

PPGI | PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

Universidade Federal do Rio de Janeiro

DISSERTAÇÃO DE MESTRADO

Debora Ramalho Barros

JOGO DO JORNAL: UM MODELO DE JOGO PARA O LETRAMENTO



Instituto de Matemática



Instituto Tércio Pacitti de Aplicações
e Pesquisas Computacionais



Universidade Federal
do Rio de Janeiro



Universidade Federal Do Rio De Janeiro
Programa de Pós-Graduação em Informática

Debora Ramalho Barros

Jogo do Jornal: Um modelo de jogo para o letramento

Programa de Pós-Graduação em Informática
PPGI-UFRJ

Orientadoras:

Profa: Claudia Lage Rebello da Motta, D.Sc.

Profa: Leonor Werneck dos Santos, Ph.D

Rio de Janeiro, RJ – Brasil

2011

Debora Ramalho Barros

***JOGO DO JORNAL: UM MODELO DE JOGO PARA O
LETRAMENTO***

Dissertação de Mestrado, apresentada ao Programa de Pós-Graduação em Informática (PPGI), Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Informática.

Orientadoras:

Prof^a. Claudia Lage Rebello da Motta, D.Sc.

Prof^a. Leonor Werneck dos Santos, Ph.D

Rio de Janeiro, RJ – Brasil

2011

B277 Barros, Debora Ramalho.

Jogo do Jornal: Um modelo de jogo para o letramento/ Debora Ramalho Barros.. – Rio de Janeiro, 2011.
páginas f.150: il.

Dissertação (Mestrado em Informática) – Universidade Federal do Rio de Janeiro, Instituto de Matemática, Técio Parcitti de aplicações e pesquisas computacionais, Programa de Pós-graduação em Informática 2011.

Orientador: Claudia Lage Rebello da Motta, Leonor Werneck dos Santos.

1. Neuropedagogia – Teses, 2. Elaboração dirigida e mineração de texto – Teses, I. Motta, Claudia Lage Rebello da (Orient.), II. Santos, Leonor Werneck, III. Universidade Federal do Rio de Janeiro. Instituto de Matemática. Núcleo de Computação Eletrônica. IV. Título.

CDD

Debora Ramalho Barros

JOGO DO JORNAL: UM MODELO DE JOGO PARA O LETRAMENTO

Dissertação de Mestrado, apresentada ao Programa de Pós-Graduação em Informática (PPGI), Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Informática.

Rio de Janeiro, 19 de Dezembro de 2011.

Aprovada por:

Profa. Claudia Lage Rebello da Motta, D.Sc., iNCE e PPGI/UFRJ (Presidente)

Profa. Leonor Werneck do Santos, Ph.D, Letras/UFRJ

Profa. Jonice de Oliveira Sampaio, Dsc., UFRJ

Carlo Emmanuel, Ph.D, iNCE e PPGI/UFRJ

Profa: Fernanda Araújo Baião, D.Sc, UNIRIO

Dedicatória

Aos meus alunos: meus grandes incentivadores, minha família que são meus pilares, incluindo parentes, colegas e amigos e a professora Carla Verônica Marque que plantou uma semente dentro de mim, que agora eu sinto, germinou, cresceu, virou botão e floresceu. Aquela que foi o pegureiro das minhas ações, das minhas determinações. Sua voz que me levantou nas horas difíceis, que me deu novas forças, que mostrou que cada dia é uma nova renovação.

Agradecimentos

Em primeiro lugar, agradeço enormemente aos meus pais, Álvaro Sérgio e Denise Ramalho, por me darem sempre total apoio em todos os momentos de minha vida e contribuírem muito para minha aptidão à pesquisa.

Agradeço ao meu marido, Paulo Gustavo Tavares, pela paciência e apoio incondicional, e por ter transformado todo o difícil e longo caminho em algo mais prazeroso de ser vivido.

Agradeço as minhas orientadoras Claudia Motta, Leonor Werneck dos Santos e aos professores Carla Verônica Marquez, Carlo Emmanuel Tolla de Oliveira e Marcos Elia, pelas orientações e incentivo mesmo nos momentos que pareciam mais difíceis, fazendo com que todas as dúvidas se esclarecessem.

Agradeço aos meus sogros, Paulo e Dina, que me receberam em sua casa durante essas minhas idas e vindas à UFRJ e são pessoas maravilhosas.

Agradeço às minhas irmãs pelo amor e carinho que sempre tiveram comigo. Agradeço à colega de curso Samantha Vrabl pelo companheirismo, apoio e incentivos sempre presentes ao longo dessa trajetória. Agradeço a minha equipe da Cultura Inglesa pelo apoio para a realização desse trabalho.

Agradeço aos programadores e estagiários que me ajudaram a desenvolver essa pesquisa em diferentes momentos, Ronald Kaiser, Maíta Carvalho e Damien Challou, pela dedicação, incentivo e liberação de horas para realização deste trabalho. Por fim, agradeço a Deus por tudo que tenho e por ter me suprido a ótima saúde e coragem de que tanto precisei nesta gloriosa etapa da minha vida e a enfrentar os desafios para realizar um projeto de pesquisa em uma área diferente daquela da minha formação original.

A realização dessa pesquisa é mais uma prova de que trabalhar interdisciplinarmente é tão prazeroso quanto desafiador. Minha trajetória de vida faz lembrar a frase de S. Francisco de Assis: "Comece por fazer o que é necessário, depois faça o que é possível e em breve estará fazendo o que é impossível". Pois realizar este trabalho muitas vezes me pareceu impossível e ambicioso por demais.

RESUMO

BARROS, Débora Ramalho. *Jogo do Jornal: Um modelo de jogo para o letramento*. Rio de Janeiro, 2011. Dissertação (Mestrado em Informática) – Programa de Pós-Graduação em Informática, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2011.

Jogos propiciam uma profícua maneira de lidar com problemas de letramento, definida por Soares (2000) como um “estado ou condição que adquire um grupo social ou um indivíduo como consequência de ter-se apropriado da escrita” e complementa: “Letramento é prazer, é lazer, é ler em diferentes lugares e sob diferentes condições.”. Os princípios da Web 2.0 de autonomia, flexibilidade, interação social, criatividade, e ubiquidade, configuradas com a questão relevante sobre o letramento no Brasil e o uso de jogos neuropedagógicos, levaram-nos a desenvolver o Jogo do Jornal, uma ferramenta que guarda as características de alguns gêneros textuais, afim de identificá-los automaticamente na escrita daqueles que se propõem a utilizar a ferramenta para o desenvolvimento da leitura e da escrita, contribuindo para contenção do fracasso escolar. O jogo se insere como um projeto de dissertação de Mestrado, que articula conteúdos interdisciplinares de psicolinguística, linguística textual, sociolinguística, sociologia, psicopedagogia cognitiva e computação. Esse conjunto de saberes permite compor novas possibilidades de trabalho em sala de aula e propulsiona o uso da atividade lúdica. O jogo aqui apresentado é uma ferramenta que auxiliará o trabalho do professor dentro e fora de sala de na tarefa de produção e avaliação textual de seus alunos num universo finito de gêneros textuais. Para a avaliação foram realizados dois estudos de caso que visaram reproduzir com alunos e professores uma simulação de utilização real desse jogo em sala de aula.

Palavras-chave: letramento, gênero textual, identificação automática de texto, jogos neuro pedagógicos.

ABSTRACT

BARROS, Débora Ramalho. *Jogo do Jornal: Um modelo de jogo para o letramento*. Rio de Janeiro, 2011. Dissertação (Mestrado em Informática) – Programa de Pós-Graduação em Informática, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2011.

Games provide a useful way to deal with problems of literacy, defined by Soares (2000) as "state or condition from a social group or an individual as a result of having taken up writing". She adds: "Literacy is pleasure, and leisure, and reading in different places and under different situations". The Web 2.0 principles of autonomy, flexibility, social interaction, creativity, and ubiquity, configured with the issue about literacy in Brazil, and the use of neuropedagogics games, led us to develop the newspaper game, whose objective is to evaluate and develop reading and writing through cognitive stimulation, contributing to containment failure in school. The developing of this game combines interdisciplinary content of psycholinguistics, text linguistics, sociolinguistics, sociology, educational psychology and cognitive computing. This set of knowledge allows composing new writing opportunities in the classroom and the use of a playful activity. The game here presented is a tool which shall assist the teacher's work in and out of classroom with textual production and evaluation of their students texts in a finite universe of text genres. For the evaluation, two case studies were carried out in real classroom environment with teacher-student interaction.

Keywords: literacy, genre, automatic identification of text, neuro educational games.

LISTA DE FIGURAS

Figura 1.1.: Campanhas de incentivo ao Letramento.....	19
Figura 1.2: Organograma da dissertação.....	16
Figura 2.1: Influência dos teóricos na teoria de Seminário.....	20
Figura 2.2: Mapa Conceitual do Conceito.....	31
Figura 2.3: Mapa explicativo sobre Metalinguagem.....	34
Figura 2.4: Definição de Metaprocesso.....	35
Figura 2.5: Mapa Conceitual da Técnica.....	36
Figura 2.6: Alternativas de trabalho.....	37
Figura 2.7. Técnica "before-after".....	38
Figura 2.8: Fases do Fio Condutor.....	39
Figura 3.1: Etapas do Processo de Mineração de Textos.....	43
Figura 3.2: Técnica <i>Stemming</i>	46
Figura 3.3 Processo de Categorização.....	49
Figura 4.1 Etapas da Metodologia proposta.....	74
Figura 5.1 Textanalyser: Detalhe do processo de extração de texto.....	77
Figura 5.2: arquitetura do TextSeeker.....	79
Figura 5.3: Diagrama UML de caso de uso.....	82
Figura 5.4: Diagrama de classes Conceitual do protótipo.....	83
Figura 5.5: Cadastro do administrador.....	84
Figura 5.6: Cadastro do aluno.....	85
Figura 5.7: Cadastro do Professor.....	86
Figura 5.8: Tela inicial do Jogo do Jornal.....	88
Figura 5.9: Tela de login do professor.....	89
Figura 5.10: Tela da tarefa de management.....	89

Figura 5.11: Tela de login de aluno.....	90
Figura 5.12: Tela de login de aluno.....	90
Figura 5.13: Tela da tarefa de Results.....	91
Figura 5.14: Tela da tarefa de Results.....	92
Figura 5.15: Tela da tarefas do administrador.....	93
Figura 6.1: Estudo de Caso Piloto.....	94
Figura 6.2: Jornal Legal, da aluna, 9º ano instituição particular.....	95
Figura 6.3: Texto antes da intervenção dialogada.....	96
Figura 6.4: Texto depois da intervenção dialogada.....	102

LISTA DE QUADROS

Quadro 3.6: <i>inputs</i> de notícia.	64
Quadro 3.7: <i>inputs</i> de artigo.	66
Quadro 3.8: <i>inputs</i> de receita.	67
Quadro 3.9: <i>inputs</i> de horóscopo.	67
Quadro 3.10: <i>inputs</i> de crônica.	69
Quadro 3.11: <i>inputs</i> de manchete.	70
Quadro 3.12: <i>inputs</i> de resenha crítica.	71
Quadro 3.13: Critérios de gêneros.	72
Quadro 3.14: Perguntas de mediação para cada critério dos gêneros.	73
Quadro 3.1: Conteúdo Temático.	68
Quadro 3.2: Estrutura Composicional.	69
Quadro 3.3: Objetivos e funções sociocomunicativas.	70
Quadro 3.4: Características da superfície linguística.	71
Quadro 3.5: Condições de produção.	72
Quadro 6.6: Produção de textos por alunos e gêneros.	108
Quadro 6.7: Intervenção dialogada conteúdo temático.	108
Quadro 6.8: Intervenção dialogada estrutura composicional.	109
Quadro 6.9: Intervenção dialogada dos objetivos e funções sociocomunicativas.	131
Quadro 6.10: Intervenção dialogada características da superfície linguística.	107
Quadro 6.11: Intervenção dialogada da condição de produção.	107
Quadro 6.12: Resultado da superfície linguística.	108
Quadro 6.13: Resultado da estrutura composicional.	108
Quadro 6.14: Resultado do conteúdo temático.	108
Quadro 6.15: Resultado das condições de produção.	108
Quadro 6.16: Resultado dos objetivos e funções sociocomunicativas.	109

Quadro 6.17: Resultado da intervenção	109
Quadro 6.19: Gêneros produzidos pelo E.M	107
Quadro 6.20: Gêneros produzidos pelo E.F. I	107
Quadro 6.21: Gêneros produzidos pelo E.F. I	108
Quadro 6.22: Gêneros produzidos pelo E.F. II	108
Quadro 6.23: Gêneros produzidos pelo E.F. II	108
Quadro 6.24: gêneros produzidos pelo E.F.I	108
Quadro 6.25: gêneros produzidos pelo E.F. I	109
Quadro 6.26: Gêneros produzidos pelo E.M	107
Quadro 6.27: Gêneros produzidos pelo E.F. I	107
Quadro 6.28: Gêneros produzidos pelo E.F. I	108
Quadro 6.29: Gêneros produzidos pelo E.F. II	108
Quadro 6.30: Resposta ao questionario	108

SUMÁRIO

1 INTRODUÇÃO	15
1.1 Contexto	15
1.2 Problema	16
1.3 Relevância.....	17
1.4 Objetivo	17
1.5 Hipótese	17
1.6 Organização da Dissertação	18
2 REFERENCIAL TEÓRICO	20
2.1 Letramento	21
2.2 Gêneros Textuais e Ensino	24
2.3 Concepções de Aprendizagem	28
2.4 Técnica metacognitiva da Elaboração Dirigida.....	30
2.4.1 Metacognição	32
2.4.2 Linguagens-Código: L3 - L4	33
2.4.3 Metalinguagem.....	33
2.4.4 Estocagem de Modelos.....	34
2.4.5 Diálogo e Motivação.....	34
2.4.6 Metaprocesso Cognitivo.....	35
2.4.7A técnica da Elaboração Dirigida, seu legado e procedimentos.....	36
2.5 Fio Condutor.....	39
2.6 Identificação automática de texto	40
2.7 Mineração de Texto.....	42
2.7.1 Etapas no processo de Mineração	42
2.7.2 Coleta de Documentos	43
2.7.3 Pré-processamento	43
2.7.3.1 Remoção de caracteres indesejados	44
2.7.3.2 Transformação das letras (Case Folding)	44
2.7.3.3 Toqueização	44
2.7.3.4 Retirada de palavras desnecessárias (Stopwords)	45
2.7.3.5 Redução ao menor radical de cada palavra (Stemming).....	45
2.7.3.6 Dicionário de dados (Thesaurus)	46
2.7.4 Extração de Padrões.....	47
2.7.5 Avaliação e Interpretação dos Resultados	47
2.8 Tarefas no processo de Mineração	47
2.8.1 Categorização de Textos.....	48
2.8.1.2 Classificador Bayesiano	49
2.8.1.3 Classificador ORANGE	50
2.8.2 Agrupamento.....	51

3. A CONSTRUÇÃO DO OBJETO DE MEDIDA.....	53
3.1 Indentificando o Gênero	54
3.1.2 Critérios Estruturais	54
3.1.3 Critérios Lexicais	54
3.1.4 Critérios Derivacionais.....	54
3.1.5 Características da superfície linguística	55
3.1.6 Notícia	63
3.1.7 Artigo	64
3.1.8 Receita	66
3.1.9 Horóscopo	67
3.1.10 Crônica	68
3.1.11 Manchete.....	69
3.1.12 Resenha	70
3.1.13 Implementação	71
4 METODOLOGIA.....	74
4.1 Pré-Processamento.....	75
4.2 Processamento dos dados	75
4.3 Extração dos dados.....	76
4.4 Categorização	76
4.5 Pós-processamento	76
5 AMBIENTE COMPUTACIONAL.....	77
5.1 Arquitetura Geral do Sistema	77
Analizador de Texto ou TextAnalyser.....	78
5.2 Modelagem do Sistema.....	80
5.2 Diagrama de caso de uso.....	81
5.2.1 Aplicar o Jogo do Jornal	81
5.2.3 Carregar diretório de texto.....	82
5.2.4 Inserir texto no banco	82
5.2.5 Jogar	83
5.3 Diagrama de caso de Classe	83
5.4 Especificações do Protótipo	83
5.4.1 Cadastro de Administrador.....	84
5.5.3 Cadastro de Aluno.....	85
5.4.3 Cadastro de Professor.....	85
5.4.4 Algoritmo	86
5.4.5 O rastreador ou Crawler	87
5.5 Ferramentas utilizadas	87
5.6 Visão Geral do protótipo.....	88
5.7 Cadastro dos Usuários	88
5.8 Jogar o Jogo do Jornal	90
5.9 Tarefas do Professor.....	91
5.10 O Administrador	92
6 ESTUDO DE CASO	94
6.1 Estudo de caso piloto	94
6.1.1 Resultados	96
6.1.2 Escopo	96
6.1. 3 Perguntas de intervenção ou API.....	97
6.1.4 Critérios selecionados para composição do crivo	100
6.1.5 Conclusão	103
6.2 Estudo de caso Final.....	104

6.2.1 Eficiência da Ferramenta.....	105
6.2.2 Relevância dos critérios	107
6.2.3 Efeito na Aprendizagem	109
6.2.4 Eficiência das recomendações automáticas.....	111
6.2.5 Conclusão do estudo de caso Final	111
REFERÊNCIAS	117
ANEXOS	122
6. Ferramentas utilizadas	127
6.1 Linguagem de programação Python	127
6. Bibliotecas Feedparser e BeautifulSoup	127
6.3 Naive Bayesian classification	128
6.4 Tornado web server + nginx.....	128
6.5. Library Orange the input file	129

1 INTRODUÇÃO

“Gosto de ser gente porque a História em que me faço com os outros e de cuja feitura tomo parte é um tempo de possibilidades e não de determinismos.”

Paulo Freire, Pedagogia da autonomia, 1997

Nesse Capítulo é apresentada a pesquisa realizada no âmbito dessa dissertação, abordando o contexto de sua realização, o problema e a hipótese, os objetivos da dissertação e a organização do texto.

1.1 Contexto

Atualmente o Letramento é preconizado pelos documentos oficiais (PCN e PCNEM, PISA, ENEM etc.) “A escola tem sua prática de produção textual sedimentada nessas concepções, ver-se uma oportunidade de trabalhar essa deficiência na produção textual” (Silva, 2008), mas como dar conta de letrar os alunos de sociedades de cultura predominantemente oral e sob forte influência de fatores de exclusão? “(...) é necessário reconhecer o fracasso da escola e, no interior desta, do ensino da língua portuguesa (e de produção textual) tal como vem sendo praticado na quase totalidade de nossas aulas... sabemos que a educação tem muitas vezes sido relegada à inércia administrativa, a professores mal pagos e mal remunerados, a verbas escassas e aplicadas com falta de racionalidade.” (Herreira, 2000). Muitos autores acreditam que a solução está no trabalho com produção textual em sala de aula.

Essa discussão tem sido assunto recorrente dos principais veículos de comunicação sobre educação nacional, como a revista Brasil Escola. A Figura 2.1 trás fotos tanto dessa revista como do programa de Formação Continuada de professores das séries iniciais do ensino fundamental, ambos fazendo alusão à urgência desse tema.



Figura 1.1: Campanhas de incentivo ao Letramento
Fonte: portal MEC.gov

1.2 Problema

Os textos escritos produzidos funcionam como artefatos sociais e culturais. A nossa familiaridade com textos escritos evidencia nossa participação em comunidades. O acesso social a estruturas e comunidades a que desejamos pertencer é mediado pelo uso efetivo e competente do processo literato da leitura e da escrita. “A literacia é o uso de práticas situadas no contexto social, histórico e cultural que nos permite criar e interpretar significados através do uso de textos” (Kern, 2000p. 16).

A literária é um dos aspectos fundamentais da participação social. No Brasil, no entanto, as avaliações realizadas pelos diferentes órgãos oficiais têm demonstrado que a maioria dos estudantes termina o ensino médio com dificuldade para ler um texto de média complexidade e para redigir textos adequadamente.

A partir do cenário mostrado, defrontamo-nos com as seguintes questões: Como deverá ser feito o trabalho de letramento na sala de aula? Quais aspectos da língua devem ser observados ao elaborar uma atividade para esse fim? É de suma importância observar que levar em consideração somente a exposição dos alunos a diferentes tipos de gêneros textuais através da leitura descomprometida e a tarefa de produção desses gêneros é talvez mais eficiente para sua compreensão e aprendizado do que análise de composição das características língüísticas desses gêneros.

Um problema do ensino é o tratamento inadequado, para não dizer desastroso, que o texto recebe. Mas o problema não reside somente nas formas de acesso ao texto, e sim, na forma de sua apresentação.

1.3 Relevância

Como resultado dessa pesquisa, espera-se que o tema do letramento através do gênero textual contribua para o desenvolvimento da qualidade de vida dos seres humanos, podendo até ser considerada como uma minúscula contribuição para este fim. No entanto, como o relatório do IPEA sugere: “vê-se que mesmo pequenas diminuições no grau de desigualdade poderiam reduzir a pobreza significativamente” (TD 1000). Instrumentalizar o professor e torná-lo capaz de mediar às tensões, negociar significados e construir novos contextos de inserção social de maneira mais prática e divertida. Para aperfeiçoar um trabalho sem o auxílio da ciência da computação, seria enfadonho, e muitos, devido a esse alto grau de esforço, não conseguem incluir em suas práticas e alcançar um grande número de alunos.

1.4 Objetivo

O objetivo principal dessa dissertação é o desenvolvimento de um modelo de utilização do conhecimento da lingüística textual, que possibilite a construção de uma ferramenta computadorizada que auxilie na classificação automática de textos de acordo com seu gênero e promova a vivência com gêneros textuais de maneira moderna e prazerosa, dentro e fora do contexto de sala de aula. Tendo como propósito uma análise da produção textual rápida, objetivando sua intervenção para o letramento, independente do número de alunos que estiverem em sala.

Os objetivos específicos são a implementação dos mecanismos propostos que pretendem validar a hipótese desta pesquisa, mapeando e combinando os seguintes pontos:

- Análise, teste e relevância das características exclusivas de cada gênero estudado;
- Intervenção na aprendizagem e produção desses;
- Os fatores responsáveis pela efetivação da aprendizagem por intermédio da ferramenta do jogo.

1.5 Hipótese

No processo de produção, análise e compreensão de gêneros textuais, o estudo da interação da linguagem é essencial para se entender o funcionamento da linguagem. O ensino de língua hoje acontece através de textos, essa é orientação central dos PCNs. A questão não reside no consenso ou na aceitação deste postulado, mas no modo como isso é posto em prática, já que existem muitas formas de se trabalhar o texto.

Neste contexto, este estudo possibilitará a construção de uma ferramenta lúdica para o letramento, e minha hipótese é a de que se é possível organizar as características determinantes de cada gênero textual à luz das teorias de lingüística textual, então é possível projetar uma ferramenta para a identificação automática desses para um fim educativo, voltada para o ensino de gênero

textual e o letramento, possibilitando que todas as vezes que utilizada repercutirá sobre a escrita, e ao mesmo tempo alimentar-se dela.

1.6 Organização da Dissertação

Este trabalho está dividido em sete capítulos da seguinte forma:

- O Capítulo 1 apresenta a motivação, que serviu como base para a pesquisa de uma metodologia de descoberta de gêneros textuais em dados disponíveis em texto, bem como toda uma justificativa para a construção de um analisador automático de gêneros para auxiliar no letramento de alunos do ensino fundamental e médio;
- O Capítulo 2 apresenta a teoria em que se baseia a importância da diversidade textual para o letramento, as teorias da linguística textual que auxiliaram na construção da ferramenta. Como também a fundamentação teórica escolhida para a aplicação dessa e as técnicas que foram selecionadas na implementação do sistema; ·
- O Capítulo 3 trata da construção do objeto de medida e define as etapas necessárias para desenvolver as tarefas de classificação de uma coleção de documentos;
- O Capítulo 4 trata da metodologia proposta e define as etapas necessárias o desenvolvimento dessa pesquisa.
- No Capítulo 5, o ambiente computacional é descrito e detalhado, assim como também, toda a modelagem da ferramenta e sua implementação.
- O Capítulo 6 apresenta dois estudos de casos realizados, o que demonstra a utilização do sistema, os resultados obtidos. E Finalmente, as considerações finais, definindo as conclusões do trabalho e sugestões de desenvolvimentos futuros que podem ser efetuados estão no Capítulo 7.

