em anexo	anexo como adjetivo	O adjetivo "anexo" concorda com o substantivo a que se refere.
xml:35	Em anexo	anexo como adjetivo	O adjetivo "anexo" concorda com o substantivo a que se refere.
xml:36	Em anexo	anexo como adjetivo	O adjetivo "anexo" concorda com o substantivo a que se refere.
xml:37	Em anexo	anexo como adjetivo	O adjetivo "anexo" concorda com o substantivo a que se refere.
xml:38	Uso de meio	meio no sentido de "um pouco"	A palavra "meio" usada no sentido de "um pouco" (advérbio) é invariável.
xml:39	Uso de meio	meio no sentido de "um pouco"	A palavra "meio" usada no sentido de "um pouco" (advérbio) é invariável.
xml:40	Uso de meio	meio como adjetivo	"Meio" usado no sentido de "metade" (numeral) concorda com o substantivo.

xml:41	Uso de meio	meio como adjetivo	"Meio" usado no sentido de "metade" (numeral) concorda com o substantivo.
xml:42	Verbo fazer	fazer indicando tempo	"Fazer", quando indica tempo, deve permanecer no singular.
xml:43	Verbo fazer	fazer indicando tempo	"Fazer", quando indica tempo, deve permanecer no singular.
xml:44	Verbo fazer	verbo auxiliar + fazer indicando tempo	"Fazer", quando indica tempo, deve permanecer no singular.
xml:45	Verbo fazer	verbo auxiliar + fazer indicando tempo	"Fazer", quando indica tempo, deve permanecer no singular.
xml:46	Verbo haver	haver + denotação de tempo	Opte pelo verbo "haver", no singular, para indicar tempo decorrido.
xml:47	Verbo haver	haver + denotação de tempo	Opte pelo verbo "haver", no singular, para indicar tempo decorrido.
xml:48	Verbo haver	haver + denotação de tempo	As formas do verbo "haver" ficam no singular quando indicam tempo decorrido.

xml:49	Verbo haver	haver + denotação de tempo	As formas do verbo "haver" ficam no singular quando indicam tempo decorrido.
xml:50	Verbo haver	haver no sentido de existir	"Haver" no sentido de existir é usado na 3a. pessoa do singular.
xml:51	Verbo haver	verbo auxiliar + haver no sentido de existir	"Haver" (existir), precedido de verbo aux., é usado na 3a. do singular.
xml:52	Emprego do mim e ti	mim + verbo no infinitivo	Use "eu" ao invés de "mim" como sujeito de verbo no infinitivo.
xml:53	Emprego do mim e ti	eu regido por preposição	O pronome "eu" não pode ser regido de preposição.
xml:54	Emprego do mim e ti	eu regido por preposição	O pronome "eu" não pode ser regido de preposição.
xml:55	Emprego do mim e ti	eu regido por preposição	O pronome "eu" não pode ser regido de preposição.
xml:56	Emprego do mim e ti	eu regido por preposição	O pronome "eu" não pode ser regido de preposição.
xml:57	Emprego de mau e mal	uso de mau	O contrário de "bem" é "mal", e o de "bom" é "mau".
xml:58	Emprego de mau e mal	uso de mal	O contrário de "bem" é "mal", e o de "bom" é "mau".

xml:59	Verbo preferir	preferir + redundância	"Preferir mais" é redundante. Suprima o "mais".
xml:60	Verbo preferir	regência do verbo preferir	Não use "preferir isso do que aquilo", mas sim "a aquilo".
xml:61	Colocação pronominal	palavras de sentido negativo + verbo + pronome oblíquo	Palavras negativas atraem o pronome átono para antes do verbo.
xml:62	Colocação pronominal	palavras de sentido negativo + verbo + pronome oblíquo	Palavras negativas atraem o pronome átono para antes do verbo.
xml:63	Colocação pronominal	palavras de sentido negativo + substantivo + verbo + pronome oblíquo	Palavras negativas atraem o pronome átono para antes do verbo.
xml:64	Colocação pronominal	palavras de sentido negativo + substantivo + verbo + pronome oblíquo	Palavras negativas atraem o pronome átono para antes do verbo.
xml:65	Colocação pronominal	pronome relativo ou conjunção subordinativa + verbo + pronome oblíquo	Pron. relativos e conj. subordinativas atraem o pronome para antes do verbo.
xml:66	Colocação pronominal	pronome relativo ou conjunção subordinativa + verbo + pronome oblíquo	Pron. relativos e conj. subordinativas atraem o pronome para antes do verbo.
xml:67	Colocação pronominal	advérbio + verbo + pronome oblíquo	Certos advérbios atraem o pronome para antes do verbo.