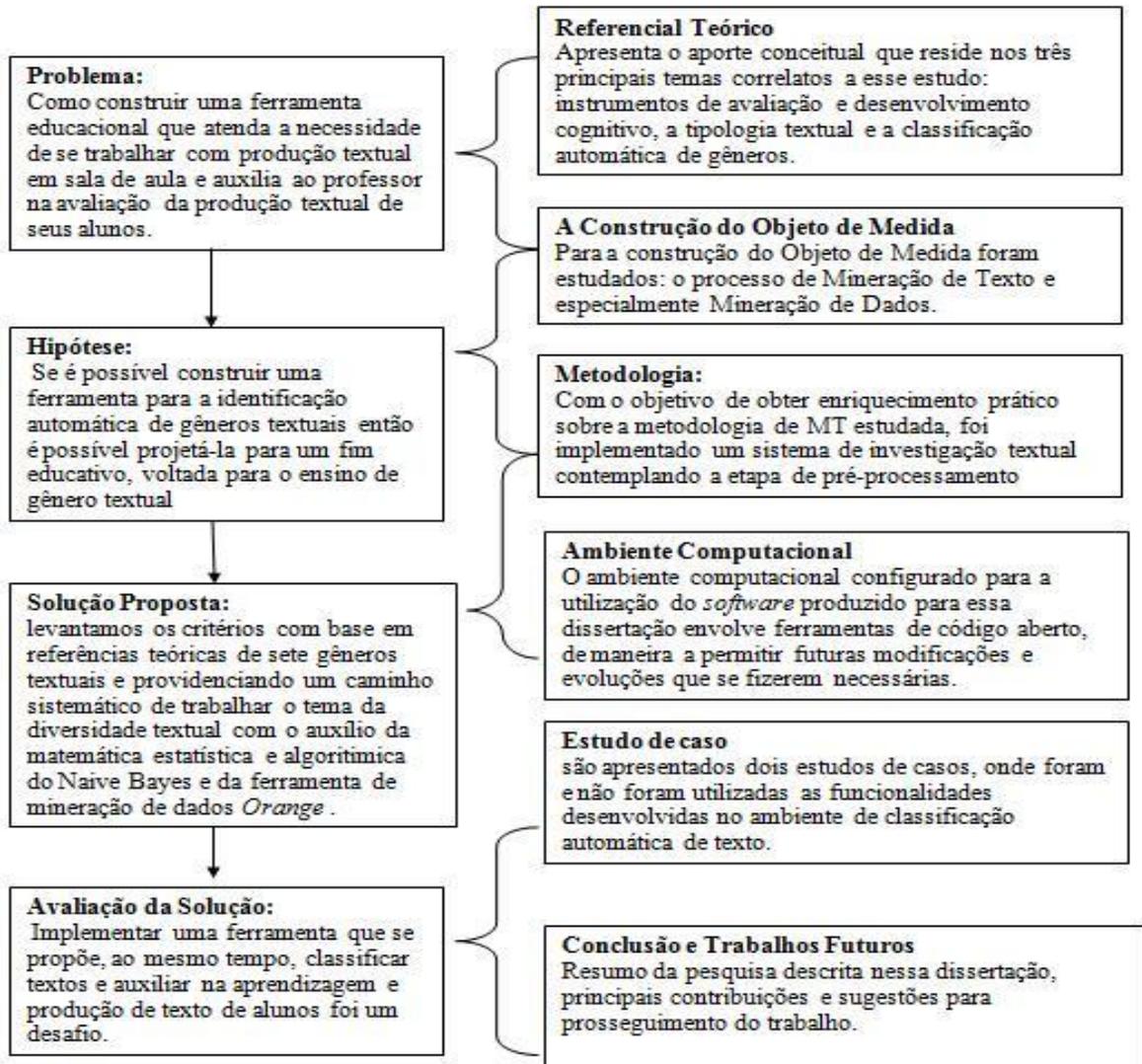


Figura 1. 2: Organograma da dissertação.
Fonte: Autora.

2 REFERENCIAL TEÓRICO

“Uma abordagem com base no letramento representa um estilo de ensino que os educadores devem considerar se querem preparar os aprendentes para uma participação completa em sociedades que progressivamente exigem competência em nível multilinguístico, multicultural e multitextual” (Kern, 2000, p. 16).

Nesse capítulo é feita uma revisão bibliográfica onde é definido o que é letramento, suas principais características e o domínio em que se baseia essa pesquisa.

O aporte conceitual desta pesquisa reside nos três principais temas correlatos a esse estudo: o desenvolvimento cognitivo, a tipologia textual e a classificação automática de gêneros. Na primeira, destacamos a teoria Metaprocessual por Franco Lo Presti Seminário (1997), desenvolvida com base nos teóricos apresentados na Figura 2.3.

BANDURA	Modelação (Transmissão de Modelos ou regras) como meio de promover a aprendizagem, sendo esta resultante da estocagem de modelos que não precisam ser reforçados na hora de sua aquisição. O indivíduo aprende a partir de seu conhecimento adquirido.
BRUNER	Poder ou valor gerativo no sentido de gerar novas hipóteses e combinações. Também denominado de regras gerativas.
CHOMSKY	A regra inata da recursão é a base do desenvolvimento da lógica e da recursão.
FLAVEL	A estratégia metacognitiva e modelo global de controle cognitivo.
GESTALT	Dinamismo inerente à reestruturação cognitiva. Salto do estado de cognição para metacognição através de <i>insight</i>
VYGOTSKY	Participação do educador e o meio social na aprendizagem

Figura 2. 2: Influência dos teóricos na teoria de Seminário.

Fonte: (Barros, 2010)

Existe também contida na pesquisa a produção de uma listagem de critérios que dizem respeito à estrutura de textos. O segundo aporte conceitual utilizado para construção dessa lista de critérios foi retirado do trabalho de pesquisa de Travaglia (2007) com contribuições de outros autores no mesmo. Segundo Travaglia (2007), existem cinco critérios para caracterizar um tipo de texto, que são: a) conteúdo temático, b) estrutura composicional, c) objetivos e funções sociocomunicativas, d) superfície linguística e e) condição de produção.

Uns dos objetivos da Linguística Textual é identificação, distinção e caracterização das diferentes categorias de texto, como pode ser verificado a seguir no item 2.2. Todavia ao entrar em contato com os textos circulantes em uma sociedade e cultura, foi visto que esta não é uma tarefa fácil. Essa identificação ou até mesmo a distinção das categorias de textos dependem diretamente de sua caracterização, porque somente o nome atribuído pelos usuários dos textos nem sempre é

suficiente para identificar e diferenciar as categorias de texto, embora seja o primeiro passo para fazê-lo.

2.1 Letramento

Nessa seção, propomo-nos a aferir os conceitos de letramento e analfabetismo funcional, destacando a importância desse assunto para a inserção do indivíduo na sociedade que o circunda (Soares, 1998). Posteriormente, apresentaremos alguns dados estatísticos acerca de desempenho escolar dos alunos em língua portuguesa dentro do contexto da escola brasileira, reconhecendo que “a grande diferença entre alfabetização e letramento, entre alfabetizado e letrado, é o fato de que um indivíduo alfabetizado não é necessariamente um indivíduo letrado; alfabetizado é aquele indivíduo que sabe ler e escrever; já o indivíduo letrado, o indivíduo que vive em estado de letramento, é não só aquele que sabe ler e escrever, mas aquele que usa socialmente a leitura e a escrita, pratica a leitura e a escrita, responde adequadamente às demandas sociais de leitura e de escrita” (Tfouni, 1988). Por fim, discutiremos algumas abordagens à escrita relevantes para essa pesquisa, bem como aspectos sobre sua realização no contexto escolar.

Atualmente, a discussão acerca do analfabetismo e do chamado analfabetismo funcional tem sido foco de interesse nos meios acadêmicos e de comunicação e político. Em 2009, a secretaria municipal de educação do estado do Rio de Janeiro decidiu que os alunos da rede municipal de ensino deveriam ser submetidos a uma avaliação para diagnosticar o grau de conhecimento deles em Língua Portuguesa e Matemática. O resultado da avaliação diagnóstica realizada pela prefeitura pode ser considerado ruim, já que os números divulgados apontam que, entre o 4º e o 6º ano do ensino fundamental, 10.500 alunos se encontram defasados em relação à idade e à série que estão cursando. O mais alarmante, no entanto, são os dados de 28.870 alunos dentre os cerca de 211.000 da rede terem sido considerados analfabetos funcionais, ou seja, incapazes de fazer uso efetivo da leitura e da escrita nas diferentes esferas da vida social (Ribeiro, Vóvio & Moura, 2002).

Ribeiro, Vóvio & Moura colocam que, apesar de a definição sobre alfabetismo vir sofrendo revisões significativas ao longo dos anos, refletindo assim transformações socioeconômicas, em 1958 a UNESCO definia como alfabetizada a pessoa capaz de ler ou escrever um enunciado simples, relacionado à sua vida diária. Cerca de 20 anos mais tarde, propôs ainda a definição de alfabetismo funcional, que é aquele indivíduo capaz de utilizar a escrita e a leitura para responder as demandas de seu contexto social e usar essas habilidades para continuar aprendendo e se desenvolvendo ao longo da vida (Soares, 1995). Todavia, a precariedade do processo escolar a que conseguem ter acesso as camadas mais populares da sociedade brasileira fez despertar uma preocupação em seus responsáveis. Ao mesmo tempo em que em países desenvolvidos onde se conseguiu universalizar uma escolarização básica mais estendida, também surgem questionamentos sobre a escola ser capaz ou não de fazer frente às demandas do novo contexto socioeconômico e cultural, com a modernização da sociedade e desenvolvimento tecnológico e ampliação da participação social e política.

Na década de 1990, o IBGE – Instituto Brasileiro de Geografia e Estatística –, por recomendação da UNESCO, passou a divulgar os índices de analfabetismo funcional ao lado dos índices de alfabetismo. Entretanto, o critério adotado deixou de ser aquele baseado na auto-avaliação dos respondentes e passou a se basear no número de séries escolares concluídas (IBGE 2001), sendo considerados como analfabetos funcionais aqueles com menos de quatro anos de estudo. Contudo, se quatro anos de estudo nas escolas nacionais da rede pública são suficientes para garantir o alfabetismo funcional, não nos cabe aqui dizer, pois são escassos os estudos acerca dessa temática. Todavia, é relevante para nossa pesquisa, que as evidências sugerem que se considere não só a quantidade de instrução necessária, mas a qualidade, além das oportunidades de uso da leitura e da escrita presentes nos contextos extra-escolares (Wagner, 1999).

Esta breve exposição sobre analfabetismo funcional é de extrema relevância para a pesquisa a qual estamos realizando na medida em que é fundamental para que possamos entender o que é letramento, conceito que comboia e permeia nossa pesquisa.

É interessante que se frise que agora é a era do despertar para o fenômeno do letramento, que está sendo incorporado ao nosso vocabulário educacional - significa que já se compreende que o problema não é apenas ensinar a ler e a escrever, mas é, também, e, sobretudo, levar os indivíduos - crianças e adultos - a fazer uso da leitura e da escrita, envolver-se em práticas sociais de leitura e de escrita. É ser amigo dos livros; dissecá-los, fazer uma bela leitura e repassar para quem não possui os ensinamentos que o letramento nos oferece (Castro, 2009).

Os estudos sobre a língua escrita a partir da década de 80 trouxeram aos educadores o entendimento de que a alfabetização envolve um complexo de elaboração e hipótese sobre a representação linguística, ou seja, é muito mais que a apropriação de um código (Colello, 2006). Nos anos seguintes, o estudo sobre a língua escrita se manteve ampliado para sua dimensão sociocultural e aprendizado, e ao mesmo tempo, rompendo com a separação daquele que aprende – o aluno – e daquele que ensina – o professor. Assim, a sala de aula deixou de ser vista como o único espaço de aprendizagem.

Na definição de Soares, letramento é “estado ou condição de quem não só sabe ler e escrever, MAS exerce as práticas sociais da leitura e escrita que circulam na sociedade em que vivem” (Soares, 1999 p.3 – grifos da autora). Ainda nesse contexto está implícita “a idéia de que a escrita traz conseqüências sociais, culturais, políticas, econômicas, cognitivas, lingüísticas, quer para o grupo social que seja introduzido, quer para o individuo que aprenda a usá-la” (Soares 1998. p.7).

Foi justamente esse aspecto da escrita que nos chamou atenção, ressaltamos que não somente propor, mas também fornecer uma ferramenta para ensino de língua, que vise a desenvolver no aluno um conjunto de habilidades e comportamentos de escrita, que lhe permitam fazer um uso e reuso maior e mais eficiente da mesma (Bagno, Sttubs & Gagné, 2002), levando-o a atingir um grau cada vez mais alto de letramento, sendo esta, uma das maiores motivações para todo esse trabalho.

Soares (1998) afirma que a palavra letramento encontra-se no mesmo campo semântico de palavras como alfabetização, alfabetizar, alfabetizado, analfabeto, letrado, iletrado e analfabetismo. Conhecemos a palavra **letrada** versado em letras, erudita – e iletrado aquele que não tem conhecimentos literários. Nesse contexto, uma pessoa letrada seria uma pessoa erudita,

experimentada em literatura e línguas, enquanto iletradas seriam aquelas que não têm conhecimentos literários, que não são eruditas; analfabetas ou quase isso. Mas esses sentidos não estão relacionados à palavra letramento. Soares (1998) propõe uma reflexão acerca do motivo de surgimento dessa terminologia. E aponta para a necessidade de se criar uma nova para dar conta de outro fenômeno que, se existia antes, não nos dávamos conta dele, já que não levávamos em conta o aluno no social (Castro, 2009).

Segundo Soares (1998), quando a pessoa muda sua condição de alfabetizado para letrado, tal mudança traz conseqüências para sua vida, nos âmbitos social, cultural, cognitivo e linguístico. Do ponto de vista social, passar à condição de letrado, ainda que não acarrete uma mudança de nível social, faz com que o sujeito passe a se relacionar com a sociedade de outra maneira: na medida em que se insere na cultura, sua relação com os outros muda. No aspecto cognitivo, acredita-se que a pessoa letrada passe a pensar diferente da iletrada. No tocante ao campo linguístico, pesquisas indicam que o indivíduo letrado fala de modo diferente do analfabeto e/ou do iletrado, mostrando a mudança no uso da língua oral, das estruturas linguísticas, e do vocabulário como conseqüência do convívio com a língua escrita.

Acreditamos ter chegado, então, a uma definição satisfatória do que é letramento, do resultado da ação de se tornar letrado. Segundo Soares (1998) letramento é a condição que um indivíduo ou grupo social adquire como consequência de ter se apropriado da escrita, bem como de suas práticas sociais, o que é diferente de saber ler e escrever.

O Instituto Nacional de Estudos e Pesquisa Educacionais, Inep, é responsável por administrar exames de caráter nacional e internacional, que têm como objetivo conhecer mais profundamente o sistema educacional brasileiro; alguns desses exames são o PISA e SAEB, fora os estudos e pesquisas que também são realizados. É fato que os resultados desses exames e estudos apontam para uma realidade um tanto desesperadora para a educação nacional. Em fevereiro 2007, o Inep publicou um estudo intitulado: *SAEB-2005- Primeiro resultados: Médias de desempenho do SAEB/2005 em perspectiva comparada*, no qual buscou avaliar a capacidade de leitura dos alunos nos últimos dez anos. Segundo os dados desse documento, em Língua Portuguesa, os alunos da 8ªsérie/9ºano tiveram uma média de 232 pontos na avaliação, quando o mínimo aceitável seria de 300 pontos. No ensino médio, a média foi de 271 pontos, sendo que de 2003 para 2005 houve uma queda de 266.7 pontos para 257.6. (Castro, 2009)

Em 2008, um estudo da UNESCO: *uma visão dentro das escolas primárias*, realizado em 11 países em desenvolvimento, mostra que o Brasil está entre os 5 (cinco) países que mais usam “métodos mecânicos” de ensino, devido ao uso excessivo da lousa, em detrimento das aulas mais focadas na capacidade de compreensão. Outros dados estatísticos levantados por essa pesquisa indicam que no Brasil 40% dos alunos em séries iniciais, na maioria das vezes, têm aulas que copiam o conteúdo; 20% aprendem recitando tabelas e fórmulas; 10% aprendem repetindo frases; 25% de alunos nas zonas urbanas estudam em prédios em más condições (em zonas rurais esse índice pode chegar a 50%); 10% frequentam escolas que carecem de água potável; 38% estão matriculados em escolas sem biblioteca; e 87%, em escolas sem computadores conectados a internet (Castro, 2009)

Todas essas adversidades certamente afetam o aprendizado dos estudantes. E cabe aqui uma reflexão acerca da escola de hoje.

De fato, a escola hoje é apontada como uma instituição defasada em relação à sociedade moderna por diversos especialistas de educação e áreas afins. A escola é caracterizada como uma instituição que se mantém fossilizada, em meio a um mundo que não para de mudar, visto que “muda a sociedade e somente mais tarde muda a educação” (Libâneo, 2003^a, p.43).

Sendo assim, aqui reportamos um micro panorama do sistema educacional do Brasil através dos dados estatísticos acerca das condições físicas das escolas e do desempenho de alunos brasileiros em avaliações aplicadas nesses ambientes. Ressalta-se que essas avaliações, assim como muitas outras que se propõem em algum tipo de diagnose em língua portuguesa, privilegiam apenas um aspecto: a leitura. A escrita é deixada de fora e não tem sido avaliada nesses contextos. Tal fato é bastante interessante para o conteúdo de nossa pesquisa, pois gera reflexões e questionamentos acerca das razões pelas quais a escrita é excluída desses tipos de avaliação. Por um lado, consideramos as questões políticas que podem haver. Ora, se os resultados referentes à avaliação de leitura, obtidos através das provas objetivas, não são muito otimistas, atividades que envolvessem uma efetiva produção da parte dos alunos não seriam menos pessimistas. Por outro lado, pode-se argumentar ainda que, considerando a dificuldade de correção e avaliação de textos, tarefa que, além de muito mais trabalhosa requer mais tempo, isso nem sempre é viabilizado.

No entanto, se a dificuldade de avaliação de textos escritos está presente na não opção de avaliação da produção escrita de alunos em exames desse porte, então sua existência na escola pode estar fadada a nem existir, visto que numa sala de aula a produção textual significará mais de 50 textos diários para serem avaliados por um único professor, responsável pela turma, e isso pode vir a ser um obstáculo para o trabalho com produção escrita em sala de aula. É, portanto, bastante relevante propor uma ferramenta que auxilie e ajude na realização dessa tarefa, como o jogo aqui em questão.

2.2 Gêneros Textuais e Ensino

Nessa seção, propomo-nos a aferir os conceitos de gêneros textuais e o ensino, destacando que hoje o ensino de línguas anda bastante centrado em gêneros. Portanto, uma dificuldade natural no tratamento desse tema acha-se na abundância e diversidade das fontes e perspectivas de análise. Posteriormente, apresentaremos a perspectiva por nós aqui escolhida para o estudo dos gêneros e sua justificativa. Por fim, discutiremos algumas abordagens relevantes para essa pesquisa, bem como aspectos sobre sua realização no contexto escolar.

A palavra *gênero* advém, na tradição ocidental, aos gêneros literários cuja análise se inicia com Platão, se firma com Aristóteles, mais precisamente no Capítulo três da Retórica [1358a], no qual surge uma teoria mais sistemática sobre os gêneros e a natureza do discurso composto por três elementos do discurso, três tipos de ouvinte e três gêneros de discurso retórico. Segue passando por Horácio na Idade Média, seguindo por outros autores no Renascimento rumo à Modernidade até os

primórdios do século XX. Todavia, hoje gênero não está vinculado apenas à literatura. A terminologia *gênero* refere-se a uma categoria distintiva de discurso de qualquer tipo, falado ou escrito, com ou sem aspirações literárias (Swales, 1990 p.33).

A teorização sistemática de Aristóteles sobre as estratégias e estruturas da epopéia, da tragédia e da comédia foi um marco precursor do estudo sobre gêneros, mas hoje predomina uma perspectiva diferente da aristotélica. Cada vez mais multidisciplinar, a análise do gênero engloba uma análise do texto e do discurso, uma descrição da língua numa perspectiva sociocognitiva e interacional e ainda tenta responder questões de natureza sociocultural no uso da língua de maneira geral (Marcuschi, 2008). Nesse sentido, há muito a discutir sobre a definição de gênero textual – conceito essencial no ensino de língua portuguesa atualmente.

Gênero Textual é definido por Marcuschi (2008) como uma abstração teórica dos textos encontrados no dia a dia, que apresentam características sociocomunicativas definidas pelos conteúdos, propriedades funcionais, estilo e composição características. O conceito de gênero foi cunhado pela primeira vez por Bahktin (1959) e vem sendo retomado com frequência por pedagogos e linguistas. Para Travaglia (1998), o Gênero Textual se caracteriza por exercer uma função social específica, pressentida e vivenciada pelos usuários. Isso equivale dizer que, intuitivamente, sabemos que gênero usar em momentos específicos de interação, de acordo com a sua função social.

Há, porém, mais um conceito importante quando se considera o gênero textual: o de sequência (ou tipo) textual. Swales (1990, p.58) afirma que o conceito de gênero compreende “uma classe de eventos comunicativos, cujos participantes compartilham algum conjunto de propósitos” – corroborando a linha que costuma nortear outros pesquisadores. Já Biber (1998) compara os dois conceitos, explicando que o gênero se fundamenta no uso (aspectos externos) dos textos, enquanto o tipo seria definido com base na forma dos textos (aspectos internos).

Marcuschi (2008), levando em conta as diversas propostas já elaboradas pelo tema, sugere que um tipo textual é algo mais diverso do que uma forma textual empiricamente realizada. Uma forma textual se dá como um evento de fala e é uma ocorrência empírica de um tipo possível, mas raramente se realiza como um tipo puro numa dada circunstância, já que as relações textuais são em geral heterogêneas quanto às formas manifestadas.

Nesse trabalho, partilhamos dos conceitos apresentados por Marcuschi (2008), gênero textual e tipo textual, da seguinte maneira: o primeiro refere-se a textos estruturados internamente por um ou mais tipos textuais; o segundo é a própria estruturação interna, raramente encontrada em “estado puro” nos textos de circulação social. A maioria dos autores (cf. Marcuschi 2008) admite, tendo em vista a natureza do texto, quatro tipologias textuais: narração, argumentação, exposição e descrição. Por sua vez, o gênero textual é observado a partir de seu uso social, de sua realização empírica, estando estreitamente relacionado aos sujeitos como produtores e receptores de texto; são inúmeros os exemplos: carta, bilhete, lista de compras, receita, folheto, panfleto, notícia de jornal etc. A estruturação interna dos gêneros é essencialmente heterogênea, por isso ele se compõe, geralmente, de mais de uma tipologia.

Há um número considerável de trabalhos voltados para a discussão de tipologias textuais, além de diversas pesquisas envolvendo os gêneros textuais, como se observa em Adam (1993) e Marcuschi (2002), dentre outros.

Para esses autores, é impossível se comunicar verbalmente, a não ser por algum texto representativo de um gênero textual. Essa posição, defendida por Bakhtin [1997] e também por Bronckart (1999), é adotada pela maioria dos autores que tratam a língua em seus aspectos discursivos e enunciativos, e não apenas em suas peculiaridades formais. Esta visão segue uma noção de língua como atividade social, histórica e cognitiva, privilegiando a natureza funcional e interativa, e não o aspecto formal e estrutural da língua. Afirma-se o caráter de indeterminação e ao mesmo tempo de atividade constitutiva da língua, o que equivale a dizer que a língua não é vista como um espelho da realidade, nem como um instrumento de representação dos fatos (Travaglia, 1991).

Apesar de as discussões teóricas ainda estarem em andamento e ser esse tema ainda controverso, percebem-se seus reflexos nos livros didáticos de português (LDP) a partir dos anos 90, pois houve inclusão de textos variados para a leitura (veiculados por livros, jornais, revistas, televisão etc.) e solicitação da escrita de textos mais diversos (carta, notícia, reportagem, resumo, diálogo, relatório etc.). Entretanto, mesmo que essa atitude demonstre uma tentativa dos autores de LPD de se manterem atualizados com as pesquisas de linguística, não necessariamente aparecem pressupostos teórico-metodológicos adequados à proposta textual, como comprovou pesquisa de Bezerra (2006).

A proposta dos Parâmetros Curriculares Nacionais (PCN) de fundamentar o ensino de língua materna, tanto oral quanto escrita, nos gêneros textuais desencadeou diversas pesquisas que visam a descrever uma quantidade considerável de gêneros com a finalidade de apresentar sugestões didáticas para o uso de textos como exemplares e fonte de referência de determinado gênero.

Em relação à Língua Portuguesa, os PCN propõem que seja trabalhada em função do eixo USO-REFLEXÃO-USO, tendo como um dos seus objetivos “expandir o uso da linguagem em instâncias privadas e utilizá-la com eficácia em estâncias públicas sabendo assumir a palavra e produzir textos, tanto orais como escritos, coerentes, coesos, adequados a seus destinatários, objetivos a que se propõem e assuntos tratados” (PCN de Português, 1997, p.41). Com isso, passou-se a considerar, como unidade de ensino, o texto em suas variadas formas, denominadas pelos PCN de gênero.

Os PCN enfatizam a responsabilidade de todas as disciplinas ensinarem o aluno a ler e produzir todos os textos de que fazem uso, embora caiba à língua portuguesa uma abordagem mais sistemática (cf. PCN de Português, 1997, p.31). No entanto, ao referirem-se ao trabalho com gêneros, inclusive apresentando uma notável distribuição por ciclos (cf. PCN de Português 1997 p.111-112, p.128-129), os Parâmetros não são auto-explicativos:

Um escritor competente é alguém que, ao produzir um discurso, conhecendo possibilidades que estão postas culturalmente, sabe selecionar o gênero no qual seu discurso se realizará escolhendo aquele que for apropriado os seus objetivos e circunstancia enunciativa em questão. Por

exemplo, se o que deseja é convencer o leitor, o escritor competente selecionará um gênero que lhe possibilite a produção de um texto predominantemente argumentativo; se é fazer uma solicitação a determinada autoridade, provavelmente redigirá um ofício (...) (1997, p.65).

Ou seja, no documento consta somente a instrução de que o aluno escolha o gênero textual mais adequado sem nenhuma informação mais precisa sobre o que é gênero. Os Parâmetros Curriculares não são um manual, e sim diretrizes, mas não auxiliam os professores em relação à compreensão de conceitos à luz de teorias linguísticas ou quanto ao tratamento didático a ser dado em sala de aula ao desenvolverem atividades com textos.

Com base, portanto, nas pesquisas que defendem a aplicação da teoria sobre gêneros e tipologias textuais no ensino, o Jogo do Jornal foi elaborado como tentativa de possibilitar a execução de uma tarefa no ensino de uma maneira geral, e em sala de aula de modo particular, tratando os textos na perspectiva dos gêneros, para levar os alunos a produzir ou analisar eventos linguísticos os mais diversos, tanto escritos como orais, e identificar as características de cada gênero.

Acreditamos que seria produtivo possibilitar ao aluno a criação de um jornal diário que, implicitamente, o faça refletir acerca dos gêneros textuais presentes, as suas características centrais em termos de conteúdo, composição, estilo, nível linguístico e propósitos. É evidente que essa tarefa pode ser reformulada de muitas maneiras, de acordo com os interesses de cada situação de ensino. Acreditamos, porém, que, por mais modesta que seja a análise, ela será sempre muito promissora. E mais promissor ainda será possibilitar ao aluno a prática de produção desses gêneros através do jogo do jornal.

O trabalho com gêneros textuais é uma extraordinária oportunidade para lidar com a língua em seus mais diversos usos diários. Afinal, como lembra Marcuschi (1996), “Dentro de uma perspectiva linguística tudo o que fizermos será feito em algum gênero. Portanto, tudo o que fizermos linguisticamente pode ser tratado em um ou outro gênero”. Por isso, o Jogo do Jornal escolheu sete gêneros, como podem ser verificados a seguir, produzidos de maneira sistemática e com grande incidência na vida diária que aparecem nas diversas mídias hoje existentes, inclusive a mídia virtual, para serem trabalhados.

No entanto, acreditamos que os alunos não aprendam naturalmente a produzir os diversos gêneros escritos de uso diário. Nem é comum que se aprendam naturalmente os gêneros orais mais formais.

Por outro lado, é de se indagar se há gêneros textuais ideais para o ensino de língua. Tudo indica que a resposta seja não. Mas é provável que se possam identificar gêneros com dificuldades progressivas, do nível menos formal ao mais formal, do mais privado ao mais público e assim por diante. Nós resolvemos escolher os gêneros de maneira aleatória, apenas respeitando como modelo o universo de gêneros encontrados no jornal impresso hoje no país (Dolz & Scneuwly, 1998).

A escolha e denominação dos gêneros escolhidos para fazerem parte do Jogo do Jornal nos levaram a diversas questões, como a “intergenericidade” (Marcuschi, 2008). Inicialmente

achamos que seríamos capazes de abordar uma variedade de gêneros, mas logo a dificuldade de classificação desses nos obrigou a diminuir mais da metade o nosso escopo de pesquisa. Diminuído o escopo, a dificuldade agora era de nomear os gêneros escolhidos. Com certeza, a denominação que usamos não foi pessoal, mas histórica e socialmente construída, presente em diversas outras fontes e denominados da mesma maneira. Contudo, é difícil atribuir nome para gêneros de texto, como já notaram inúmeros autores, desde Bakthin (1979), que afirma que os gêneros muitas vezes se imbricam e interpenetram para constituírem novos gêneros, até mesmo fazendo a função um do outro. Porém neste trabalho foram escolhidos gêneros básicos, já classificados e definidos.

Enfim, o trabalho com gêneros é uma forma de dar conta do ensino de leitura e escrita, seguindo a proposta oficial dos Parâmetros Curriculares Nacionais que defendem esta perspectiva. Tem-se a oportunidade de observar tanto a oralidade como a escrita em seus usos culturais mais autênticos sem forçar a criação de gêneros que circulam apenas no universo escolar. Por isso, os gêneros selecionados para fazerem parte do Jogo do Jornal são: notícia, artigo, resenha crítica, crônica, horóscopo, receita e manchete.

2.3 Concepções de Aprendizagem

Esse estudo busca seguir a teoria da diversidade textual segundo a maneira como foi utilizada por Josette Jolibert e Mabel Condemarín e Felipe Allende em suas obras respectivas "Formando crianças produtoras de texto" e "Leitura, teoria, avaliação e desenvolvimento". Essas duas obras possuem um vasto repertório de estudos e pesquisas realizados por esses três autores e colaboradores que desenvolveram uma maneira de pensar a leitura como um instrumento prático e hábil para o ensino da escrita, dentro da sala de aula, utilizando os professores e o material didático para sua realização.

A criança precisa experimentar não só a utilidade e funções da escrita como também o poder de se ter um domínio suficiente da escrita e o prazer que isso possa lhe proporcionar. A criança precisa entender que a escrita corresponde às intenções e difere de acordo com as situações, permitindo-lhe comunicar-se, contar histórias, criar etc. Escrever não deve ser sinônimo de trabalho enfadonho, bloqueio e fracasso; a criança deve sentir prazer de compreender como um texto funciona, de buscar as palavras, de encontrar o tipo de escrita adequado para a situação e ver a finalização de seu trabalho num texto escrito. (JOLIBERT, 1994).

A partir da revisão da literatura, observou-se que seria necessária a criação de uma lista de critérios sobre a diversidade textual, que auxiliaria na classificação automática de textos. Sabe-se que ao construir seu texto, quem o escreve deve ser capaz de contemplar seus principais níveis linguísticos e para isso, a melhor maneira de aprendê-lo é através de uma imitação ou observação de regras que conseqüentemente se tornarão intrínsecas nas crianças à maneira que vão tendo contato com elas, esses principais níveis linguísticos a serem adquiridos são:

- a) Superestrutura: definição dos blocos de texto e dinâmica interna;

- b) Linguística do texto: funções de linguagem dominantes, enunciação de pessoas e espaço/tempo, coerência do texto;
- c) Linguística da frase: ordem das palavras, relações sintáticas e gestão de frases complexas (JOLIBERT, 1994)

Tendo em vista esse arcabouço linguístico que precisa ser adquirido e ensinado para a criança, é importante utilizar uma metodologia que seja capaz de transformar em realidade toda essa tarefa, assim chegamos a teoria psicopedagógica da elaboração dirigida do teórico Franco Lo Presti Seminário.

Segundo Seminário (1997), as teorias de aprendizagem vigentes não são suficientemente eficazes na tarefa de ensinar. Devido a diversas razões, não colaboram com a possibilidade de proporcionar saltos cognitivos nas crianças que só uma abordagem metacognitiva é capaz de fazer.

A elaboração dirigida atua por encadeamentos sucessivos na abstração e no imaginário da criança, a partir de uma metacoginição de cada uma das aquisições operadas, e a avaliação prévia imaginativa das conseqüências, a partir de diálogo e motivação, com objetivo de fazer a criança estocar modelos – instrumentos mentais representativos, para resolver situações novas de compreensão e domínio da realidade.

Ao observar as ações dos outros, o indivíduo sem dúvida considera para si características do modelo, as suas próprias características e se compara com ele, atentando ainda para as conseqüências positivas ou negativas daquelas ações, e poderá vir a adotar aquelas condutas como referências para si. No caso da televisão, o telespectador é um observador exposto a padrões diversos de conduta que poderá vir adotar por meio do mesmo mecanismo de identificação na observação de modelos. No caso aqui, esse modelo é o *software* que apresentará um único formato de texto possível a ser desenvolvido pela criança.

Foi investigando o comportamento imitativo e aprendizagem incidental que Albert Bandura chegou às teorizações sobre processo de aquisição de comportamentos a partir de modelos programados ou incidentais. A aprendizagem decorrente da observação foi rotulada como imitação e outros termos, tais como aprendizagem observacional, cópia, facilitação social, contágio, identificação, aprendizagem sem ensaio e jogo de papéis. No processo de aprendizagem sem ensaio, Bandura (1965a) postula que o desenvolvimento de respostas mediadoras, sob a forma de representações simbólicas e verbais implícita dos estímulos percebidos, pode desempenhar papel relevante: "novas respostas são adquiridas ou respostas já existentes são modificadas em decorrência da observação do comportamento de outros e de suas conseqüências reforçadoras, sem que haja desempenho aberto do observador durante o período da observação do modelo." Bandura (1965ª p.15).

A aprendizagem por meio de modelo, segundo Bandura, é dependente de quatro subprocessos inter-relacionados:

- d) Atenção;
- e) Retenção ou lembrança do comportamento;
- f) Reprodução motora;
- g) Reforço e motivação.

Bandura (1971^a) considera que a principal função do estímulo modelo é transmitir informação aos observadores sobre como organizar os componentes das respostas dentro de novos padrões de comportamento.

O modelo pode ser apresentado por vários veículos: demonstrações físicas, representações pictóricas, descrições verbais e por um *software* não mencionado por ele devido à ignorância tecnológica da época.

Aqui o *software* se utilizará da formatação das características comuns a um determinado tipo textual e guiará a criança através de suas tentativas a internalizar as regras existentes para a construção de determinado tipo e gênero de texto.

A teoria de modelação cognitiva descrita poderá ser aplicada futuramente com nossa ferramenta como suporte ao jogo, que será utilizado pelas crianças com o intuito de ensiná-las a aprender. A ferramenta disponibilizará as regras de diversos tipos de texto, porém isso nunca será disponibilizado para as crianças de maneira descritiva. Por outro lado, crianças trabalharão com os gêneros que por sua vez estão contidos nos tipos, e ao reproduzir os textos estarão aprendendo as regras inconscientemente, ou seja, por detrás do texto.

2.4 Técnica metacognitiva da Elaboração Dirigida

Seminério foi um grande estudioso da cognição humana, “abarcando tanto dimensões teóricas e técnicas quanto políticas. (...) Possui a intenção declarada de criar, em última instância, ferramentas efetivas de intervenção sobre o desenvolvimento intelectual das crianças oriundas de classes socialmente desfavorecidas” (Dittrich *et al*, 2001). Reconhecido como um dos pioneiros da psicologia no Brasil questionou a cognição tal como se apresentava na década de 60 – voltada para diagnóstico através de instrumentos que, para ele, não ofereciam confiabilidade suficiente para o que almejava em termos práticos (*idem*). Seminério foi buscar bases teóricas existentes sobre o processo cognitivo, tal era a sua necessidade de “estudar, criticar e reestruturar” (*idem*).

Oferecer uma proposta de mudança para o ensino brasileiro foi a meta perseguida na vida de Seminério em mais de 40 anos de estudo que ficam evidentes em grande parte de suas publicações. Durante suas pesquisas sobre a psicologia humana, notadamente no campo da cognição, Franco Lo Presti Seminério percebeu que o desenvolvimento da inteligência não ocorre sempre na mesma profundidade, variando de acordo com o meio onde o indivíduo cresça.

A criança que se desenvolve em um nível social menos favorecido, podendo ter um acesso muito restrito ao conhecimento necessário para o desenvolvimento de sua cognição, apresentará consequente *déficit* e se tornará um adulto intelectual e cognitivamente defasado. Este é um problema não só político, mas principalmente ético - pois ele perpetua as diferenças entre pessoas de classes sociais diferentes. Como esclarece Chiarotino (1987) após pesquisas realizadas em São Paulo: “Temos um 'déficit' e não uma 'diferença'. Isso não significa que sejam (crianças) inferiores: superando esse déficit elas podem ir até mais longe do que as outras”.

Seminário desenvolve um trabalho fundamentado na teoria da metacognição, capaz de intervir efetivamente sobre o desenvolvimento intelectual das crianças, principalmente das oriundas de classes socialmente desfavorecidas, a fim de diminuir as disparidades sociais. Acredita que este *déficit* possa ser, dentre muitos outros fatores, instrumento para as desigualdades entre as classes sociais. O que reforça a tese ambientalista, que defende a ideia de que a aprendizagem, desde que uniformizada, garante oportunidades. Sugerindo assim, que não é a inteligência que produz as classes sociais, e sim, o contrário, ou seja, as classes sociais produzem a inteligência.

O estudo de Seminário indaga como é possível remover o efeito social efetuando modificações substanciais preventivas em crianças na primeira infância. As técnicas pedagógicas vigentes, segundo Seminário, mostram não ser capazes de resolver este problema. A Escola Tradicional centraliza a figura do professor, que detém o saber, com alunos identificados tão somente como receptores passivos do conhecimento acadêmico. Trata-se de um modelo conservador e hierarquizado. A Escola Nova defende a autonomia do aluno, que possui atuação direta no processo de aprendizagem. A Escola Operatória, apesar de fundamentar no princípio de construção de estruturas cognitivas como base de conhecimento, postula que essa construção é dada por etapas sequenciais, de acordo com os estágios de desenvolvimento da criança, impossíveis de aceleração.

Uma quarta técnica, posterior à Escola Operatória, é criada por Seminário, em que propõe um método que privilegia o uso metaprocessual para aquisição de conhecimentos, efetivada por intervenções específicas do educador que ampliaria a capacidade da criança refletir sobre seus processos cognitivos, permitindo a construção de uma sociedade mais democrática e participativa. Esta técnica recebe o nome de Elaboração Dirigida, que consiste em fornecer um modelo lógico ao aluno por intermédio de uma modelagem cognitiva como ilustra a Figura 2.4.

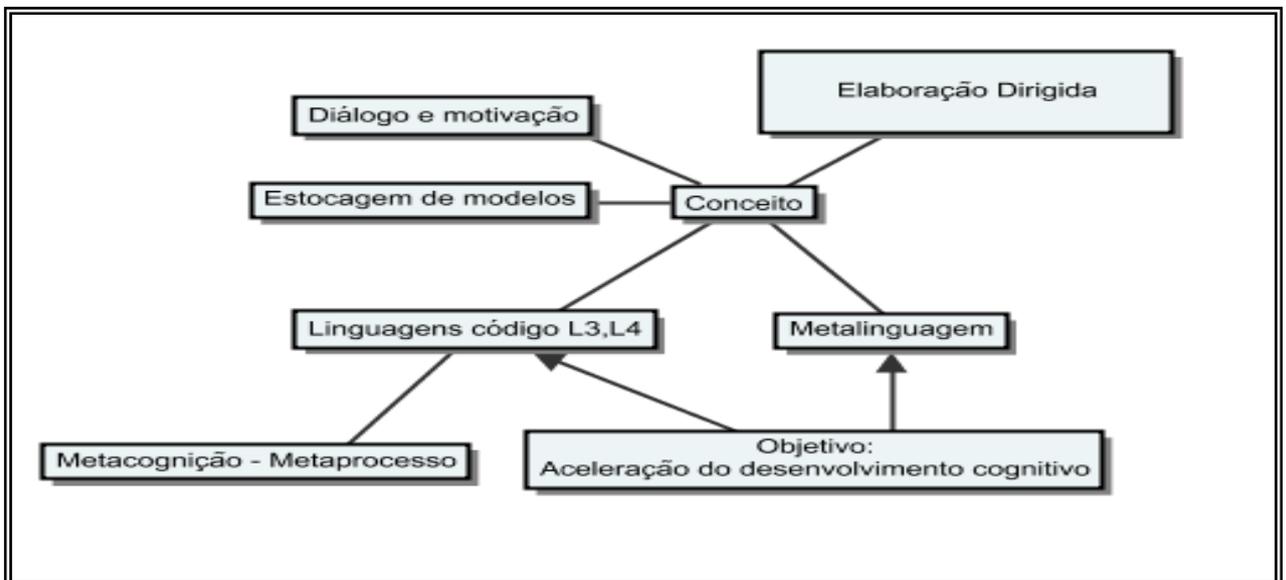


Figura 2. 3: Mapa Conceitual do Conceito
Fonte: (Carla *et al*, 2009b)

Possibilitado pelas inferências do professor, a técnica possibilita ao aluno se apropriar da regra generativa de um determinado conhecimento e transferi-la do inconsciente para a consciência. Esse nível de identificação e consciência das regras preestabelecidas através da Elaboração Dirigida

- percebidas pelo aluno como sendo derivativas de um conjunto de invariantes - só pode ser decodificado através das estruturas de linguagem que fornecem a via de acesso aos diferentes mecanismos cognitivos com os quais cada aluno é dotado.

É nesse aspecto que se insere a proposta da aplicação pedagógica da infraestrutura teórica deste projeto: uma vez hipotetizadas as “linguagens básicas”, item 2.4.2, como competências inatas, torna-se viável o controle dos respectivos desempenhos através do manejo consciente do próprio sujeito, exercido sobre os mecanismos destas competências. Em termos práticos, trabalhando-se uma criança para a compreensão da existência de regras ou invariantes, em quaisquer situações, e a implícita capacidade de descobri-las, espera-se alcançar um salto na escala de suas possibilidades para “aprender a aprender”.

Essa possibilidade metaprocessual inverte até certo ponto um aspecto aplicativo da teoria piagetiana: em vez de esperar que a criança, apenas contando com suas possibilidades, realize assimilações, acomodações e equilibrações, pensa-se ser viável trabalhar com um modelo lógico que possa ser elaborado pela criança através da imitação provocada, encurtando a passagem de regras potenciais do inconsciente cognitivo para a consciência para ampliar as possibilidades de aprendizagem.

Em última análise, na Elaboração Dirigida, busca-se levar o aluno a refletir sobre a regra que gerou um determinado resultado, num processo recursivo - o que significa entender a regra que gerou a regra e assim sucessivamente. Mesmo assim entendido, considera-se impossível conceitualizar a Elaboração Dirigida sem contextualizá-la segundo a teoria metaprocessual, ampliada por Seminério – o que se faz a seguir de forma resumida.

2.4.1 Metacognição

Seminério, em seus estudos, complementa a definição de metacognição, acrescentando o entendimento de que os processos e produtos cognitivos são passíveis de serem monitorados de forma sistemática e integrada, tanto intuitiva quanto instigada. Neste ponto, distingue-se da teoria continuísta piagetiana - da construção do saber ao longo do tempo de forma espontânea, ao considerar que os patamares do desenvolvimento cognitivo podem ser provocados e também acelerados. Seminério encontra, ainda, na teoria de Bandura, uma consistente argumentação relacionada ao conceito da estocagem de modelos – que cita, por exemplo, não ser necessário que um engenheiro construa, por ensaio e erro, tantas pontes, até conseguir uma que não caia; da mesma forma que nenhum motorista aprende a dirigir um carro à custa de reiterados acidentes. Em qualquer dos casos, as regras fundamentais foram veiculadas e incorporadas antes de efetivamente serem utilizadas. Todavia, todo esse processo de ensino fundamentado no fornecimento da regra não é o bastante para explicar o metaprocesso. Para Seminério, não basta informar a regra, é preciso se chegar à regra que gerou essa regra, a regra que gerou a regra que gerou essa regra e assim sucessivamente, num monitoramento consciente do processo mental, firmando-se na certeza de que não é necessário esperar que a criança construa o seu saber autonomamente. Ao contrário, convém

facilitar sua compreensão através de metamodelos, mediante uma ação reflexiva, mostrando-lhe com clareza as regras generativas (conceito de regra que gerou a regra e sucessivamente, mencionado anteriormente), possibilitando o salto metaprocessual.

2.4.2 Linguagens-Código: L3 - L4

O fluxo constante das informações é traduzido para a realidade através de quatro linguagens morfogenéticas, L1,L2,L3 e L4 possibilitando a atividade cognitiva do homem. Nelas fundamentada, a técnica metaprocessual da Elaboração Dirigida ancora-se mais precisamente nas duas últimas (L3 e L4), pois as primeiras (L1 e L2) dizem respeito à estrutura fundamental da formação cognitiva do ser humano, na formação dos canais viso-motor e audio-fonético, receptores de informação e conhecimento.

A Elaboração Dirigida é um processo reflexivo consciente que ocorre através da metalinguagem correspondente ao mecanismo cognitivo utilizado, assim fica fácil perceber a importância das referidas linguagens-código. Para Seminério, qualquer criança que tenha acesso a determinados paradigmas lógicos e imaginativos (L4 e L3) terá condições de alcançar os estágios pré-operatório, operatório concreto e operatório formal, descritos por Piaget.

A L4 tem um aspecto que a diferencia qualitativamente das anteriores: a recursividade - que permite o controle metacognitivo de todas as outras linguagens. Seminério afirma que quando a criança tem acesso a modelos oferecidos por adultos, através de um diálogo intencionalmente proposto, conseguirá suprir lacunas consideradas insuperáveis. Esse processo dialético tem na L4 uma ferramenta fundamental para a deflagração do desenvolvimento potencial do raciocínio lógico – já que é uma linguagem que se relaciona às regras lógicas e sua flexibilidade. Já o uso da L3 implicará a avaliação prévia imaginativa das conseqüências, das estratégias por encadeamentos sucessivos. Isso significa que, ao aplicar a técnica da Elaboração Dirigida, o diálogo ajudará na construção de instrumentos mentais representativos das regras, ou seja, das estratégias que serão traduzidas para regras e que serão reutilizadas para resolver situações novas a fim de dominá-las.