xml:68	Colocação pronominal	advérbio + verbo + pronome oblíquo	Certos advérbios atraem o pronome para antes do verbo.
xml:69	Colocação pronominal	advérbio + verbo + pronome oblíquo	Certos advérbios atraem o pronome para antes do verbo.
xml:70	Colocação pronominal	advérbio + verbo + pronome oblíquo	Certos advérbios atraem o pronome para antes do verbo.
xml:71	Colocação pronominal	pronome indefinido + verbo + pronome oblíquo	Alguns pronomes indefinidos atraem o pronome para antes do verbo.
xml:72	Colocação pronominal	pronome indefinido + verbo + pronome oblíquo	Alguns pronomes indefinidos atraem o pronome para antes do verbo.
xml:73	Colocação pronominal	só, ou, ora ou quer + pronome oblíquo	"só, ou, ora e quer"atraem o pronome para antes do verbo.
xml:74	Colocação pronominal	só, ou, ora ou quer + pronome oblíquo	"só, ou, ora e quer"atraem o pronome para antes do verbo.
xml:75	Se eu ver	conjugação de um verbo irregular no futuro do subjuntivo	Conjugação de um verbo irregular no futuro do subjuntivo.

xml:76	Se eu ver	conjugação de um verbo irregular no futuro do subjuntivo	Conjugação de um verbo irregular no futuro do subjuntivo.
xml:77	Se eu ver	conjugação de um verbo irregular no futuro do subjuntivo	Conjugação de um verbo irregular no futuro do subjuntivo.
xml:78	Crase	regência verbal	"Ir"constrói-se com prep. "a". Há crase com compl. feminino.
xml:79	Crase	regência verbal	"Aderir"constrói-se com prep. "a". Há crase com compl. feminino.
xml:80	Crase	regência verbal	"Pentecer"constrói-se com prep. "a". Há crase com compl. feminino.
xml:81	Crase	regência verbal	"Candidatar-se"constrói-se com prep. "a". Há crase com compl. feminino.
xml:82	Crase	regência verbal	"(Des)obedecer"constrói-se com prep. "a". Há crase com compl. feminino.
xml:83	Concordância do sujeito com o adjetivo predicativo	pronome + verbo de ligação + adjetivo predicativo	O adjetivo na função de predicativo concorda com o sujeito.

xml:84	Crise	Crise - regência de alguns nomes	Alguns nomes regem a preposição "a", logo há crase aqui.
xml:85	Crise	regência verbal - crise	"Reagir"constrói-se com prep. "a". Há crase com compl. feminino.
xml:86	regência verbal	regência do verbo obedecer/-desobedecer	"(Des)obedecer"constrói-se com preposição "a".
xml:87	Crise	à + pronomes pessoais	Não há crase porque alguns pronomes pessoais não admitem artigo.
xml:88	emprego de eu e mim	a + eu	"Eu"não deve ser preposicionado. Use "a mim".
xml:89	Crise	à + pronomes pessoais	Não há crase porque alguns pronomes pessoais não admitem artigo.
xml:90	regência verbal	regência do verbo namorar	Não use a preposição "com"na regência do verbo namorar.
xml:91	Crise	crise - regência de alguns nomes - compl. plural	Alguns nomes regem a preposição "a", logo há crase aqui.
xml:92	Concordância adjetivo-substantivo	meio-dia e meia	"Meia"concorda com o substantivo "hora", subentendido.

xml:93	Crase	regência verbal - crase	"Equivalente"rege a preposição "a". Com compl. feminino, há crase.
xml:94	Crase	regência verbal - crase	"Equivaler"constrói-se com prep. "a". Há crase com compl. feminino.
xml:95	Concordância Substantivo	artigo plural + substantivo singular	Os artigos concordam com o substantivo a que se referem.
xml:96	regência verbal	regência do verbo evitar, usufruir.	Os verbos "evitar"e "usufruir"não regem preposição "de".
xml:97	regência verbal	regência de demorar, torcer, votar	Regência do verbo incorreta.
xml:98	regência verbal	regência do verbo arrasar	O verbo "arrasar"não rege preposição "com".
xml:99	regência verbal	regência verbo habituar-se	Regência do verbo incorreta. Use "habituar-se a".
xml:100	regência verbal	regência habituar com pró-clise.	Regência do verbo incorreta. Use "habituar-se a".
xml:101	regência verbal	regência verbo habituar-se	Regência do verbo incorreta. Use "habituar-se a".
xml:102	regência verbal	regência verbo habituar-se	Regência do verbo incorreta. Use "habituar-se a".
xml:103	Concordância Substantivo	artigo singular + substantivo plural	Os artigos concordam com o substantivo a que se referem.