2.4.3 Metalinguagem

A metalinguagem, segundo Seminério, é a via que torna possível o desenvolvimento cognitivo, uma representação da ação para a construção da operação. Ainda em Seminério, constitui-se como instrumento de uso do metaprocesso – já que pela intermediação do diálogo obtém-se como resultado a recursão de regras.

Em estudos linguísticos realizados com o propósito de descobrir o que faz da comunicação humana diferente da comunicação dos outros seres vivos, chegou-se ao entendimento de que a primeira incide sobre a capacidade de recursão com a língua. Apenas para ampliar a informação, ser recursivo linguisticamente significa se utilizar de construções frasais com pronomes relativos, por

exemplo “aquela é a menina **que** veio aqui ontem e quando partiu não disse adeus” o que possibilitaria argumentar infinitas vezes sobre a mesma ideia. A Figura 2.4 ilustra essa interpretação a partir da teoria das linguagens códigos de Seminário (1997) detalhado na seção 2.4.2. As setas marcam os processos, por exemplo, as linguagens código são acionadas para a interpretação da ação, essa pode subir a escada que representa os estágios de desenvolvimento, degrau por degrau, ou pular de um degrau para outro mais em cima, ou até mesmo ir direto lá para cima, assim quando isso ocorre, chamamos de salto metacognitivo que leva a representação da ação a construção da operação.

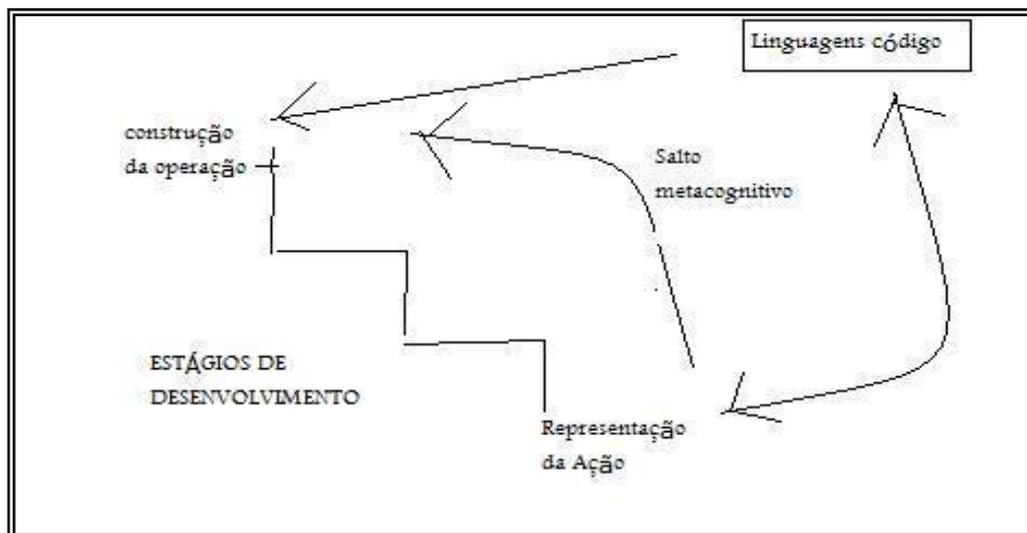


Figura 2. 4: Mapa explicativo sobre Metalinguagem
Fonte: (Marques *et al*, 2009b)

2.4.4 Estocagem de Modelos

Segundo Bandura, toda a aprendizagem resulta da estocagem de modelos, que não precisam ser necessariamente estocados na hora de sua aquisição, podendo fornecer assim regras de conduta úteis para serem aplicadas de modo diferido, cuja simples aquisição pode até mesmo ser reforçada pela percepção satisfatória e motivadora de dominar e controlar situações anteriormente incertas. Em concordância, Seminário atribui ao fator motivação o sentimento de realização que a criança demonstra ao compreender e ser capaz de reutilizar determinada regra.

2.4.5 Diálogo e Motivação

Ao comparar a possibilidade da criação de uma escola, cuja ação estivesse amparada na técnica da Elaboração Dirigida, com a escola tradicional, a escola nova e a escola operatória,

Seminário enfatiza o diferencial de sua técnica ao apontar duas características fundamentais: o diálogo e a motivação. E explica: o metaprocesso deveria ser provocado através de diálogos entre aqueles que já possuem um modelo estocado e os que ainda não, possibilitando que esses construam e estoquem seus próprios modelos – num diálogo funcionando como processo facilitador da metalinguagem e do desenvolvimento cognitivo. Outra característica importantíssima é o fator motivacional, que atuaria de forma essencial para despertar o interesse da criança em aprender determinada tarefa, na conseqüente compreensão e domínio de determinada regra, acelerando seu desenvolvimento.

2.4.6 Metaprocesso Cognitivo

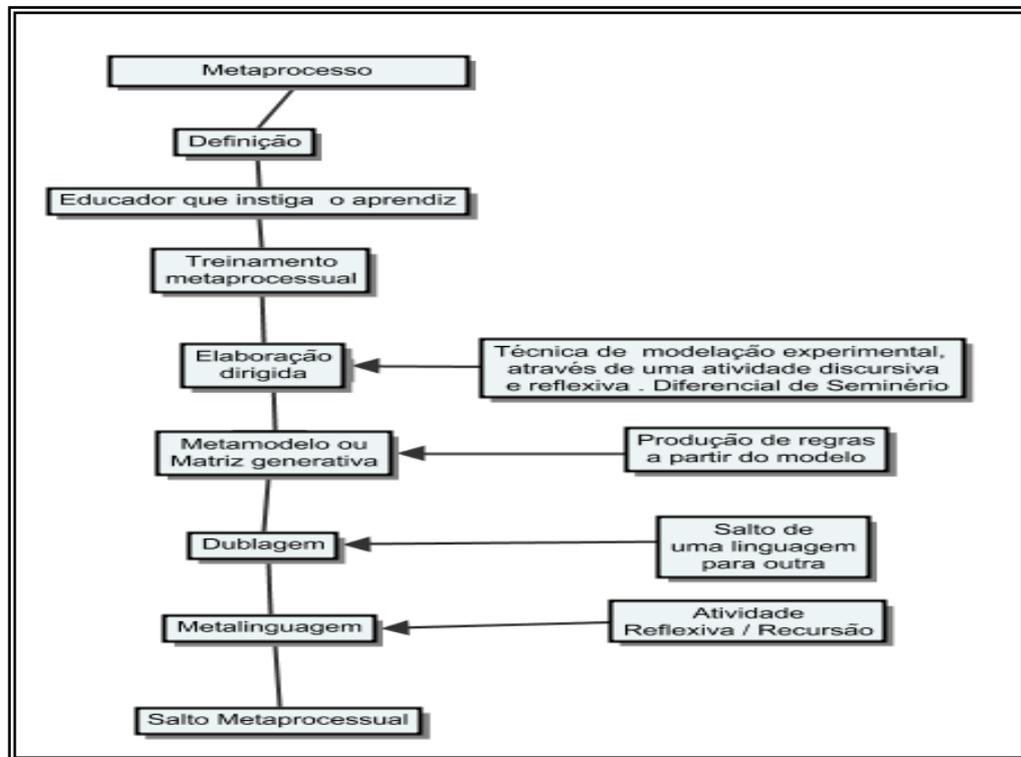


Figura 2. 5: Definição de Metaprocesso.
Fonte: (Marques *et al*, 2009b)

É o Metaprocesso que torna o ser humano um agente intencional; é o seu uso que torna capaz a diferenciação entre os humanos, principalmente quanto ao raciocínio lógico. Seminário se dedica à elaboração de uma técnica pedagógica para promover o Metaprocesso e modernizar a educação. Embora admita que a recursão metaprocessual que pode ocorrer de modo espontâneo, argumenta que o nível de elaboração mais elevado ocorre essencialmente pela relação interpessoal, valorizando assim a função do adulto no desenvolvimento cognitivo da criança. Em sua proposta, o educador instiga a criança, possibilitando-a, desde muito cedo, a fazer a dublagem de suas experiências, ajudando-a a raciocinar sobre o que fez, viu, ouviu, percebeu e até pensou, como ilustra a Figura 2.6.

É importante que este adulto-educador faça um treinamento de reflexão na criança para que esta possa aproveitar cada momento possível de inferência, levando a dublagem e ao treinamento de reflexão. Este tipo de treinamento desde a tenra idade é um fator de diferenciação primordial e o mais poderoso agente proposto por Seminério para o desenvolvimento cognitivo.

Durante a aplicação prática de sua teoria, Seminério comprova que, através da técnica “Elaboração Dirigida”, os educadores podem provocar o uso da atividade metaprocessual, possibilitando que as crianças desfavorecidas socialmente tenham a chance de ampliar seus códigos, melhorando sua capacidade de aprender e de refletir sobre suas experiências.

2.4.7A técnica da *Elaboração Dirigida*, seu legado e procedimentos

Nesta proposta, o desenvolvimento é visto como uma estocagem de paradigma nas sintaxes disponíveis, permitindo a metacognição, ou seja, um constante metaprocessamento ao nível da quarta linguagem. O propósito é promover o diálogo na educação, dando ênfase à atuação relacional. De forma mais específica, aponta-se o ensino da linguagem de programação com base num trabalho metaprocessual, com o objetivo de criar um ambiente provocador que possibilite o sujeito passar de um processo de síntese para abstração, através das relações interpessoais e de modelos cognitivos, esse tipo de modelação é mencionado por Bandura (2008), como transmissão de modelos que estabelece um modo amplamente raciocinado através do diálogo como ilustra a Figura 2.7.

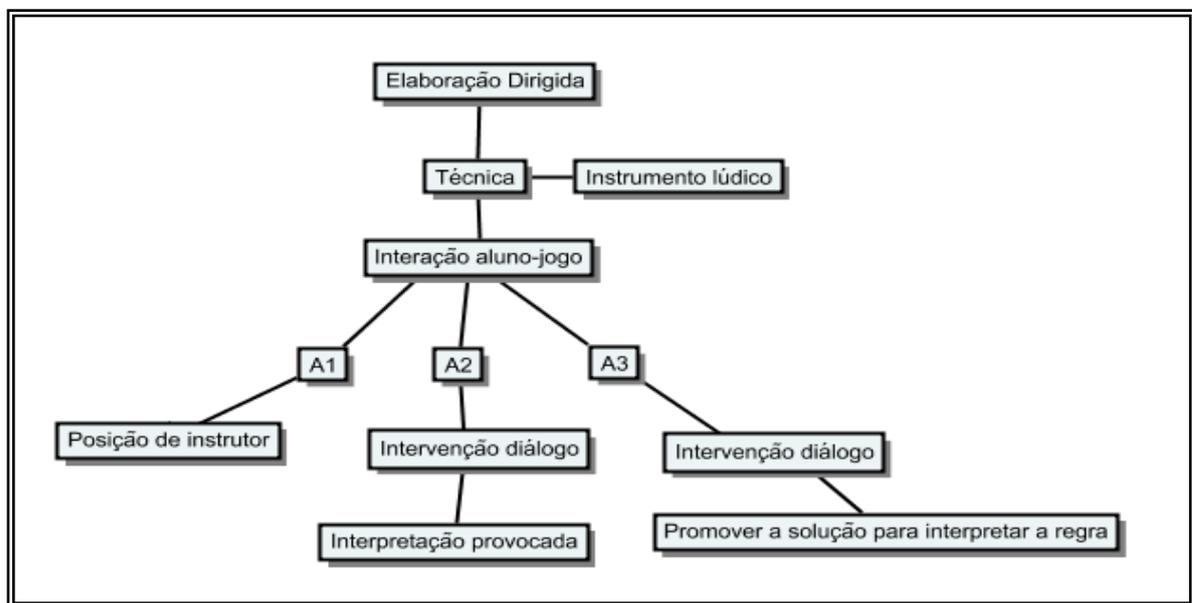


Figura 2. 6: Mapa Conceitual da Técnica.
Fonte: (Marques *et al*, 2009b)

Como instrumentos para o trabalho com a Elaboração Dirigida, Seminério estruturou jogos utilizando principalmente material de sucata, como tampinhas, palitos de fósforo, pedaços de papel, que por sua possibilidade lúdica, gerava significativa motivação nas crianças. Mas, para além da

simplicidade, os objetos escolhidos possibilitavam o trabalho com as regras lógicas clássicas, como seriação, classificação e inclusão. Os jogos de classificação visavam a trabalhar este aspecto como consequência da constância específica: os de inclusão como decorrência do encadeamento coordenado e os de seriação como implicação da diferença constante.

A ação do jogar envolvia três etapas:

- h) Apresentação do material a ser utilizado no jogo;
- i) Atividades são propostas, e o participante desenvolve as atividades sem nenhum tipo de intervenção. Exposição das regras para a correção das atividades; e
- j) Encaminhamento para a situação problematizada, reforçando e verificando o aprendizado do participante.

Em seguida, era solicitado à criança que solucionasse o problema proposto a fim de que fosse possível ao aplicador verificar sua atuação e, então, dar andamento ao trabalho de acordo com três alternativas previamente estabelecidas. As alternativas seriam:

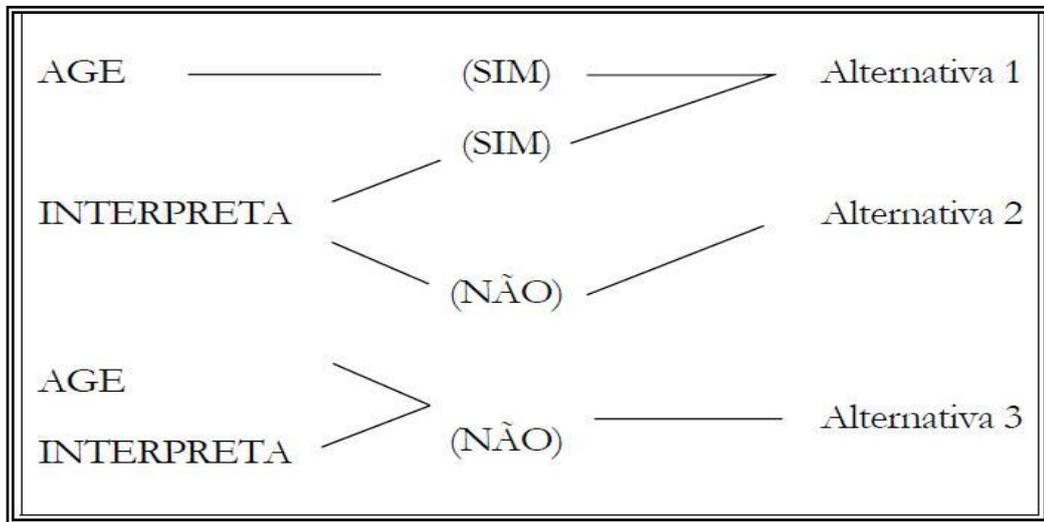


Figura 7: Alternativas de trabalho.
Fonte: (Marques *et al*, 2009b)

Alternativa 1 – Age e Interpreta:

A criança verbaliza as razões das soluções encontradas. Neste caso, é considerada a variação progressiva da complexidade do jogo.

Alternativa 2 – Age e Não Interpreta:

O experimentador solicita o critério genérico à criança, provoca a interpretação através de um diálogo capaz de desenvolver heurísticamente a descoberta das razões.

Alternativa 3 – Não Age e Não Interpreta:

A ação preliminar deve ser do experimentador que deve promover a solução do problema dado - por exemplo: CLASSIFICAR, SERIAR ou fazer a INCLUSÃO -, solicitando que a criança reflita sobre a operação realizada e auxiliando a interpretação desde até que mediante verbalização este

consiga entender a lógica do modelo que deverá lhe ser claramente detalhado e totalmente explicitado.

No caso desse trabalho ser realizado numa turma escolar, é esperado uma não homogeneidade. Sendo assim, o nivelamento deverá ser feito de baixo para cima, da não ação e não interpretação para ação e interpretação. As crianças mais adiantadas (as que alcançarem as duas primeiras alternativas), para que não haja desmotivação, se tornarão instrutoras, podendo, com ajuda do experimentador, trabalhar os passos progressivos do processo de elaboração e conscientização dos demais. Isto representa um eficaz trabalho de elaboração metaprocessual das crianças que assumirem esta posição.

Seminário procedeu a uma testagem desta natureza, com crianças entre sete e dez anos de três escolas, uma particular que atendia crianças da classe média e outras duas da rede municipal que atendiam crianças menos favorecidas. Foram realizados pré-testes e pós-testes, constituídos das provas clássicas de diagnóstico operatório de Piaget. A figura abaixo demonstra a utilização da técnica "before-after". O aprendizado é demonstrado a partir da diferença entre o pré-teste (Δ_1) e o pós-teste (Δ_2).

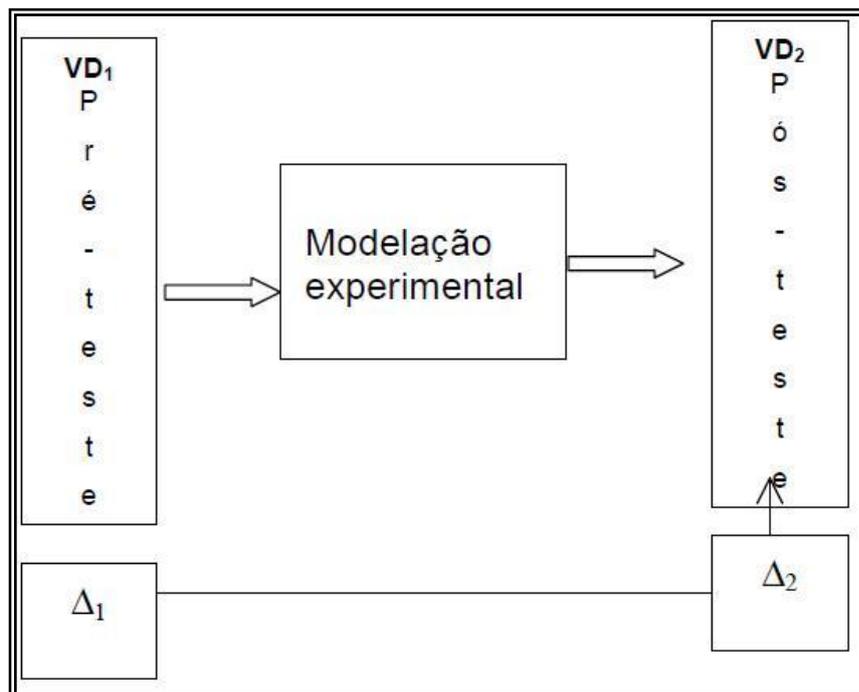


Figura 8: Técnica "before-after".
Fonte: (Marques *et al*, 2009b)

Os resultados mostraram avanços no desenvolvimento cognitivo das duas populações. Entretanto, o mais significativo ocorreu justamente com as crianças desfavorecidas de uma bagagem informacional, diferente das crianças da classe média que possuíam um acesso superior à

informação, não somente pela família, mas também pela própria escola, demonstrando a eficácia da Elaboração Dirigida na aprendizagem.

2.5 Fio Condutor

Formulado pela Prof^a. Carla Verônica Marques (2009b), o Fio Condutor é uma forma prática, mais completa, de se aplicar a Elaboração Dirigida no contexto de avaliação e reabilitação cognitiva. Facilitado por jogos especialmente criados para o trabalho de desenvolvimento metacognitivo, o Fio Condutor envolve sete momentos, chamados de fases ou versões. A sessão do jogo pode ser composta por dois ou mais atores, constituindo-se como mínimo 1 jogador e 1 aplicador. A imagem a seguir mostra uma tabela com as fases do Fio Condutor, e logo abaixo a descrição dos procedimentos envolvidos em cada uma.

FASES DE CONSTRUÇÃO E APLICAÇÃO DO PROTOCOLO/JOGO						
VERSÃO 1	VERSÃO 2	VERSÃO 3	VERSÃO 4	VERSÃO 5	VERSÃO 6	VERSÃO 7
AVALIAÇÃO ESTÁTICA	AVALIAÇÃO DINÂMICA	AVALIAÇÃO DINÂMICA	AVALIAÇÃO DINÂMICA	AVALIAÇÃO ESTÁTICA	INTERVENÇÃO	REPETIÇÃO DAS VERSÕES 1 E 2 COM TAREFAS DIFERENTES
APRESENTAÇÃO DO ESTÍMULO	MEDIAÇÃO PARA A CONSTRUÇÃO DA REGRA	CONSTRUÇÃO DA REGRA COM OBJETOS MANIPULÁVEIS	ELABORAÇÃO DIRIGIDA	RETESTE MEDIADO		

Figura 9: Fases do Fio Condutor.
Fonte: (Marques *et al*, 2009b)

Fase 1: Início do jogo, no qual o jogador tem seu primeiro contato com o jogo. Nenhuma regra é dada ao jogador, o jogo é iniciado a partir do entendimento intuitivo do jogador. O aplicador acompanha todas as ações do jogador. Nesta fase, o aplicador testa o comportamento e o estágio da criança.

Fase 2: O aplicador solicita ao jogador que explique o que fez anteriormente. O aplicador, então, registra as explicações dadas.

Fase 3: É dividida em duas partes (A e B). Na primeira, todas as peças do jogo são soltas para que o jogador arrume o tabuleiro para iniciar suas jogadas. Na segunda, o aplicador solicita ao jogador que explique o que fez anteriormente.

Fase 4: O jogador precisa verbalizar a regra. O aplicador é o mediador com intuito de que o jogador formule a regra do jogo e verbalize-a. O aplicador, então, registra tudo o que for falado. Esta fase caracteriza a aplicação da Elaboração Dirigida.

Fase 5: O jogo é iniciado novamente, e o aplicador verifica se o jogador conseguiu alcançar os objetivos. Esta fase funciona como um pós-teste para verificar o resultado da aplicação da Elaboração Dirigida.

Fase 6: Esta fase se caracteriza pela intervenção do especialista a partir dos dados coletados na aplicação da Elaboração Dirigida.

Fase 7: Nesta fase, o mesmo jogo é aplicado, mas com tarefas diferentes para confirmar o resultado da aplicação. Nesta fase, também é verificada a complexidade do jogo, podendo haver um aumento da complexidade.

A ação neuropedagógica promovida pela técnica da Elaboração Dirigida tem assegurada sua continuidade com as pesquisas da Prof^a. Carla Verônica e pela prática intermediada pelos mais de 1.000 jogos por ela criados sob a ótica da Revolução Cognitiva proposta por Franco Lo Presti Seminário (Marques, 2009).

2.6 Identificação automática de texto

Gênero tem recebido muita atenção nas escolas brasileiras devido à sua grande importância na aprendizagem de língua (Marcuschi, 2008). De acordo com o MEC, trabalhar com gênero e produção de texto é um caminho possível para resolver a questão do analfabetismo funcional (Soares, 1998). Quanto mais você está em contato com diferentes tipos de textos, seja produzindo ou lendo-os, você se torna letrado. Assim, este projeto tem como propósito, ajudar na produção de gêneros diferentes de modo a criar contexto e oportunidade para a prática da escrita, bem como fornecer *feedbacks* sobre o desempenho, que inicialmente, apenas dizem respeito à composição do gênero produzido.

Em primeiro lugar, gênero de texto ou o estilo de texto, neste trabalho e nos trabalhos pesquisados, caracteriza a finalidade para a qual o texto foi escrito. Exemplos de gênero são: artigos de pesquisa, romance, poema, reportagem, editorial, *homepage*, propaganda, manual, decisão judicial etc. Aplicações baseadas em texto estão se tornando cada vez mais comuns hoje em dia, diferentes aspectos do texto, tais como gênero, pode ser úteis para várias finalidades. Para não mencionar que, caracterizar texto diferente do assunto usual ou conteúdo preposicional, tem sido o foco de recuperação de informação e muitas pesquisas de classificação. Nessa seção, vamos abordar a questão da detecção automática da classe gênero de texto.