xml:104	Concordância Substantivo	Artigo-feminino + substantivo masculino	Os artigos concordam com o substantivo a que se referem.
xml:105	Concordância Substantivo	Artigo-masculino + substantivo feminino	Os artigos concordam com o substantivo a que se referem.
xml:106	à medida em que/à medida que	vícios de expressão	A expressão correta é "à medida que".
xml:107	regência verbal	verbo acarretar	É inadequado o uso da preposição "em".
xml:108	Crase	segunda a sexta	Não há crase pois o "a" é apêndice nas preposições.
xml:109	regência verbal	assistir com o sentido de presenciar.	Há crase pois o verbo "assistir" (presenciar) rege preposição "a".
xml:110	regência nominal	valorização de	"Valorização" rege preposição "de" e não preposição "a".
xml:111	emprego de vírgulas	expressões entre vírgulas	A expressão "ou seja" deve ser isolada por vírgulas.
xml:112	emprego de vírgulas	expressões entre vírgulas	A expressão "ou seja" deve ser isolada por vírgulas.
xml:113	emprego de vírgulas	expressões entre vírgulas	Deve haver vírgula antes de "no entanto".
xml:114	Concordância Determinante-Substantivo	determinante singular + substantivo plural	Os determinantes concordam com o substantivo a que se referem.

xml:115	Concordância Numeral-Substantivo	Concordância Substantivo	Numeral-	Os numerais concordam com o substantivo a que se referem.
xml:117	Concordância sujeito-verbo	sujeito plural + verbo singular		Verificou-se erro de concordância entre o sujeito e o verbo.
xml:118	Concordância sujeito-verbo	sujeito singular + verbo plural		Verificou-se erro de concordância entre o sujeito e o verbo.
xml:119	Verbo haver	haver no sentido de existir		"Haver" no sentido de existir é usado na 3a. pessoa do singular.
xml:120	Verbo haver	verbo auxiliar + haver no sentido de existir		"Haver" (existir), precedido de verbo aux., é usado na 3a. do singular.
xml:121	Gerundismo	Gerundismo		Verifique o excesso de verbos em sequência.
xml:122	Uso do verbo haver	Redundância semântica		"Atrás" é redundante, devido ao uso do verbo "haver".
xml:123	Expressões redundantes	Redundância semântica		Expressões redundante. Suprima "junto" ou "juntos".
xml:124	Concordância do sujeito com o adjetivo predicativo	pronome + verbo de ligação + adjetivo predicativo		O adjetivo na função de predicativo concorda com o sujeito.

xml:125	Concordância do sujeito com o verbo do predicado	núcleo do sujeito + verbo	O sujeito concorda em número com o predicado.
xml:126	Crase	à + pronomes de tratamento	Pronomes de tratamento não admitem artigo, logo não há crase aqui.
xml:127	Concordância do sujeito com o predicativo	pronome + verbo de ligação + predicativo	O adjetivo na função de predicativo concorda com o sujeito.
xml:128	Concordância do sujeito com o adjetivo predicativo	pronome + verbo de ligação + adjetivo predicativo	O adjetivo na função de predicativo concorda com o verbo.
xml:129	Concordância sujeito-verbo	sujeito + verbo + .	Verificou-se erro de concordância entre o sujeito e o verbo.
government:government	Erros sintáticos	Regência verbal	Regência verbal.
probs:paronyms	Enganos ortográficos	Ortografia	Possível confusão entre %s e %s.
punctuation:before sentences	Erros mecânicos	Pontuação	Pontuação no início da frase.
punctuation:extra punctuation	Erros mecânicos	Pontuação	Excesso de sinais de pontuação.
repetition:duplicated token	Erros mecânicos	Repetição de palavras	Repetição de palavras.

space:extra	after	left	Erros mecânicos	Uso do espaço	Excesso de espaço depois de símbolo.
punct					
space:extra	before		Erros mecânicos	Uso do espaço	Excesso de espaço antes de símbolo.
right punct					
space:extra	between		Erros mecânicos	Uso do espaço	Excesso de espaços entre as palavras.
words					
space:missing	space af-		Erros mecânicos	Uso do espaço	Falta espaço entre símbolo e palavra à direita.
ter punct					