Classes de gênero são claramente diferentes das classes de assunto que a pesquisa de classificação mais tem trabalhado. Um conjunto de documentos pode pertencer à mesma classe porque eles compartilham um tema comum, mas muitas vezes estes servem a propósitos diferentes, classificados na classe gênero diverso. Assim, a classificação de documentos com base em gênero resultaria em um resultado totalmente diferente do que o da classificação por assunto comum. Do ponto de vista tradicional de recuperação de informações, uma consulta de recuperação sobre um determinado tema, como "esportes", deve recuperar muitos documentos relacionados, de fontes diversas, quando submetido a um motor de busca na Internet, que podem ser de gêneros diferentes,

como página sobre um esporte de canal de TV, esportes, notícias, publicidade, produto ou revisão crítica de um determinado jogo. O Gênero fornece uma nova dimensão para a recuperação e classificação de texto, e ajuda os usuários a se familiarizarem com a estrutura intrínseca necessária em textos de formatos diferentes.

Classificação automática de gênero foi estudada por Bretan *et al.* (1997), Dewe *et al.* (1998) e Dillon *et al.* (2000). Karlgren e Cutting (1994) exploraram o uso de pistas estruturais e sugestões bastante simples, como acusações de pronomes de terceira pessoa no texto com a análise discriminante. Em trabalho posterior (1996), eles investigaram a relação entre o gênero de documentos recuperados contra não recuperados e documentos relevantes e não relevantes.

Recursos utilizados são estatísticas simples, tais como o comprimento da sentença e extensão da palavra, e complexidade sintática, tais como profundidade média de uma árvore de análise. Identificar gênero de texto seria benéfico para muitas aplicações baseadas em texto. Por exemplo, se o gênero de cada documento é conhecido *a priori*, os resultados de recuperação de informação poderiam ser melhor apresentados ao usuário, dependendo da preferência do usuário. Como apontado por Kessler *et al.* (1997), o desempenho de muitas ferramentas de processamento de linguagem natural, como parte do discurso, marcação de sentido, analisador de palavra, podem ser reforçadas uma vez que alguns usos da linguagem embutidos em construções gramaticais e os sentidos da palavra são relacionados com o gênero de texto. Em aplicações Web, detecção de gênero ajudaria *wrappers* que tentam extrair as informações especificadas de semiestruturada. Kessler *et al.* (1997) identificaram pistas em quatro categorias: pistas estruturais (por exemplo, a contagem de *tags* POS), pistas lexicais (palavras por exemplo, usadas para expressar as datas), a personagem de nível de sinais (sinais de pontuação, por exemplo), e pistas de derivativos. Eles decidiram não utilizar as pistas estruturais devido ao alto custo computacional. Seus métodos computacionais foram regressão logística e redes neurais (uma percepção simples e multicamada de percepção) que combinam 55 pistas.

Mais recentemente, Stamatatos *et al.* (2000) relataram em seu trabalho para a detecção de gênero usando frequências de palavras e sinais de pontuação. Em vez de usar sofisticadas pistas linguísticas, eles tentaram desenvolver um método que funciona para texto irrestrito em qualquer domínio e da linguagem com o mínimo custo computacional na extração de pistas.

Lee e Myaeng (2004) tomou a posição mais relacionada com a recuperação de informação tradicional e abordagens de categorização de texto do que ao processamento de linguagem natural profundo para a identificação de gênero. Nossa análise de texto é baseado em seu trabalho.

Recursos gerados a partir da análise de texto são usados para a classificação baseada em gênero dos documentos. Nossa abordagem é semelhante ao de Lee e Myaeng (2004) na medida em que desenvolveram seus próprios desvios baseados em método de seleção estatística, utilizando característica assunto baseado em informações de classificação dos documentos de treinamento, e estamos seguindo seus passos. Por literatura doém, utilizando uma lista de critérios diferente, baseado na literatura de classificação de gêneros textuais.

Tendo desenvolvido um classificador de gênero, eles começaram a investigar sobre a questão de como gêneros textuais ajudam a classificação de documentos com base no conteúdo

temático dos documentos. Este é um corolário da nossa hipótese de que a estrutura linguística ajudaria a identificar a classe de gênero de um documento automaticamente, ajudando assim a melhorar a escrita, fornecendo um *feedback* de acordo com o gênero que optar por escrever.

2.7 Mineração de Texto

Nossa pesquisa não se limita à busca de padrões, portanto os processos de MT serão adaptados às nossas necessidades, visando única e exclusivamente a construir um identificador automático de gêneros textuais específicos para fins educacionais. Todavia, é importante descrever nosso entendimento sobre a área de Mineração textual, assim como detalhar suas etapas para por ventura a partir delas explicar nosso trabalho.

Mineração de textos refere-se ao processo não trivial de extração de padrões úteis e interessantes (conhecimento) a partir de um conjunto de documentos textuais não estruturados (FELDMAN, R. & H. HIRSH, 1997). A Mineração Textual inclui métodos inteligentes e ferramentas automáticas para auxiliar pessoas na análise de grandes volumes de textos a fim de minerar o conhecimento útil. Portanto, uma característica importante do processo de mineração de textos é que o conhecimento extraído seja compreensível a humanos. A Mineração Textual utiliza técnicas das áreas de Extração de Informação, Processamento de Língua Natural (PLN) e Recuperação de Informação, juntamente com algoritmos e métodos de *Knowledge Discovery in Database* (KDD), Aprendizado de Máquina e estatística. Será dada ênfase à área Aprendizado de Máquina e estatística, uma vez que esse é o caminho escolhido para a construção do objeto de medida desta dissertação.

TAN (1999) destaca que a Mineração de Texto é um processo complexo, pois trabalha com dados textuais que são inerentemente não estruturados e que, eventualmente, possuem ambiguidade. DÖRRE *et al.* (1999) consideram como o primeiro desafio da mineração de textos a manipulação e análise de informações textuais expressas em língua natural, que, inicialmente, não foram projetadas para serem processadas por computadores, mas para serem interpretadas por humanos. Pesquisas na área de Mineração de Texto têm focado em problemas de representação de textos, classificação, *clustering*, extração de informação e busca ou modelagem de padrões.

2.7.1 Etapas no processo de Mineração

O processo de Mineração de Texto pode ser dividido em quatro etapas fundamentais: coleta de documentos, pré-processamento, extração de padrões e avaliação dos resultados (pós-processamento), como apresentado na figura 11.



Figura 3.1: Etapas do Processo de Mineração de Textos.
Fonte: Morgado, 2009.

2.7.2 Coleta de Documentos

A primeira etapa consiste na busca de documentos relevantes ao domínio de aplicação do conhecimento a ser extraído. Esses documentos podem estar em livros impressos, na internet ou qualquer outra fonte de texto possível de ser digitalizada. Entretanto, é possível que os documentos coletados estejam em uma grande variedade de formatos e, dessa forma, pode ser útil converter tais documentos a um formato padrão como, por exemplo, o XML, antes de prosseguir para a etapa de pré-processamento.

Mais adiante, é apresentado em maiores detalhes o pré-processamento de textos, a extração de padrões, bem como as tarefas de categorização e análise de *link*.

2.7.3 Pré-processamento

A segunda etapa após a coleta de documentos denomina-se pré-processamento. Os documentos selecionados são formatados, pois eles serão submetidos aos algoritmos de extração automática de conhecimento.

O pré-processamento de textos consiste em um conjunto de transformações realizadas sobre alguma coleção de textos com o objetivo de fazer com que esses passem a ser estruturados em uma representação atributo-valor. De modo geral, a etapa de pré-processamento tem por finalidade melhorar a qualidade dos dados já disponíveis e organizá-los. A etapa de pré-processamento demanda a maior parte do tempo do processo de extração de conhecimento. Além disso, ela exige planejamento e processamento, pois durante a transformação dos textos em formato estruturado existe a possibilidade de que informação intrínseca ao conteúdo dos textos seja perdida. Um desafio, nesse caso, é obter uma boa representação minimizando a perda de informação. A etapa de pré-processamento em um processo de Mineração de Texto é, portanto, fundamental para o desempenho de todo o processo (MARTINS, 2003).

2.7.3.1 Remoção de caracteres indesejados

Em um texto são encontrados sinais, pontuações, numerações, caracteres matemáticos (“%”, “+”, “<” etc.), monetários (“\$”, “€” etc.), números, caracteres de formatação (retorno de carro, *newlines*, caráter de tabulação etc.), entre outros que devem ser retirados do texto por serem indesejáveis dentro do objetivo de uma análise. Para executar esta tarefa, é necessário identificar os caracteres que deseja excluir do texto, ou o contrário: identificar os que não devem ser excluídos e remover os demais. Com a extração de textos de *sites* através de um buscador automático essa tarefa se torna mais fácil, já que geralmente os caracteres indesejáveis são vestígios de HTML, que se repetem, portanto fáceis de identificar e excluí-los automaticamente.

2.7.3.2 Transformação das letras (Case Folding)

Essa etapa consiste na integração de todos os caracteres da etapa anterior dentro de uma mesma caixa para a utilização posterior: letras maiúsculas, letras minúsculas. No sistema desenvolvido, essa etapa foi excluída, pois a linguagem Python, como diversas outras linguagens de programação, acompanha em suas bibliotecas padrão mecanismos automáticos para a integração da caixa. Caso contrário, é possível alterar facilmente a caixa fazendo simples cálculos utilizando o código ASCII de cada caráter.

2.7.3.3 Toquenização

A tarefa de *toquenização* consiste na identificação dos termos dentro do texto e é utilizada para decompor o documento em cada termo que o compõe. Os delimitadores utilizados para *toquenização* geralmente são: o espaço em branco entre os termos, quebras de linhas, tabulações e alguns caracteres especiais, como por exemplo: (“ ”), que é o único caráter fora do alfabeto restante da etapa de remoção de caracteres indesejados, e que por natureza é um caráter separador de palavras. Cada grupo de caracteres entre espaços é denominado *token*, que é o termo em si. O texto, que é visto como uma sequência de caracteres, será transformado em uma coleção de *tokens*, como mostra o exemplo em (MORGADO 2009) “Venda de celulares cairá em 2009” Utilizando os espaços como separadores, os *tokens* (termos) da sentença acima serão os seguintes: “venda”, “de”, “celulares”, “cairá”, “em”, “2009”.

Esta tarefa é muito importante, uma vez que todas as etapas posteriores a este ponto utilizarão os textos como termos. Muitas linguagens de programação já acompanham bibliotecas e classes que auxiliam nesta etapa (biblioteca `strtok.h` em C, classe `StringTokenizer` em Java, `Tokenizer` for Python source etc.).

2.7.3.4 Retirada de palavras desnecessárias (Stopwords)

No processo de análise, dependendo do objetivo de análise, pode ser relevante a eliminação de palavras que não possuem importância significativa no texto, no intuito de limitar a quantidade de termos-índices, com a visão de se manter apenas os termos que representam realmente o contexto de cada documento.

São exemplos destas palavras artigos, preposições, conjunções, pronomes, tais como: de, assim, a fim de, agora, onde, outro, outros, ainda, a, o, que, vários, e, do, da, uns, em, um, para, é etc.

Na maioria dos casos, estas palavras possuem apenas a finalidade de conectividade entre termos (no auxílio à formulação de frases), não havendo assim a necessidade de adicioná-las na estrutura de índices. É importante ressaltar que a finalidade de conectividade entre termos pode não possuir importância significativa em pesquisas onde o objeto de análise não é o gênero textual, pois do contrário as conjunções em especial são termos que ajudam a representar realmente o contexto de cada documento e possuem, portanto, importância significativa no texto.

Outras palavras que podem ser tidas como irrelevantes são aquelas que aparecem com frequência na coleção de documentos. Desta forma, são consideradas incapazes de discriminar os mesmos, tornando-se desnecessária a permanência destas na estrutura de índices (WIVES, 2002). A existência dessas palavras no texto implicaria a perda de desempenho e qualidade nas etapas seguintes das tarefas de mineração de texto, por isso a importância de execução dessa etapa.

2.7.3.5 Redução ao menor radical de cada palavra (Stemming)

A tarefa de *stemming* consiste em reduzir um termo ao seu radical. Este processo é de grande importância para as etapas posteriores.

Um *stem* é um grupo natural de termos que compartilham interpretações semânticas iguais ou similares. O processo de *stemming* é responsável por reduzir as diversas formas de um termo a uma forma comum (raiz) ou *stem*, que é um grupo natural de termos que compartilham interpretações semânticas iguais ou similares.

Os algoritmos de *stemming* aplicam uma série de normalizações linguísticas para remover sufixos de termos, ou inclusive mapear verbos para a sua forma no infinitivo. Por exemplo, na Língua Inglesa, os termos *gets*, *got*, *getting* e *gotten* são reduzidos ao seu radical único, o *stem get*, que expressa o significado comum aos quatro termos.

Com essa técnica, diminui-se o tamanho do dicionário, isto é, o número de termos distintos que representam o conjunto de documentos fica menor, resultando em redução no espaço de armazenamento necessário e no tempo de processamento, pois termos com o mesmo radical poderão ser processados em conjunto, pois palavras cujo radical é o mesmo serão consideradas como equivalentes.

Através da Figura 3.2 pode ser visualizada a aplicação do processo de *stemming* com método mais conhecido desenvolvido para esta língua, o Removedor de Sufixo da Língua Portuguesa (RSLP) é composto de oito passos. Pode ser observado que o mesmo diminui consideravelmente a quantidade de termos, possibilitando, assim, uma melhora na etapa de criação da estrutura de termos-índices (ORENGO et al., 2001).

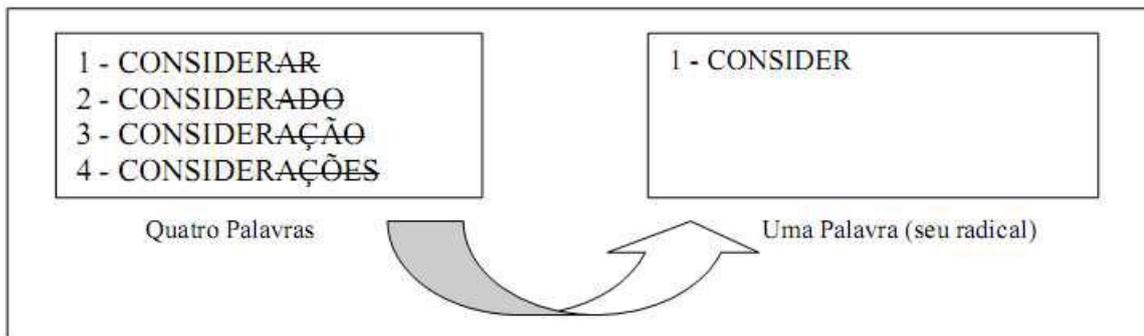


Figura 3. 10: Técnica *Stemming*.
Fonte: Morgado, 2009.

Embora seja uma ótima ferramenta para melhorar o desempenho da aplicação de mineração de texto, o RSLP, por se tratar de um processo linguístico, a fase de *stemming* pode apresentar falha, uma vez que é muito difícil englobar todas as exceções de um idioma em uma lista. Também é muito difícil estabelecer regras para remoção de sufixos que contemplem todas as palavras de uma língua. Há ainda a possibilidade de existirem palavras digitadas erroneamente, e isto acarretará em um radical incorreto.

Para contornar esse problema, verifica-se o radical ao final da etapa de *stemming* para corrigir. Isto é feito com o auxílio de uma lista que indica o radical correto caso a palavra tenha sido reduzida a um radical incorreto contido na lista.

2.7.3.6 Dicionário de dados (*Thesaurus*)

Para melhorar os resultados de uma aplicação, outra possibilidade é utilizar um dicionário de dados que correlacionam palavras diferentes e comuns a uma única palavra em todo o texto, ou seja, montar uma relação de várias palavras para uma única palavra que possa substituí-las sem alterar o contexto. Como exemplo, as palavras “data”, “Agosto”, “24/06”, que poderiam ser associadas a uma única palavra que é “data”. Outro exemplo que se relaciona com uma aplicação astrológica, são as palavras “touro”, “ascendente”, “Vênus”, “signo”, que foram padronizadas como “horóscopo” nessa pesquisa.

2.7.4 Extração de Padrões

Para cada método e aplicação de mineração existe um tipo diferente de extração de conhecimento de textos. Esses métodos precisam ser definidos para que as tarefas possam ser executadas a fim de identificar padrões e relacionamentos entre os documentos. Dentre as principais tarefas relacionadas ao processo de Mineração de Texto, escolhemos trabalhar com a categorização, uma técnica que visa a identificar os tópicos principais em um documento e a associar este documento a uma ou mais categorias predefinidas (YANG *et al.* 1997) e com o agrupamento, uma técnica que busca agrupar, em um ou mais grupos, um conjunto de exemplos de acordo com a similaridade ou dissimilaridade de seu conteúdo. Decidimos não aplicar a técnica da sumarização de Mineração de Texto, que identifica os termos e frases mais importantes de um documento, ou conjunto de documentos, gerando a partir destes um resumo ou sumário (FAYYAD *et al.*, 1996) com o objetivo de reduzir a quantidade de texto de um documento, porém mantendo seus significados-chaves. Em nosso trabalho de pesquisa, a redução da quantidade de texto poderia ainda resultar na perda de elementos significativos para a extração de conhecimento.

2.7.5 Avaliação e Interpretação dos Resultados

Por fim, a última etapa do processo de Mineração de Texto é responsável pela avaliação e interpretação dos padrões extraídos. Esta etapa visa a constatar se o objetivo almejado foi alcançado, ou se todas ou algumas etapas do processo necessitam ser refeitas. Os padrões descobertos podem ser avaliados pelo usuário final, especialista do domínio e analista de dados, com o intuito de validar o conhecimento obtido.

Várias medidas objetivas são utilizadas para estimar o desempenho de um algoritmo de mineração de textos. Frequentemente, os resultados obtidos pelos algoritmos de extração de informação são, estatisticamente, avaliados em termos das métricas *abrangência*, *precisão* e *F-measure*.

Embora o trabalho do especialista, auxiliando e resolvendo possíveis situações de conflito, é essencial ao longo do processo de extração de conhecimento em mineração de textos, as técnicas e ferramentas para visualização de dados visam a melhorar a compreensão dos resultados obtidos e a comunicação entre os usuários, visto que ferramentas de visualização que geram diversas formas de visualização (árvores, regras, gráficos 2D, 3D) combinadas com técnicas de mineração de textos podem melhorar muito o processo de Mineração de texto (MORGADO, 2009).

2.8 Tarefas no processo de Mineração

Como fora mencionado anteriormente, cada tipo de tarefa de Mineração de Texto extrai um tipo diferente de informação dos textos com o objetivo de descobrir conhecimento útil e inovador dos

textos, que sejam úteis para o objetivo da pesquisa. Nos próximos itens, iremos detalhar mais as tarefas que utilizamos nessa pesquisa.

2.8.1 Categorização de Textos

A categorização trabalha com categorias predefinidas e, através da identificação de tópicos principais em um documento, esse processo tenta identificar a qual domínio ou categoria pertence.

O trabalho do professor de português em sala de aula é exatamente esse, ao analisar os textos de seus alunos e verificar se pertencem ou não ao padrão predefinido dos gêneros em questão. Porém, esse processo manual é muito custoso. Assim como também é para o *site* de buscas *Yahoo*, por exemplo, que possui mais de duzentos especialistas para rotular manualmente ou categorizar os *sites*, que recebe centenas de páginas diariamente (AL-SHALABI *et al.*, 2008).

Então, o interesse na área de categorização automática de textos tem ganhado muito interesse também pelo fato de estar conseguindo atingir níveis de desempenho competitivos em relação às classificações realizadas por profissionais treinados (SEBASTIANI, 2002). Suas vantagens são a acurácia comparável com a alcançável por especialistas humanos, e a economia da mão de obra do especialista, pois a intervenção de um engenheiro de conhecimento ou especialista do domínio não é mais necessária para a construção do classificador ou para encaminhar a um conjunto diferente de categorias.

As categorias são escolhidas para corresponder aos tópicos ou temas dos documentos. O principal objetivo da categorização de textos é a organização automática. Pode um sistema categorizador retornar uma única categoria para o documento, ou categorias múltiplas. Em ambos os casos, um categorizador pode retornar nenhuma categoria ou algumas categorias com confiabilidade muito baixa. Nestes casos, o documento é normalmente associado a uma categoria rotulada como “desconhecida”, para posterior classificação manual. (LOPES, 2004)

A categorização de textos resume-se em uma ferramenta que classifica automaticamente um conjunto de documentos em uma ou mais categorias preexistentes, com a finalidade de recuperar informação, como pode ser visualizado na Figura 3.3.

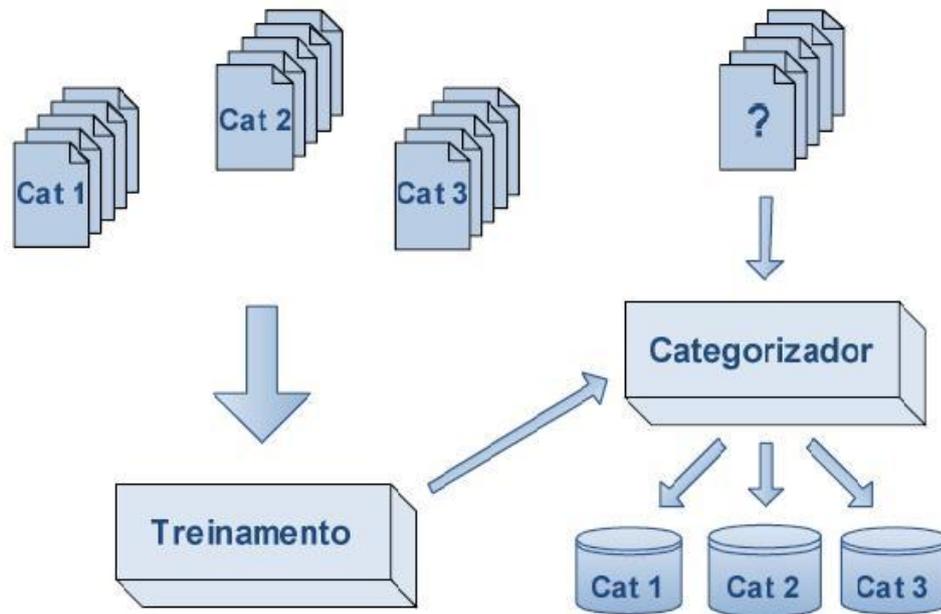


Figura 3.3.: Processo de Categorização.
Fonte: Morgado, 2009.

A categorização de documentos se dá a partir de algoritmos que têm sido desenvolvidos e dividem-se em duas categorias gerais: os algoritmos de aprendizagem de máquina (DUMAIS, 1998) e algoritmos de categorização especializados e desenvolvidos a partir da área de recuperação de informação.

No presente trabalho, são realizados experimentos com conjuntos de regras, classificadores baseados em exemplos e classificadores bayesianos.

2.8.1.2 Classificador Bayesiano

Destaca-se neste trabalho o método bayesiano. É baseado no Teorema de Bayes, que é usado na inferência de estatística para atualizar estimativas da probabilidade de que diferentes hipóteses sejam verdadeiras, baseando-se nas observações e no conhecimento de como essas observações se relacionam com as hipóteses. O Teorema de Bayes pode ser representado da seguinte forma:

$$p(A|B) = \frac{p(B|A)p(A)}{p(B)}$$

O teorema determina que, para eventos A e B, a probabilidade de A, dado que B tenha acontecido denotado por $p(A|B)$ é igual à probabilidade de B, dado que A denotado por $p(B|A)$ tenha

acontecido, vezes a probabilidade de A acontecer $p(A)$, dividido pela probabilidade de B acontecer $p(B)$ (GIBBS, 2003).

Segundo Morgado, (*apud* DORNAN, 2004, p.64) na informática, muitos dos sistemas de classificação automática são baseados no Teorema de Bayes. Inicialmente, o sistema é treinado aceitando entrada de humanos que dizem que uma dada entrada pertence a determinado grupo. Com o tempo, o sistema acumula um grande banco dessas informações e, aplicando o Teorema de Bayes, consegue estimar a probabilidade de cada novo dado de pertencer a cada grupo já classificado.

O classificador Naive Bayes utiliza dados de treinamento para estimar a probabilidade de um documento pertencer a determinada classe, os documentos com seus respectivos pesos são utilizados para a classificação, para cada termo é feita a probabilidade de esse pertencer à categoria. De acordo com a regra Bayes, é feita uma combinação de probabilidades levando em consideração o peso dos termos, o documento só será inserido na categoria quando o resultado for maior que o coeficiente determinado.

2.8.1.3 Classificador ORANGE

Segundo informações do site Orange (2010) o *ORANGE* é uma biblioteca de objetos centrais C++ e rotinas que inclui uma grande variedade da máquina padrão e não tão padrão de aprendizagem e algoritmos de mineração de dados, além de rotinas para entrada de dados e manipulação. O *ORANGE* é também um ambiente para prototipagem rápida de novos algoritmos e sistemas de testes. É uma coleção de módulos baseados em Python. Isto inclui uma variedade de tarefas, tais como impressão de árvores de decisão, subconjunto de atributos, incluindo o conjunto completo de componentes para dados de pré-processamento, classificação e filtragem de recurso, a modelagem, a avaliação do modelo e técnicas de exploração. É implementado em C (velocidade) e Python (flexibilidade).

O processo de classificação de textos baseado no *ORANGE* é experimental e orientado para a organização das categorias em hierarquias. Neste tipo de classificação, características particulares dos documentos, tais como mediana e dicionário de termos, são considerados pelo método. Os documentos são divididos em conjuntos definidos como base de treinamento e de teste. A base de treinamento é usada para o algoritmo de classificação obter as características das categorias da coleção. A base de teste valida o desempenho do classificador, determinando as categorias às quais os novos documentos pertencem.

Na fase de análise, o desempenho do algoritmo é medido de acordo com o resultado obtido na classificação original dos documentos, o que é feito através de classificação manual.

2.8.2 Agrupamento

O agrupamento tem sido amplamente utilizado em padrões de análises exploratórias de dados, tomada de decisão e situações de aprendizado de máquina, incluindo mineração de dados, recuperação de informação, segmentação de imagens e classificação de padrões. (MORGADO, 2009)

Técnicas de agrupamento buscam agrupar em um ou mais grupos, um conjunto de exemplos similares entre si. Ao final do processo, cada grupo irá conter um representante que permite identificar aquele grupo. O resultado do é dependente da medida de similaridade (critério) adotada. Conforme descrito em Morgado (*apud* Jain and Dubes, 1999), para se executar uma tarefa de agrupamento, normalmente os seguintes passos devem ser seguidos:

- Selecionar as características;
- Estabelecer uma medida de similaridade;
- Determinar um critério de agrupamento;
- Escolher um algoritmo de agrupamento;
- Validar os resultados do agrupamento;
- Interpretar os resultados.

A medida de similaridade faz-se necessária ao selecionar e identificar as características dos exemplos que serão agrupadas. Já a segunda etapa do processo de agrupamento consiste em analisar todos os exemplos com o objetivo de selecionar semelhanças entre eles, através de uma fórmula de similaridade que analisa todas as características semelhantes que os exemplos possuem e retorna um valor que indica o grau de semelhança eles.

Conforme descrito em Morgado (*apud* COLE, 1998), para agrupar objetos de acordo com sua similaridade, deve-se definir uma medida de quão próximos dois objetos estão, ou quão bem seus valores se comparam. Uma pequena distância entre os objetos deve indicar uma alta similaridade. Na literatura, existem várias medidas de similaridade, tais como Euclidiana, Manhattan e Minkowski.

Dentre essas medidas (COLE, 1998), destaca-se a distância Euclidiana, que é a mais utilizada. Depois de concluída a etapa de cálculo de similaridades (ou distância), tem-se uma matriz que indica os valores de similaridade entre os exemplos. Com base nesta matriz, é possível identificar os grupos de exemplos, especificando algum tipo de regra de relacionamento entre os mesmos e então escolher um dos tipos de agrupamentos (tais como, *K-means*, *Cobweb* etc.). Por fim, os resultados do agrupamento são validados e interpretados com o intuito de verificar o desempenho do processo.

Existem dois tipos de agrupamento quando nos referimos à MT: um baseado em termos, e outro, em documentos. No primeiro tipo, os grupos de termos similares são identificados, com o intuito de construir um dicionário de termos que definam assuntos similares. Em contraste, o agrupamento por documentos visa a identificar os documentos de assuntos similares e alocá-los em um grupo. (MORGADO, 2009)

Então, a finalidade do agrupamento é identificar os exemplos, tais como documentos ou termos que apresentem características em comum e agrupá-los em subconjuntos de exemplos similares, podendo ser o mais variável possível antes do agrupamento.

A caracterização de texto de forma diferente da usual que busca assunto ou conteúdo proposicional tem sido o foco da área de recuperação de informação (RI) e pesquisa de classificação. Nossa pesquisa aborda a questão sobre a detecção automática da classe de gênero de texto.

Classes de gênero são claramente diferentes de classes de assunto que pesquisa de classificação mais lidou. Embora um conjunto de documentos pertença à mesma classe porque eles compartilham um tema em comum, muitas vezes servem também propósitos diferentes, classificados em classes de gênero diversas. Como tal, a classificação de documentos com base no gênero levaria a um resultado totalmente diferente do que da classificação por assunto comum. Desde o tradicional ponto de vista de recuperação de informação, recuperação de uma consulta sobre um determinado tema, como "esporte", iria recuperar vários documentos relacionados com muitas fontes diferentes das coisas, quando submetidos à busca na Internet, mas eles podem ser de diferentes gêneros, como uma página de desporto, notícias de esportes na TV, propaganda, produto para venda ou até mesmo uma resenha crítica de um determinado jogo. Gênero fornece uma nova dimensão para a recuperação e classificação de texto, além de atual, ajuda os usuários a se familiarizarem com a estrutura intrínseca necessária em diferentes formatos de textos.

3. A CONSTRUÇÃO DO OBJETO DE MEDIDA

Para a construção do Objeto de Medida foram estudados diversos processos utilizados para descobrir conhecimento útil em uma coleção de documentos, como, por exemplo, o processo de Mineração de Texto, uma área multidisciplinar que incorpora técnicas de diversas áreas, como Recuperação de Informação, Aprendizado de Máquina, Estatística, Linguística Computacional, Extração de Informação, Visualização e especialmente Mineração de Dados.

Neste Capítulo, serão apresentados os estudos e trabalhos de pesquisa que abordam a classificação de gênero de uma maneira mais sistemática, como um espaço de propriedades multidimensionais, buscando identificar os gêneros como fontes confiáveis de evidências computacionais plausíveis de rastreamento.

A literatura tradicional de Linguística Textual é rica em abordagens do gênero por meio de esquemas classificatórios, material tipográfico, sobre tipos de texto. Encontrado em Koch (2002), Dolz & Scheneuwly (2004), Meurer, Bonini, & Motta-Roth (2005), Travaglia (2003), Karkoswki *et al.* (2007), Marcuschi (2008), Bakhtin(1992). Nesta pesquisa, iremos nos referir aos atributos usados para classificar os gêneros como critérios, assim como em Kessler *et al.* (1997).

Um critério é uma propriedade distintiva de uma classe de textos que responde a certos interesses práticos e é, além disso, associada a uma estrutura computacional ou linguística, ora categórica ora estatística, que descreveremos como critérios. Todavia, um pouco diferente da estratégia de Kessler (1997), que ignora o *framework* linguístico já levantado na literatura para gêneros específicos com a intenção de não restringir a categorização e criar seu próprio *framework* que auxiliará no trabalho de recuperação de informação e estender sua classificação para gêneros desconhecidos.

Nesta pesquisa, levantamos os critérios com base em referências teóricas para organizar e centralizar os critérios de sete gêneros textuais corriqueiros e essenciais para a formação de um indivíduo letrado. Por isso, também providenciamos um caminho mais sistemático de trabalhar o tema da diversidade textual, ao mesmo tempo em que fornecemos uma maneira de comprovar esses critérios com o auxílio da matemática estatística e do computador.

O primeiro trabalho de pesquisa linguística sobre gênero que utiliza métodos quantitativos é o de Biber (1986), (1988), (1992),(1995), que apresenta um trabalho de análise estilística, índice de leitura e diferenças entre linguagem escrita e oral.

Biber organiza os textos em diferentes "dimensões textuais" construídas pela aplicação da análise textual de um grupo de características sintáticas e lexicais, caracterizadas em termos como "informativo" vs "envolvidos" ou "narrativa" vs "não narrativa". Esses fatores não são usados para classificação de gênero (os valores de um texto em várias dimensões são frequentemente não informativos com respeito ao gênero). Ao invés, fatores são utilizados para validar hipóteses sobre as funções de várias características linguísticas.

Já Karlgren and Cutting (1994), ambos começam com *corpus* de textos classificados manualmente, o "Corpus Brown", e muitas das características textuais que eles usam são relacionadas à superfície linguística do texto, assim como nos propomos fazer nesta pesquisa.

3.1 Identificando o Gênero

Passemos, agora, a entender os critérios genéricos, que são propriedades observáveis de textos associados a um gênero. Em Kessler *et al.* (1997) e Karlgren e Cutting (1994), podem ser estruturais, lexicais, de pontuação e derivacionais. Nesta pesquisa, optamos por descartar o critério de pontuação, devido à dificuldade de estabelecer parâmetros fixos aptos a esse aspecto textual, para nosso propósito de pesquisa.

Os problemas essenciais que queremos abordar aqui estão ligados à composição (plano da redação) e à escolha vocabular, "técnica de uma formulação verbal que dispense os elementos extralingüísticos e elocucionais, só participantes da exposição oral" (Câmara 2001,p.62). Assim, a pontuação, a ortografia, a concordância, a acentuação, constituem os problemas secundários que, para Mattoso Câmara, são mais fáceis de resolver.

3.1.2 Critérios Estruturais

Exemplos de critérios estruturais são voz passiva e ativa, nominalizações, frases topicalizadas, e registro da frequência de características sintáticas (*tags* de discurso). Esses critérios não são muito discutidos na literatura tradicional de gêneros, mas foram colocados em evidência no trabalho de identificação automática de textos de Biber (1995), Karlgren and Cutting (1994). Porém, o processo é limitado por precisarem de textos com *tag* e *parse*. E, por isso, resolvemos não utilizá-los, pois o grau de programação seria bem maior do que estávamos preparados. E o impacto a certa dessa decisão, se limita ao nível da eficácia do analisador, independente do processo de aprendizagem do aluno.

3.1.3 Critérios Lexicais

A maioria dos critérios está relacionada a pistas lexicais, como datas (ex: 25 de maio, 1996), que predominam em notícias; ingredientes e quantidades de medida (ex: ½ xícara de açúcar) presentes em receitas; desinências modo-temporais e número pessoais (ex: -mos, -va,-sse); conectivos (ex: embora, todavia, porém); adjetivos (ex: feliz, morto etc.); e advérbios (felizmente, calmamente etc.).

3.1.4 Critérios Derivacionais

Critérios derivacionais são a mediana de variação advindas da medida das características lexicais e estruturais e de pontuação quando houver. Médias que se relacionam de alguma maneira a gêneros textuais. Representamos médias como a soma de outras características, transformando-as

em um valor numérico. Nesta pesquisa, as medianas estão relacionadas ao tamanho do texto e a frequência de aparecimento de certos critérios, ou seja, quanto mais frequentes maior será o valor atribuído a eles.

Alguns critérios estabelecidos para cada gênero, na fase de teste, não se mostraram significativos, retornando na maioria das vezes baixa ocorrência. Então, decidimos retirá-los do analisador e reforçar o valor atribuído aos critérios mais frequentes a determinados gêneros.

3.1.5 Características da superfície linguística

Diferente da maioria do referencial teórico encontrado em ciência da computação, nesse trabalho resolvemos buscar na Linguística textual o esquema de classificação dos gêneros já pesquisado por linguistas nacionais e citados em diversos tipos de trabalho, desde artigos até livros e teses de doutorado. E com o material pesquisado testar a veracidade de suas afirmações através da ciência da computação e dos mecanismos que ela nos oferece.

A primeira etapa desse trabalho foi construir um *corpus* com as características mais citadas como pertencentes aos gêneros que decidimos trabalhar. Seguiu-se à elaboração do mecanismo de teste para averiguação da verossimilhança desses dados com os textos do mundo real para que, a partir dos dados mais confiáveis, fosse construída uma regra genérica que sirva como apontador automático de gêneros baseado na superfície linguística e conteúdo desses gêneros. E finalmente testá-lo com variados textos, inclusive o *input* de textos produzidos por crianças para a verificação da acurácia desses resultados e a possibilidade de trabalhar com esse classificador como complementação escolar.

Um dos teóricos que mais contribuiu com a primeira etapa da construção desse objeto de medida foi Travaglia (2003/2007a) com sua teoria de caracterização de gêneros textuais e o trabalho de pesquisa de seus orientandos.

Segundo Travaglia, existem cinco critérios para caracterizar gênero textual, que são:

- k) conteúdo temático = o conteúdo a cerca do tema desses gêneros.
- l) estrutura composicional = como o gênero está organizado.
- m) objetivos e funções sociocomunicativas = objetivos ou funções sócio comunicativas são identificados como um ato ou macro-ato de fala.
- n) superfície linguística = elementos composicionais de formulação da seqüência linguística, do que muitos chamam de superfície linguística. Essas características podem referir-se a qualquer plano da língua (fonológico, morfológico, sintático, semântico, pragmático) ou nível (lexical, frasal, textual).
- o) condição de produção = O critério de “quem produz” inclui tanto o indivíduo (geralmente ocupando um lugar social) como a comunidade discursiva.

A caracterização das categorias de texto é feita por uma conjugação de critérios que pode ocorrer de diferentes modos e, muitas vezes, a distinção depende de uma combinação diversa dos

expressa a opinião do autor sobre um determinado fato cultural, que pode ser um livro, um filme, peças teatrais, exposições, <i>shows</i> etc.																				
11. Crítica direta.																				
12. Texto não é espontâneo nem natural.																				
13. Uma obra em que o autor cria os diálogos e as situações que envolvem os falantes, previamente.																				
14. Divide-se em estágios na maioria dos casos uma verdade, uma crença arraigada na cultura popular, que ajuda a fazer o leitor recuperar um intertexto, evidentemente, pretendido pelo autor do texto.																				
15. A relação entre texto, contexto, conhecimento de mundo e coerência.																				
16. Deve ser memorável (capacidade da frase ser lembrada sem auxílio).																				
17. A grande idéia deve ser transmitida no anúncio.																				
18. O conteúdo é sempre Algo a ser feito e/ou como ser feito, uma ou várias ações ou fatos e fenômenos cuja realização é pretendida por alguém.																				
19. Dar uma opinião, uma sugestão ou debater um tema.																				
20. O tema é o cotidiano e que leva o leitor a refletir sobre a vida comentando assuntos do dia a dia.																				
21. Divulgação de um acontecimento.																				
22. Venda ou oferta de um produto ou serviço, a fim de troca comercial ou divulgação.																				

Quadro 3.1: Conteúdo temático

ESTRUTURA COMPOSICIONAL

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. É um título no alto da primeira página de jornal ou revista.														
2. Possui quatro parágrafos.														
3. Traz informações novas e inéditas.														
4. Possui tema específico e um autor.														
5. Faz um posicionamento crítico.														
6. Traz justificativas.														
7. Traz exemplos concretos e dados factuais.														
8. Tem uma conotação negativa.														
9. Apresenta um consenso de várias opiniões sobre determinado assunto.														
10. Apresenta opiniões de terceiros.														
11. Traz o testemunho direto de fatos.														
12. Apresenta argumentos.														
13. O texto traz dados bibliográficos.														
14. Faz uma divisão em Capítulos ou seções.														
15. Utiliza de 3 a 5 parágrafos.														
16. Utiliza o quadro, o balão e as legendas.														
17. Metadiscursão, através de marcadores pessoais, hedges e enfatizadores.														
18. Evita a linguagem atitudinal normalmente associada ao significado interpessoal.														
19. Deve incluir um benefício-chave, deve diferenciar a marca, lembrar o nome da marca fazer rima com ele.														
20. Transmitir sentimentos positivos sobre a marca.														
21. Faz uma propaganda, um pedido, ou outros tipos de comunicação.														
22. Sequência ordenada de procedimentos que deverão ser realizados Informação sobre o tempo de preparo a quantidade que será produzida.														
23. Especifica o assunto e é breve; trace previamente o objetivo da carta (opinar, sugerir, debater).														
24. Narra fatos históricos em ordem cronológica, Trata de temas da atualidade.														
25. Trazem as pessoas comuns como personagens, sem nome ou com nomes genéricos sem aprofundamento psicológico.														
26. O comentário do tipo expectativa e o resultado do tipo reações verbais.														

Quadro 3.2: Estrutura Composicional

OBJETIVOS E FUNÇÕES SOCIOCOMUNICATIVAS

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Criticar.														
2. Chamar atenção para um determinado assunto.														
3. Exposição de algum assunto de interesse mútuo.														
4. Aprofundar mais em uma notícia ou determinado tema.														
5. Veículo pelo qual o autor expressa sua opinião sobre determinado assunto ou o resultado de estudos e pesquisas de determinado assunto.														
6. Posicionamento da empresa de comunicação sobre determinado assunto que direciona as matérias publicadas por ele.														
7. Formato de divulgação de um acontecimento Excepcional, anormal e de grande impacto social.														
8. Guiar o leitor pelo emaranhado da produção cultural que cresce a cada dia e que tende a confundir até os mais familiarizados com todo esse conteúdo.														
9. Expressar a opinião do autor sobre um determinado fato cultural, que pode ser um livro, um filme, peças teatrais, exposições, <i>shows</i> etc.														
10. Análise de forma crítica: sua opinião.														
11. Recomenda uma obra.														
12. Entreter, fazendo ou não críticas.														
13. Finalidades de entretenimento e aconselhamento.														
14. Deve ser memorável (capacidade de a frase ser lembrada sem auxílio).														
15. Faz uma propaganda , um pedido, ou outros tipos de comunicação que interesse ao público em geral, necessitando ou não de uma resposta para o autor.														
16. Fazer com que o destinatário possa fazer coisas a partir da leitura da receita.														
17. Opinar, sugerir ou debater um determinado tema.														
18. Analisa os fatos que acontecem e dá um colorido emocional, mostrando aos olhos do leitor uma situação comum, vista por outro ângulo, singular.														
19. Envolver, emocionar o leitor.														
20. Fazer o posicionamento da empresa de comunicação sobre determinado assunto.														

Quadro 3.3: Objetos e funções sócios comunicativas

CARACTERÍSTICAS DA SUPERFÍCIE LINGUÍSTICA

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. A omissão das palavras desnecessárias (adjuntos adnominais e adverbiais).														
2. Usa abr ATEH (para até), IMPAGO (não pago) e outras.														
3. a) introduzir um argumento apresentado como acréscimo, um argumento a favor de uma determinada conclusão, além disso, ainda mais, além de, e mais, e, também, nem, nem mesmo, ademais, não apenas... Mas, não apenas... Mas também, e não só por que... Mas também, não só... Mas também, sobretudo, e também, além, mais uma vez, mais ainda, aliás, ainda, e ainda.														
4. b) assinalar uma oposição, mas, mais ainda, mas também, entretanto, porém, contudo, todavia, do contrário, no entanto, agora, ao contrário. - embora, mesmo que, apesar de, ainda que.														
5. Conectores para os mais diferentes tipos de relações: Conjunção, contra junção ou oposição, disjunção ou alternância, causa /consequência (explicativas, justificativas, causais, condicionais, finais, consecutivas, conclusivas), comparação (comparativas, conformativas), comprovação, especificação (incluindo a exemplificação) entre outras.														
6. Uso de adjetivos e gírias.														
7. Descrição.														
8. Argumentação.														
9. Apresenta de 3 a 5 parágrafos 5.														
10. Discurso direto, onomatopéia, expressões populares.														
11. Prolongamento e intensificação de sons.														
12. Linguagem persuasiva.														
13. Construções condicionais com conectivo explícito, isto é, com se ou implícito.														
14. Lembrar o nome da marca e o ideal é que o nome da marca seja incluído na frase.														
15. Linguagem popular como um chavão ou aliteração relevante e palavras cunhadas ou compostas.														
16. Trocadilhos e rimas.														
17. Faz propaganda, um pedido, ou outros tipos de comunicação.														
18. Texto impessoal usa de 3ª pessoa (do ponto de vista morfológico) com valor de 2ª pessoa (do ponto de														

buscam explicações e previsões.																				
14. O vendedor para divulgação ou venda de algo ou conceito para o público os leitores.																				
15. Em jornais impressos e revistas, além de <i>websites</i> , os anúncios podem ser publicados junto ao conteúdo editorial ou em seções separadas, denominadas Classificados.																				
16. Algumas vezes não é identificado. Outras vezes é alguém especialista no assunto que escreve para jornais ou revistas ou públicos livros. Em todos os casos, é alguém que sabe fazer uma determinada comida e pretende ensinar como fazer.																				
17. O leitor para opinar sobre determinado assunto no jornal.																				
18. Dentro de um espaço sempre igual e com a mesma localização.																				

Quadro 3.5: Condição de Produção

Os quadros acima organizam as informações da seguinte maneira, cada número da coluna direita corresponde a um gênero do escopo, como descrito no parágrafo anterior, e a coluna da esquerda a um critério. Vejamos, portanto o Quadro 3.4 características da superfície lingüística, para o gênero manchete (1) sabemos que: *a omissão das palavras desnecessárias (adjuntos adnominais e adverbiais)* é uma característica da sua superfície lingüística, ou seja, se me propuser a escrevê-los ou estudá-lo, vai notar que é geralmente um texto curto sem muita descrição ou detalhes.

A fim de nos certificar da validade das informações coletadas no primeiro experimento, foi criado um banco de dados correspondentes aos índices encontrados. Por exemplo, Koch e Favero (1987) afirmam que o gênero notícia pertence ao mundo narrado, por isso apresenta verbos no presente do indicativo, pretérito perfeito composto, pretérito perfeito simples (retrospectiva), futuro do presente simples e composto, então buscamos dados do presente do indicativo possíveis de serem computados e listados e chegamos às DMT (desinências modo-temporais) típicas das flexões verbais dos verbos regulares.

É importante ainda salientar que as tabelas propostas a seguir de maneira alguma pretendem ser finitas e exclusivas, porém funcionam a partir do conceito clássico de modelagem computacional, que trata da aplicação de modelos matemáticos à análise, compreensão e estudo da fenomenologia de problemas complexos em áreas abrangentes, através da simulação de soluções para problemas científicos, analisando os fenômenos, desenvolvendo modelos matemáticos para sua descrição e elaborando códigos computacionais para obtenção dessas soluções.

A seguir, propomo-nos a aferir os conceitos de cada gênero utilizado nessa pesquisa, assim como a referência bibliográfica sobre sua composição, para a construção dos novos quadros com os critérios que foram utilizados na classificação automática desses gêneros, pois, assim como o escopo de gêneros estudados foi reduzido os critérios também foram restringidos em apenas um dos cinco parâmetros de Travaglia (2007), o parâmetro da superfície lingüística, que julgamos ser de maior

facilidade de automatização, visto que os dados levantados eram mais simples de serem computados, agilizando a construção do classificador. Senão seria necessário mais tempo e pesquisa, visto que cada dado de cada parâmetro levaria a equipe de pesquisa levantar metodologias diversas para sua computação.

3.1.6 Notícia

A notícia constitui-se na divulgação de um acontecimento por meios jornalísticos (Lima, 1990). É a matéria-prima do Jornalismo, normalmente reconhecida como algum dado ou evento socialmente relevante que merece publicação na mídia. Fatos políticos, sociais, econômicos, culturais, naturais e outros podem ser notícia se afetarem indivíduos ou grupos significativos para um determinado veículo de imprensa.

Um dos critérios associados ao gênero notícia são os marcadores de tempo; segundo Travaglia (2007^a, p. 24), “as narrativas nas notícias costumam ter indicações temporais mais precisas, com datas, por exemplo, como em *O presidente Lula, esteve em Cuba no dia 26/09/2003, onde estabeleceu acordos comerciais com aquele país...*”. Já os critérios de marcadores, preposições e conjunções de locuções prepositivas e conjuntivas de tempo foram explorados por nós apesar de não serem especificamente restritos ao gênero, já que o uso de conectores e de tipos de relações entre cláusulas foi revelado ligado ao tipo de texto (Travaglia 2007a p.65), encontramos no quadro 3 de Travaglia (2007a, p.109), o gênero notícia entre outros gêneros necessariamente compostos pela narrativa como tipo dominante, e segundo o mesmo, na narrativa os conectores são importantíssimos para marcar as relações temporais; assim, montamos uma lista finita, porém não exaustiva de alguns desses critérios.

Nossa tabela inicialmente contava com elementos linguísticos diversos, como por exemplo, critérios com valor temporal ou com implicações temporais, que podem marcar anterioridade, simultaneidade, posterioridade:

- p) elementos adverbiais de valor temporal: advérbios, sintagmas adverbiais, orações adverbiais;
- q) datas;
- r) preposições (após, antes de, depois de etc.);
- s) conjunções (enquanto, depois que, antes que, logo que etc.);
- t) verbos (iniciar, começar, terminar etc.) outros elementos ordenadores, tais como: primeiro, segundo, último, penúltimo, aí, daí etc.; como se pode verificar no quadro abaixo. O critério sobre predominância de 3^a pessoa do singular/plural e do tempo passado foi encontrado associado ao gênero notícia por Koch & Favero (1987), que afirmam que os tempos do mundo comentado indicam comprometimento, pois conduzem o ouvinte a uma atitude receptiva, tensa, atenta.

Koch e Fávero (1987) também afirmam que, relacionados à dimensão pragmática no tipo narrativo, destacam-se os enunciados de ação, tendo como atitude comunicativa o mundo narrado,

que se manifesta nas seguintes situações comunicativas: romances, contos, novelas, reportagens, noticiários, depoimentos, relatórios etc. Com referência à dimensão esquemática global, apresentam eventos em sucessão temporal e causal, existindo o antes e o depois. Há, ainda, os circunstancializadores e a presença do discurso relatado. Todavia, todos os estudos esclarecem que uma modalidade textual pode conter excertos de outras tipologias, isto é, um texto narrativo pode conter fragmentos descritivos, injuntivos, dissertativos etc.

Notícia	
Marcadores de tempo	Em (1-2011), no dia (1-31) de outubro de (1-2011), ontem, hoje, janeiro, fevereiro, março, abril, maio, junho, julho, agosto, setembro, outubro, novembro, dezembro.
Discurso Direto	““, -, disse:
Verbos irregulares.	
Conjunções e locuções conjuntivas de tempo	Quando, enquanto, logo que, assim que, depois/antes que
Preposições ou locuções prepositivas de tempo	Após, antes de, depois de,
Sequenciadores ou encadeadores de tempo	Aí, daí, então, e
Marcadores de tempo, advérbios e adjuntos adverbiais de tempo	Há muito tempo, à noite, em três dias, por muitos anos, dali a algum tempo, naquele momento, ontem, amanhã.
Terceira Pessoa	Ela, ele, elas, eles
Marcadores de tempo nomes (substantivos e adjetivos) indicadores de tempo	Dia, mês, semana, ano, década, atrasado, adiantado, temporário, transitório, semanal, mensal, anual, provisório etc.
Terceira pessoa	Ele, ela, eles, elas
Pretérito perfeito do indicativo (1ª conjugação)	-ei, -ou, -aram
Pretérito perfeito do indicativo (2ª conjugação)	-eu, -eram
Pretérito perfeito do indicativo (3ª conjugação)	-iu, iram
Presente do indicativo (1ª conjugação)	-a, -am,
Presente do indicativo (2ª conjugação)	-e, em
Presente do indicativo (3ª conjugação)	-i, -iem

Quadro 3. 6: *inputs* de notícia.

3.1.7 Artigo

Segundo Travaglia (2007a), o gênero artigo encontra-se entre as categorias de texto necessária e predominantemente dissertativas argumentativa (por exemplo: tese, dissertação de mestrado, artigo acadêmico-científico, editorial de jornal, monografia, conferência etc.) e assim por diante. Então procuramos as características da estrutura composicional e da superfície linguística

desse tipo de texto, porém o artigo trabalhado por nós nessa pesquisa é o artigo argumentativo encontrado no suporte de jornal e revistas, ou seja, diferente do artigo científico de publicação em seminários e congressos, que possuem uma estrutura rígida.

O artigo explorado nessa pesquisa é a exposição de algum assunto de interesse mútuo, seja através de jornais, revistas, televisão ou internet.

Nos gêneros textuais predominantemente dissertativos, aparecem todos os tempos verbais (categoria). Em Travaglia (1991), encontramos detalhadamente a proporção desses tempos verbais, com predominância do onitemporal (67,85%) ou do tempo não marcado (21,86%), seguidos do futuro (4,18%), do presente (3,21%) e do passado até o presente (1,61%) nesta ordem. A marcação de presente para o futuro não apareceu. Entende-se a predominância do onitemporal e do não marcado, tendo em vista as propriedades da dissertação de apresentar ideias vistas como válidas para todos os tempos, o conhecer abstraído do tempo. A marcação de passado até o presente parece ser característica do texto dissertativo. (Travaglia, 1991, Capítulo 5).

Já os outros critérios formam retirados das características tidas como de texto do mundo comentado (em que a perspectiva é de comprometimento do locutor/enunciador com o que diz): presente do indicativo, pretérito perfeito composto, pretérito perfeito simples (retrospectiva), futuro do presente simples e composto (Koch & Favero, 1987).

Artigo	
Enunciados de opinião	Eu acredito, eu acho, Nós entendemos que, É certo, Afirma-se, Entende-se, Entendo, É fato.
Adjetivos elucidativos	necessário, obrigatório, proibido, caro, raro, mortal etc.
Preposições ou locuções prepositivas de tempo	após, antes de, depois de
Advérbios de intensidade, dúvida, negação, modo e afirmação.	Muito, pouco, bastante, mais, menos, demais, tão, tanto, meio - Talvez, provavelmente, possivelmente - Não, Sim, certamente, realmente, efetivamente - Bem, mal, assim, depressa, calmamente, suavemente, alegremente.
Expressões de relevância	Conforme, segundo, de acordo com, como dizia, citando.
Conectores diversos	Mais que, tão, quanto, menos, por exemplo, assim, contudo, com isso, por exemplo, porque, devido a, embora, por causa, no qual, na qual, pois, porque, portanto etc.
Terceira pessoa	Ele, ela, eles, elas
Pretérito perfeito do indicativo (1ª conjugação)	-ei -ou, - aram
Pretérito perfeito do indicativo (2ª conjugação)	-eu, - eram
Pretérito perfeito do indicativo (3ª conjugação)	-iu, iram

Presente do indicativo (1ª conjugação)	-a, - am,
Presente do indicativo (2ª conjugação)	-e, em
Presente do indicativo (3ª conjugação)	-i, - iem
Futuro do indicativo (1ª conjugação)	Ará, - Arão
Futuro do indicativo (2ª conjugação)	-erá, erão
Futuro o indicativo (3ª conjugação)	-irá, - irão

Quadro 3. 7: *inputs* de artigo.

3.1.8 Receita

As receitas são um gênero predominantemente injuntivo com as seguintes características: presença de auxiliares modais de modalidades imperativas, sobretudo ordem, obrigação e prescrição; são constituídas essencialmente de verbos dinâmicos (ações); verbos enunciativos mais no discurso indireto e ligados à condição do produtor do texto de incitador e do recebedor de potencial executor das ações: mandar, ordenar, determinar, pedir, suplicar, sugerir, recomendar etc., não atualização do aspecto; presença de modalidades imperativas (obrigação, permissão, ordem, proibição, prescrição) e com a volição e que são características dos textos injuntivos; o tempo característico é o futuro, independentemente da forma verbal (cf.Travaglia, 1991).

Segundo Travaglia (1991, p.237, 239; 1992), a superestrutura de um texto do tipo injuntivo é constituída de elenco ou descrição, em que se apresentam os elementos a serem manipulados na ação a ser feita. Pode-se dar apenas uma lista desses elementos (V. ingredientes das receitas culinárias) ou pode-se listá-los e descrevê-los, como nos manuais de instrução em que, comumente, a descrição é substituída por fotos ou desenhos com indicação dos nomes das partes seguida ou não de indicação de sua função

Receita	
Presença de um elenco = ingredientes de comida	Ovo, açúcar, sal, óleo, manteiga, margarina, farinha de trigo, farinha de mandioca, azeite, cebola, alho, água, óleo.
Auxiliares modais de modalidades imperativas de ordem, obrigação e prescrição.	Dever, poder, ter de, haver de, hás de, tens de, tem que, deve, pode
Verbos dinâmicos (ações)	Corte, aqueça, doure, derreta, esfrie, misture, bata, jogue, tempere, desfie, dilua, pique, mecha, resfrie, frite, bata, esquite, refogue, unte, pegue, asse, descongele, derreta, cozinhe, ferva.
Conectores de sequenciamento de ações	Primeiro, depois, por último, depois disso, agora, daí, após, antes, segundo, terceiro.
Futuro do presente do indicativo	-ei, - ás, á, - remos, - reis, -rão
Conectores de justificativa	Porque, para que, para, assim, devido a, visto que, pois, desta forma, ora, desse modo.
Presente do indicativo (1ª conjugação)	-a, - am,
Presente do indicativo (2ª conjugação)	-e, em

Presente do indicativo (3ª conjugação)	-i, - iem
--	-----------

Quadro 3. 7: *inputs* de receita.

3.1.9 Horóscopo

O horóscopo é considerado um gênero textual que apresenta uma forma estrutural específica, com a predominância do tipo lógica preditiva e o propósito comunicativo de prever eventos cíclicos. Apresenta um conteúdo e meio de transmissão específico, como e-mail e jornais. Contudo, os critérios para captura dos dados do gênero horóscopo foram retirados das características dos gêneros preditivos (como programas de viagem ou outros programas, boletins meteorológicos e astronômicos, profecias etc.) sempre serão compostos necessariamente pelo tipo preditivo geralmente em fusão com o descritivo ou o narrativo e às vezes o dissertativo. Todavia, no caso do gênero horóscopo, encontramos citado como possuindo uma das características de estrutura composicional do texto injuntivo, assim como a receita. O horóscopo foi citado em Travaglia (1991, p.237, 239, 1992), como possuidor da superestrutura de um texto do tipo injuntivo por possuir elenco, justificativa, explicação ou incentivo, em que se dá razões para a realização das situações especificadas em *Carneiro/pessoa*: “A amizade exige às vezes *discrição e sacrifícios*.” Estas partes não têm ordem fixa e podem se intercalar. A única parte obrigatória é a determinação, mas às vezes o produtor do texto apenas dá à justificativa ou explicação, e a determinação fica implícita, sendo dedutível através de inferências. Isto é comum em horóscopos, como em *Câncer/saúde*: “A *dieta da Lua é especialmente recomendada para as cancerianas*” (Horóscopo da revista Elle. Ano 2 nº 10. São Paulo: Ed. Abril, outubro 1989:209).

Horóscopo	
Verbos preditivos	Serão, sentir-se-á, agirá, Terá, terão, poderá, poderão, trará, trarão, sentirão, sentirá, irá, irão, Haverá,vai, vão
Marcadores de tempo	Em (1-2011), no dia (1-31) de outubro de (1-2011), ontem, hoje, janeiro, fevereiro, março, abril, maio, junho, julho, agosto, setembro, outubro, novembro, dezembro
Presença de palavras-chaves	Ascendente, Vênus, Marte, Netuno, Saturno, Júpiter, Plutão, Mercúrio, Urânio, Sol, Lua, Escorpião, Peixes, Aquário, Áries, Touro, Gêmeos, Câncer, Leão, Libra, Virgem, Sagitário, água, fogo, amor, relacionamento, trabalho, família, projetos, Capricórnio, Balança
Auxiliares modais de modalidades imperativas de ordem	Dever, poder, ter de, haver de, tem, pode, deve, podem, devem, têm
Terceira pessoa	Ele, ela, eles, elas, você
Futuro do presente do indicativo (1ª 2ª e 3ª conjugação)	-ei, - ás, á, - remos, - reis, -rão

Quadro 3. 1: *inputs* de horóscopo.

3.1.10 Crônica

A crônica é um gênero textual com predominância da tipologia narrativa, (cf.Travaglia 2007). As características da tabela foram interpretações a partir da lista de características da superfície linguística do tipo narrativo, que apresenta verbos gramaticais predominantes, marcadores temporais e auxiliares aspectuais, o que são coerentes com a propriedade dada pela perspectiva de inserção no tempo e também os auxiliares semânticos (que dão detalhes ou *nuances* dos fatos narrados); são constituídos essencialmente por verbos dinâmicos (ações, fatos, fenômenos, transformativos); aparecem verbos enunciativos de contar e assistir, já que o produtor é o contador e o receptor é o assistente dos episódios: *presenciar, assistir, ver (tudo/o que acontecer/suceder/ocorrer), contar, relatar, narrar, falar/dizer (tudo/o que acontecer/ suceder/ocorrer)*; só são possíveis com o aspecto perfectivo que caracteriza a narração.

Dos aspectos de duração, os mais característicos da narração são o durativo, o iterativo e o pontual; as modalidades características desse tipo de texto são a certeza e a probabilidade, uma vez que são os textos que dão a conhecer os acontecimentos; também para a narração o tempo atualizado depende da relação entre o tempo referencial e o da enunciação: a) presente na narração presente; passado na narração passada. O presente aparece com função de relevo emocional; futuro nas narrações futuras. (Travaglia 1991, Capítulo 5)

Crônica	
Marcadores de tempo	Em (1-2011), no dia (1-31) de outubro de (1-2011), ontem, hoje, janeiro, fevereiro, março, abril, maio, junho, julho, agosto, setembro, outubro, novembro, dezembro, atualmente, hoje em dia, esses dias, essa semana, nos fins de semana, aos sábados, segunda-feira, terça-feira, quarta-feira, quinta-feira, sexta-feira, domingos, todo dia, diariamente, atualmente, sempre, nunca, às vezes
Auxiliares aspectuais	Continuar, começar, estar a, acabar de, andar
Verbos dinâmicos	Acordar, tomar café, caminhar, ir, vir, voltar, pensar, correr, sonhar, limpar, falar, rezar, viajar, namorar, beijar, morrer, viver, cortar, cozinhar etc.
Pretérito perfeito do indicativo (1ª conjugação)	-ei -ou,-amos, -endo - aram
Auxiliares semânticos	Leu/está a ler/vinha lendo/tem de ler/começou a ler/acabou de ler/deixou de ler
Enunciativos de contar e assistir	Presenciar, assistir, ver (tudo/o que acontecer/suceder/ocorrer), contar, relatar, narrar, falar/dizer (tudo/o que acontecer/ suceder/ocorrer)

Aspecto perfectivo (durativo, o iterativo e o pontual)	Tenho, tenha, estou, está
Modalidades de certeza e probabilidade	Dever, poder, ter de, haver de, tem que, deve, pode
Presente do indicativo (1ª conjugação)	-a, -am,
Presente do indicativo (2ª conjugação)	-e, em
Presente do indicativo (3ª conjugação)	-i, -iem

Quadro 3. 2: *inputs* de crônica.

3.1.11 Manchete

Esse gênero em português se aproxima muito das características da superfície linguística do gênero notícia, porém sua diferença é óbvia e se resolve pelo aspecto de estrutura composicional que é a dimensão: o tamanho médio dos textos daquele gênero. Embora nunca se possa estabelecer e nunca se estabeleça um tamanho exato para um gênero, há um padrão esperado de dimensão.

Assim, por exemplo, uma epopeia (como “Os Lusíadas”, de Camões) é sempre muito maior do que um “poema heroico” (como o “Uruguai”, de José Basílio da Gama). Do mesmo modo se espera que um romance seja muito maior do que um conto. São também narrativas curtas os apólogos, fábulas e parábolas que, geralmente, são menores do que um conto. Nas programações de cinema, os textos do gênero resumo de filme são bem curtos, do mesmo modo que os classificados.

Certamente este critério é problemático, mas também não se pode deixar de reconhecer que a dimensão do texto de dado gênero é caracterizadora do mesmo. Basta pensar que ninguém imaginará ser uma manchete uma notícia.

Manchete	
Marcadores de tempo	Em (1-2011), no dia (1-31) de outubro de (1-2011), ontem, hoje, janeiro, fevereiro, março, abril, maio, junho, julho, agosto, setembro, outubro, novembro, dezembro.
Conjunções e locuções conjuntivas de tempo	Quando, enquanto, logo que, assim que, depois/antes que
Preposições ou locuções prepositivas de tempo	Após, antes de, depois de
Auxiliares modais	Pode, deve, tem poder, dever, ter, pudesse, tivesse, devesse
Sequenciadores ou encadeadores de tempo	Aí, daí, então
Marcadores de tempo, advérbios e adjuntos adverbiais de tempo	Há muito tempo, à noite, em três dias, por muitos anos, dali a algum tempo, naquele momento.
Tamanho do texto (número de palavras)	<i>“deve ser menor que a notícia”</i>
Terceira Pessoa	Ele, ela, eles, elas
Marcadores de tempo, nomes (substantivos e adjetivos) indicadores de	Dia, mês, semana, ano, década, atrasado, adiantado, temporário, transitório, semanal,

tempo	mensal, anual, provisório etc.
Pretérito perfeito do indicativo (1ª conjugação)	-ei -ou, - aram
Pretérito perfeito do indicativo (2ª conjugação)	-eu, - eram
Pretérito perfeito do indicativo (3ª conjugação)	-iu, iram
Presente do indicativo (1ª conjugação)	-a, - am,
Presente do indicativo (2ª conjugação)	-e, em
Presente do indicativo (3ª conjugação)	-i, - iem

Quadro 3. 3: *inputs* de manchete.

3.1.12 Resenha

Esse gênero textual foi bem explorado por Motta Roth (1998), porém em seu aspecto acadêmico e argumentativo. Nesse trabalho, exploraremos o universo de resenhas críticas sobre cinema, teatro, música do universo de crianças produtoras de texto, assim achamos também importante criar um elenco como característico da estrutura composicional do gênero em questão, palavras-chaves que devem ser mencionadas num texto que possui como objetivo e função sociocomunicativa, descrever, apontar pontos positivos e negativos e recomendar ou não. Assim, acreditamos seja um gênero predominantemente do tipo descritivo, que se faz sobretudo com verbos dinâmicos.

Resenha crítica	
Elenco	Ator, atriz, trilha sonora, cenário, diretor, roteiro, música, história, artistas, elenco, dublês, efeitos especiais, <i>script</i> , herói, mocinho, vilão, vilã, mocinha, palco, luz, indumentária, figurino, horário, preço, dublagem, personagem, cenário, cena, <i>performance</i> , figurante, maquiagem, relevância, Oscar, prêmio etc.
Presença de descrição	Bonito, legal, interessante, rico, pobre, de qualidade, sensacional, belo, chato, horrível, ruim, longo, curto, retrô, clássico, moderno, barroco, renascentista, original etc.
Presente do indicativo (1ª conjugação)	-a, - am,
Presente do indicativo (2ª conjugação)	-e, em
Presente do indicativo (3ª conjugação)	-i, - iem
Pretérito perfeito do indicativo (1ª conjugação)	-ei -ou, - aram
Pretérito perfeito do indicativo (2ª conjugação)	-eu, - eram
Pretérito perfeito do indicativo (3ª conjugação)	-iu, iram

Futuro do presente do indicativo	-ei, -ás, á, -remos, -reis, -rão
Auxiliares modais	Dever, poder, ter de, haver de, hás de, tens de, tem que, deve, pode
Ordenadores textuais	Em resumo, até aqui, como já foi falado, em primeiro lugar, em segundo lugar, sinceramente, cá entre nós,
Verbos enunciativos de pensar	Achar, saber, parecer, acreditar, julgar, supor, adivinhar, imaginar

Quadro3. 4: *inputs* de resenha crítica.

3.1.13 Implementação

A fim de testar a relevância dos critérios tabelados acima, montamos, a partir de expressões regulares Quadro 3.8, uma busca por esses critérios no banco de textos de aproximadamente 5.000 textos já previamente classificados entre os gêneros. A caráter científico, procuramos e numeramos a ocorrência de todos os critérios em todos os tipos de texto e chegamos à validação da relevância dos critérios selecionados para cada gênero a partir da frequência com que apareciam no texto; com isso, foram eliminados os critérios com baixa frequência e para aqueles de alta frequência foi adicionado um valor, que é o número de vezes que ele é encontrado multiplicado por 1000 para que esse pese como característica marcante do gênero em questão. Com isso, muitos critérios foram desativados e cada gênero obteve uma média de cinco critérios significativos.

Gênero	Qt.	Critério
Manchete	03	"prepdetempo_manchete" "presente_do_indicativo" "auxiliares_manchete"
Receita	02	"ingredientes_receita" "verbos_dinâmicos_receita"
Resenha	05	"Presença_de_descrição" "Pretérito_perfeito_do_indicativo3_notícia" "verbos_irregulares_notícia" "verbos enunciativos de pensar" "léxico"
Notícia	10	"locuções_de_tempo_notícia" "prepdetempo_notícia" "Advérbios_notícia" "Pretérito_perfeito_do_indicativo1" "Pretérito_perfeito_do_indicativo2" "Pretérito_perfeito_do_indicativo3" "verbos_irregulares_notícia" "Marcadores_de_tempo_nome_notícia" "Terceira_pessoa" "discurso_direito"
Artigo	05	"Adjetivos_elucidativos" "Advérbios_artigo" "Expressões_de_relevância_artigo" "Conectores_diversos_artigo" "enunciados_de_opinião"

Horóscopo	05	"verbos_preditivos_horóscopo" "palavras_chave_horóscopo" "futuro_do_presente1_artigo" "Terceira_pessoa_horóscopo" "marcadores_de_tempo"
Crônica	06	"verbos_dinâmicos_crônica" "auxiliares_aspectuais_crônica" "enunciativos_contar_crônica" Aspecto_perfetivo "modalidade_certeza_crônica" "primeira_pessoa"

Quadro 3. 5: critério dos gêneros

O nosso desafio foi encontrar a melhor técnica de implementação dessas perguntas que auxiliasse o professor, fazendo com que a maior parte do processo de intervenção dialogada, fundamental da técnica psicopedagógica da elaboração dirigida, fosse feito somente da interação aluno e máquina.

Portanto, optamos em fazer um *Wizard*, pois decidimos que a intervenção deveria acontecer em formato de perguntas associadas a cada critério, para que assim a característica de diálogo fosse mantida, no formato pergunta e resposta, como mostra o Quadro 3.9. Todavia, essas perguntas se propõem a acionar regras já conhecidas pelos alunos, ou quando novas, a levar à reflexão da superfície linguística desses gêneros a fim de construir essas novas regras que ajudam a classificá-los, por isso devem não somente testar o conteúdo, mas levar a uma reflexão sobre ele.

Como as perguntas da elaboração dirigida estão associadas aos critérios, essas são acionadas quando os critérios não são encontrados no texto dos alunos. Assim, os alunos, têm a chance de revisar seus textos e alterá-los repetidas vezes até serem aprovados. O número de tentativas para revisão dos textos é limitada, três tentativas por critério, segundo a premissa de que a possibilidade de tentativas infinitas poderia desmotivar os alunos e escamotear a eficiência das perguntas de elaboração dirigida.

Gênero	Qt.	Perguntas de Mediação
Manchete	03	<i>Como você ligou os acontecimentos? Sua notícia está numa ordem cronológica? Como que eu sei que uma coisa aconteceu depois da outra? Você colocou isso no seu texto? "Sua manchete chama atenção do leitor para algo que está acontecendo? Como você faz isso? Usa o tempo presente?" "Como você indica certeza, obrigação ou dúvida das ações em seu texto?</i>
Receita	02	<i>Você escreveu sobre como fazer uma comida? Ou uma sobremesa? Quais ingredientes você colocou em sua receita? Explicou como prepará-los? Especificou as quantidades?" Em suas orientações você especificou o que eu devo fazer com os ingredientes? Como, por exemplo, como cortar ou mexer a massa?</i>
Resenha	05	<i>Você descreve em seu texto os detalhes do objeto analisado, dando sua opinião? Utiliza adjetivos para fazer isso?" Sua resenha analisa uma obra de arte que está acontecendo, que já aconteceu ou que ainda irá acontecer? Como você faz isso? Usa o tempo presente, passado ou futuro? Como você indica em seu texto sua ideias, opiniões etc.? Em seu texto você menciona as áreas técnicas que avalia? Separando o projeto final da obra nas partes que a compõe, como, por exemplo, o elenco e a direção de um filme, ou a trilha sonora, o cenário, etc?</i>

Notícia	10	Como você ligou os acontecimentos? Sua notícia está numa ordem cronológica? Como que eu sei que uma coisa aconteceu depois da outra? Você colocou isso no seu texto?" "Como você ligou os acontecimentos? Como indica intensidade da certeza ou incerteza de suas afirmações? Como indica a dúvida quando há? Como afirma as incertezas, as negações? A maneira como afirma ou fazem as coisas? "Sua notícia conta algo que está acontecendo? Como você faz isso? Usa o tempo presente? "existe ação em seu texto? Utiliza verbos regulares e irregulares para descrevê-las? Será que não ficou faltando? Existe em seu texto referências a quando sua notícia aconteceu?" Você foi o protagonista de sua notícia? Você estava lá e conta para gente o que você viu, ou outra pessoa te contou e você está contando para gente? Como essas outras pessoas aparecem na sua notícia? Como você reporta o que foi dito por outras pessoas?
Artigo	05	Como você classifica o conteúdo das informações que coloca em seu texto? Usa bastante adjetivos? Como indica intensidade da certeza ou incerteza de suas afirmações? Como indica a dúvida quando há? Como afirma as incertezas, as negações? A maneira como afirma ou fazem as coisas? Você em algum momento deu alguma informação que na verdade não foi você quem afirmou ou disse primeiro? Você por algum acaso mencionou em seu texto quem é a pessoa ou a fonte de onde pegou essa informação, e quando? "Você utiliza palavrinhas para unir as suas ideias, com objetivo de contrastar, explicar, concluir e adicionar informações em seu artigo? Como você indica que o que está afirmando é uma opinião, e não uma certeza
Horóscopo	05	eu horóscopo é uma previsão de como será o dia, a semana ou o mês da pessoa daquele signo e conta sobre o que ainda vai acontecer?" Você menciona os elementos e ascendentes e planetas, amor, relacionamento, saúde, trabalho e projetos relacionados à sua previsão? Como você dá a sua previsão, e descreve o que ainda irá acontecer você usa o tempo presente? Como você indica em seu texto que está se referindo a terceiros, e não a si mesmo? "Você acha importante colocar uma data para a sua previsão, como uma semana, um dia específico ou um mês?
Crônica	06	Em sua crônica você especificou as ações de rotina feitas pelo protagonista?" "Como você indica as ações em sua crônica? ="Você utiliza verbos que especificam estar contando ou assistindo aos fatos em sua crônica? Isto está claro?" Como você indica certeza e probabilidade das ações em sua crônica? ="Você é o protagonista de sua história? Você narra o seu cotidiano ou de outra pessoa? Como essa outra pessoa aparece no seu texto? Como eu sei que ela não é você?"

Quadro 3. 6: Perguntas de mediação para cada critério dos gêneros

4 METODOLOGIA

Nesse Capítulo é apresentada a metodologia de pesquisa realizada no âmbito da Mineração de Texto descrito no capítulo anterior.

O desenvolvimento desse sistema tem como objetivo atuar na etapa de pós-processamento de documentos e será alimentado da seguinte forma: Obtenção de conteúdos colhidos referentes aos gêneros, os quais serão armazenados em uma base de dados central, e assim, será criada uma interface de acesso que permitirá ao usuário escrever textos referentes aos gêneros do jornal.

Na etapa de processamento, será utilizado o *software* Orange, para dar suporte às tarefas de classificação e categorização entre os documentos. Os resultados da etapa de processamento serão analisados através dos módulos de visualização do *software*. A metodologia é apresentada em cinco etapas, conforme a Figura 4.1. A etapa identificada pelo número I requer envolvimento direto do usuário, a de número II são atividades desenvolvidas no escopo deste trabalho e a de número III utiliza *software*.

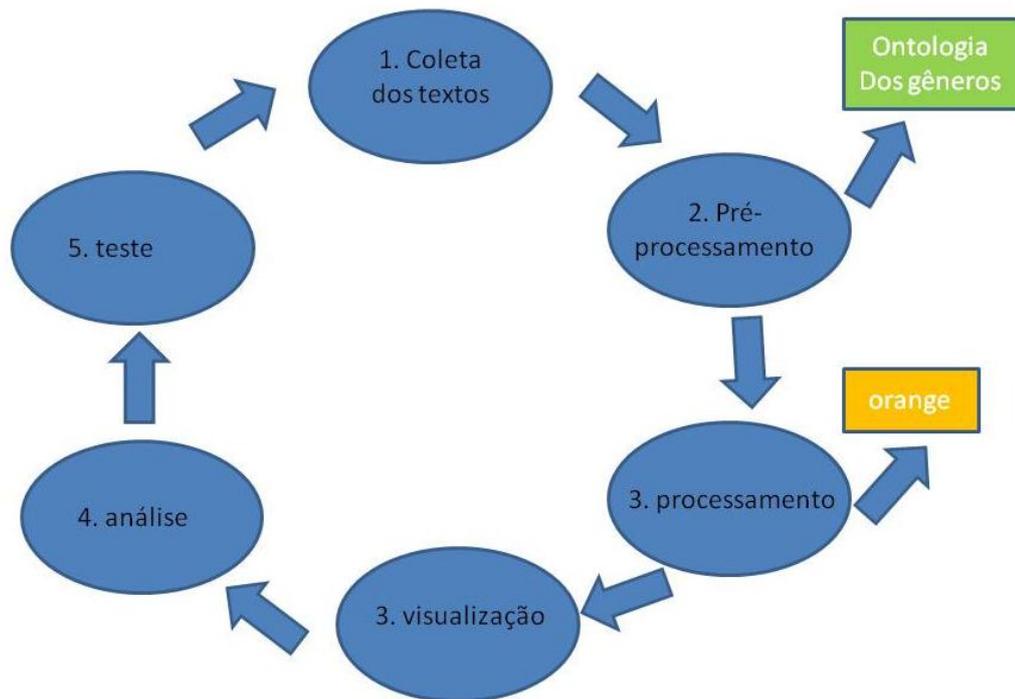


Figura 11. 1: Etapas da Metodologia proposta.

Cada etapa tem suas atividades definidas e é realizada sequencialmente, mas a retroação entre as etapas é considerada sempre que for necessário melhorar os resultados finais.

A busca de documentos relevantes ao domínio de aplicação do conhecimento a ser extraído que são sete gêneros textuais foram retirados da internet, de *sites* específicos que divulgam textos dos gêneros escolhidos para o desenvolvimento dessa pesquisa. Os textos coletados foram

divulgados entre os anos de 2009 e 2011. O primeiro passo foi converter textos em *http* para o formato de texto xml e, em seguida, inseri-los textos no banco de dados utilizado na aplicação. Será apresentado um estudo de caso neste trabalho. Inicialmente foram selecionados aproximadamente 10.0000 textos de pré-classificados a partir de seus *sítes* de origem, nos seguintes gêneros: notícia, manchete, horóscopo, artigo, resenha, crônica e receita. O problema desta abordagem está na perda de formatação da fonte original, pois caracteres costumam vir juntos com as letras e números, muitas vezes quebrando palavras, porém não tivemos muitos problemas e com a quantidade de textos armazenados decidimos não atentar para a correção do texto, visto que a interferência dos caracteres seria insignificante. Então, para poder trabalhar com os textos, montou-se um sistema de cadastro de gêneros junto a uma estrutura de banco de dados com o intuito de organizá-los textos pelos gêneros correspondentes.

4.1 Pré-Processamento

A segunda etapa após a coleta de documentos de Mineração de Texto é denominada de pré-processamento. Os documentos selecionados são formatados, pois eles serão submetidos aos algoritmos de extração automática de conhecimento, quando se tem como objetivo a extração de padrões. Nessa fase, compreende-se a identificação de todos os termos existentes nos documentos, bem como o armazenamento de informações referentes aos termos que serão utilizados na fase seguinte, pelas ferramentas de descoberta de conhecimento. O objetivo desta fase é identificar cada termo existente no documento. Sendo assim, são executadas: a análise léxica, efetuada por um *parser* que identifica cada termo como sendo uma sequência de letras e/ou dígitos; a análise morfológica, efetuada por um *stemmer*, que reduz cada termo ao seu radical; e a retirada de *stopwords* ou termos que não agregam valor, tais como: preposições, conjunções, artigos etc.

Contudo, para nossa pesquisa não são interessantes nem justificáveis as duas últimas tarefas anteriores: a redução de cada termo ao seu radical e a retirada de *stopwords* ou termos como: preposições, conjunções, artigos etc. Seria como seguir no sentido contrário da proposta dessa pesquisa, visto que o conteúdo pesquisado aqui, da diversidade textual, engloba todos os organizadores textuais como suas partes mais significativas, responsáveis pela distinção da composição, estruturação, objetivo e conteúdo dos gêneros.

4.2 Processamento dos dados

A configuração do ambiente é apresentada no Capítulo seguinte. A listagem de critérios levantados teoricamente sobre cada um dos gêneros textuais foi utilizada e testes foram gerados para verificar a veracidade das informações pesquisadas. Surpreendentemente, a frequência de muitas características foi muito baixa e insignificante para auxiliar na classificação dos gêneros. Muitas outras características não se mostraram exclusivas para cada gênero. Assim, a listagem de

critérios/características feitas somente com base em pesquisa teórica e na literatura foi reduzida a uma média de quatro critérios por gênero, critérios esses que só permaneceram por serem frequentes e constantes em determinados gêneros, porém ainda estes tiveram que ser multiplicados para que ficassem ainda mais evidentes a cada gênero.

4.3 Extração dos dados

Nesta etapa, buscou-se conhecer melhor a base de dados através de análise estatística dos dados e a distribuição dos registros na base. Como primeira análise, procurou-se entender a distribuição das características linguísticas por gêneros. Em seguida, foram encontrados os termos com maior frequência nos textos. Analisou-se também a frequência das frases mais significantes. Esta técnica foi adotada para encontrar características repetidas e exclusivas em diferentes gêneros.

4.4 Categorização

A primeira tarefa realizada após a exploração dos dados foi categorizar os documentos. O objetivo maior desta etapa é encontrar nos textos do banco de dados características dos gêneros que possam estar de acordo ou não com a listagem de critérios teóricos. Além disso, buscaram-se avaliar os melhores critérios. A classificação auxiliou o especialista do domínio a categorizar automaticamente os gêneros dos textos, Os textos aparecem pré-classificados, no jogo, o que torna o trabalho do especialista mais simples. A classificação foi realizada em algoritmos disponíveis no sistema Orange e o bayesiano.

4.5 Pós-processamento

Essa etapa envolve a apresentação, a análise e a interpretação dos resultados, a fim de validar as descobertas obtidas na etapa anterior. Nela, o especialista em mineração de dados e o especialista no domínio da aplicação podem, a partir da avaliação dos resultados alcançados, definir novas alternativas de investigação dos dados. Nesta etapa, buscam-se responder algumas perguntas iniciais sobre a base de dados, que motivaram o desenvolvimento deste trabalho: Qual a probabilidade de um critério se repetir em outros gêneros? Quais gêneros apresentam critérios mais parecidos?

A proposta de utilização da metodologia descrita foi crucial para o desenvolvimento dessa pesquisa, pois inicialmente foram experimentados métodos muito ingênuos para o objetivo desse trabalho. A seguir, será apresentado todo o ambiente computacional utilizado.

5 AMBIENTE COMPUTACIONAL

Neste capítulo é apresentado o ambiente computacional produzido, a arquitetura geral do sistema e seus módulos de execução, as ferramentas utilizadas, a modelagem do sistema e a visão geral do protótipo.

5.1 Arquitetura Geral do Sistema

Foi implementado um sistema de investigação textual contemplando a etapa de preparação dos dados descrita no Capítulo 3. O ambiente de desenvolvimento do sistema é composto pelo sistema operacional Linux, pelo ambiente de programação *Komodo tools*, pela linguagem de programação python com *xml* e pelo banco de dados *Apache couchdb*.

O sistema foi desenvolvido de forma clássica, isto é, com acesso às funcionalidades principais através de uma barra de menus que levam a formulários de preenchimentos, consultas, ações e visualização de resultados.

Para um melhor desenvolvimento e evolução mais fácil, foi criada uma arquitetura modular. Uma aplicação parcelada em 4 módulos: TextSeeker, dbConnector, Tipos e TextGrabber como mostra a Figura 5.1, ou Anexo 1.

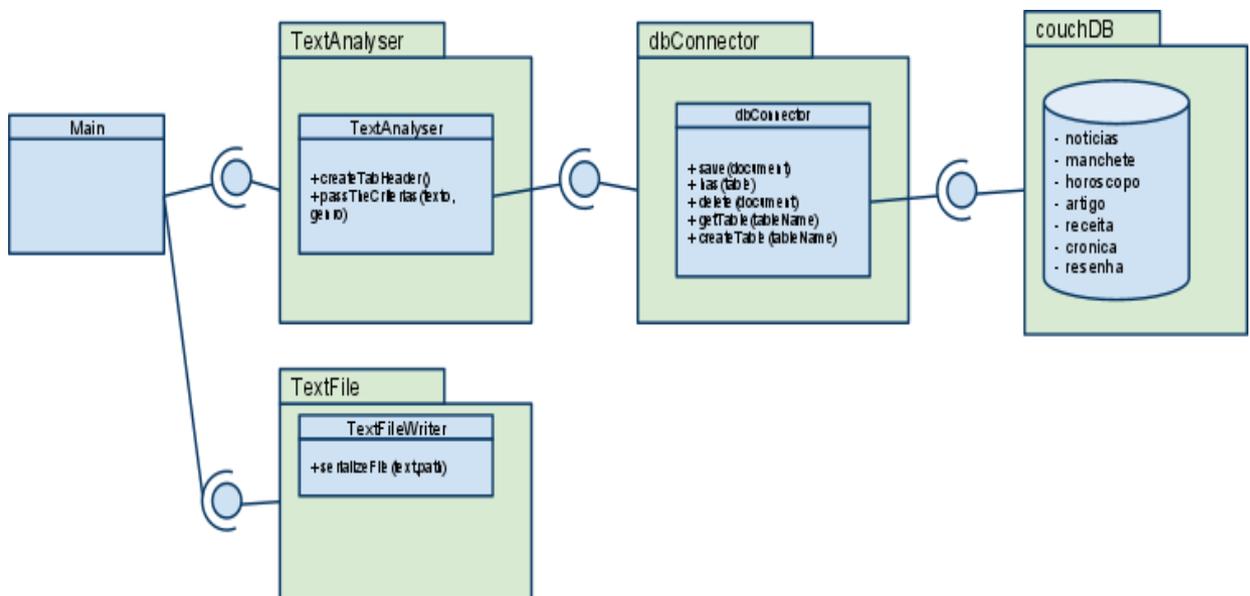


Figura 5. 12: Textanalyser: Detalhe do processo de extração de texto.

Classe principal ou *Main*

A classe principal é a classe *executável* deste aplicativo, que faz a ligação dos diferentes módulos. O usuário tem a possibilidade de executar essa classe como um *daemon*, e executar a atualização das tabelas, todos os dias, por si só, ou o usuário pode executá-lo uma vez. Para correr como *deamon*, o usuário tem que adicionar o parâmetro de *boot* quando executa o *software*.

dbConnector

DbConnector é um invólucro que permite a comunicação e a gestão da base de dados. Aqui usamos um banco de dados couchDB. Um *wrapper* permite ao utilizador mudar facilmente o banco de dados sem perturbar os outros módulos.

TextFile

Esse módulo permite a seriação das tabelas de exemplo num formato de biblioteca Orange.

A classe de critérios XMLFileReader permite a leitura e a extração dos dados do arquivo xml, descrevendo as expressões regulares de alguns critérios. A classe TextFileWriter pega o texto que vem do módulo TextAnalyser e salva num formato de arquivo de texto, cujo caminho foi especificado em parametros.

Analizador de Texto ou *TextAnalyser*

O *software* TextAnalyser irá criar o arquivo de exemplo, as tabelas que serão posteriormente utilizadas para o *software* TextTrainer. O arquivo será criado em um formato que, a biblioteca de mineração de dados, Orange pode compreender e tomar como entrada.

O arquivo foi dividido em duas partes: o cabeçalho estático e os exemplos. O método createTabHeader () cria este cabeçalho em função dos arquivos de critérios e outros métodos. O outro método "*passTheCriterias*" analisa e calcula os atributos diferentes para cada texto, para o método Bayesiano. Os dados provêm de diferentes textos, guardados na base de dados. Cada texto é conhecido com a sua classe (notícia, manchete, horóscopo, etc.), que pode ajudar a criar a tabela de exemplo bayesiana. O teste de critérios refere-se a métodos de processamento de linguagem natural, tais como expressões regulares, formatos de texto, etc.

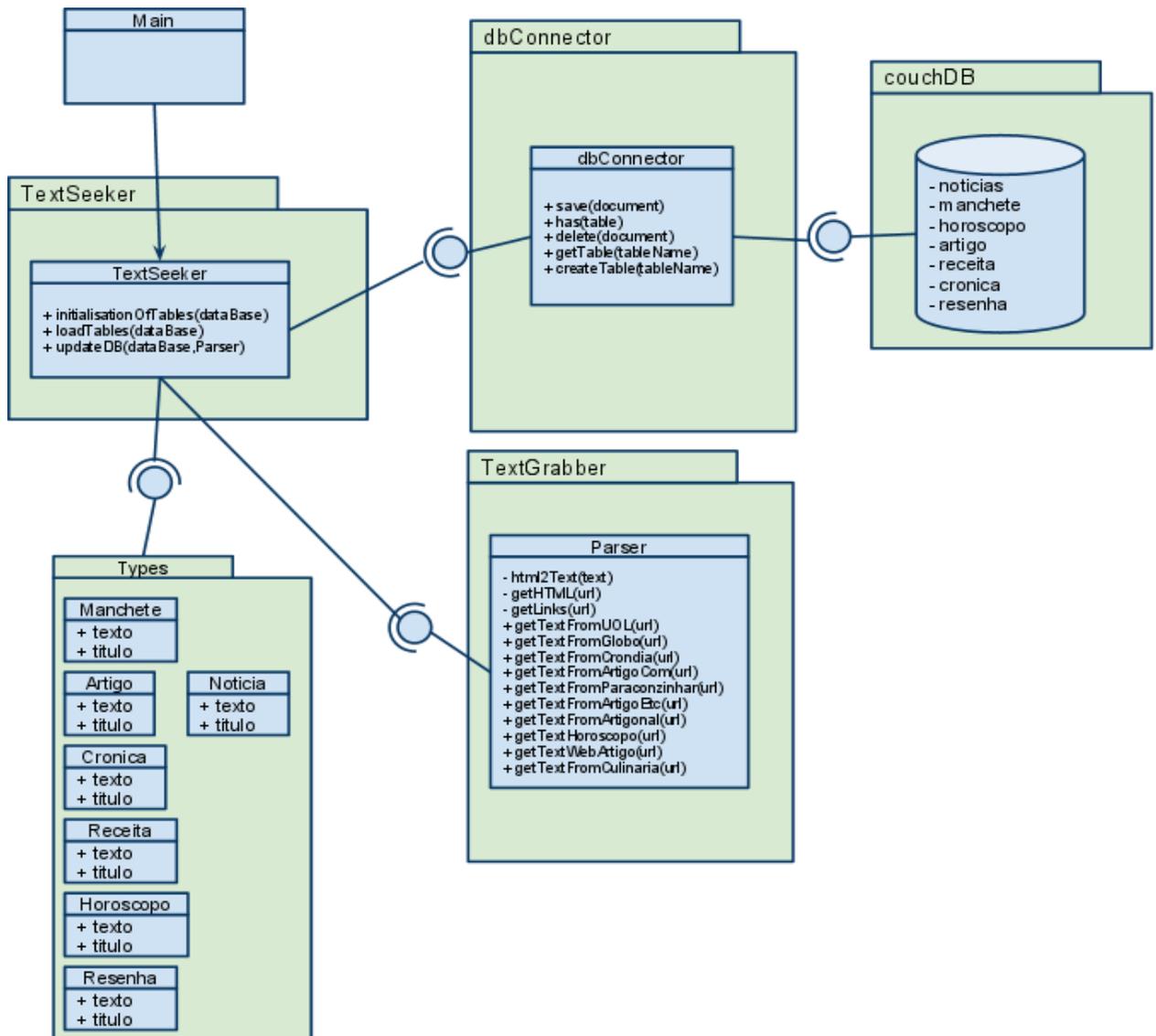


Figura 5. 13: arquitetura do TextSeeker .

TextGrabber

Este módulo oferece todas as funções para analisar os diferentes *websites*, que contém cada um dos gêneros. Este módulo contém todas as definições de classes (interfaces) dos dados guardados no banco de dados. Cada tipo tem os atributos de texto e título.

TextSeeker

Assim é chamada a classe principal: Ele irá inicializar as tabelas, o que significa criar as tabelas se elas não existem, e colocá-las na memória do sistema. Então TextSeeker tem uma função

para atualizar o banco de dados com novos textos provenientes de diferentes *feeds* RSS, chamando o analisador e, em seguida, o conector de base de dados.

TextFileWriter

Este módulo permite a serialização do arquivo de exemplo da tabela em um formato de biblioteca Orange. A classe *criteriaXMLFileReader* permite ler e extrair os dados do arquivo XML que descreve as expressões regulares a partir de alguns dos critérios. O *TextFileWriter* irá obter o texto vindo do módulo *TextAnalyser* e salvá-lo em um arquivo de texto, que o caminho tenha sido especificado no parâmetro.

Software "TextTrainer"

Esta aplicação é o cliente final. Ele irá analisar um texto de entrada e identificar o gênero do texto. (Manchete, notícia, Artigo, resenha ou horóscopo). Mais, *TextTrainer* pode dar a lista de atributos de um texto de entrada que completará os critérios de um gênero textual específico.

O ambiente computacional configurado para a utilização do *software* produzido para essa dissertação envolve ferramentas de código aberto, de maneira a permitir futuras modificações e evoluções que se fizerem necessárias.

5.2 Modelagem do Sistema

Ao observar as ações dos outros, o indivíduo sem dúvida considera para si características do modelo, as suas próprias características e se compara com ele, atentando ainda para as consequências positivas ou negativas daquelas ações, e poderá vir a adotar aquelas condutas como referências para si. No caso da televisão, o telespectador é um observador exposto a padrões diversos de conduta que poderá vir a adotar por meio do mesmo mecanismo de identificação na observação de modelos. No caso aqui, esse modelo é o *software* que apresentará um único formato de texto possível a ser desenvolvido pela criança.

Aqui o *software* utiliza da formatação das características comuns a um determinado gênero textual e guiará o aluno através de suas tentativas a internalizar as regras para a construção de determinado tipo e gênero de texto.

5.2 Diagrama de caso de uso

Nesta seção, são descritas as principais funcionalidades executadas no protótipo, de maneira a narrar sequência de intenções entre usuário (humano ou máquina) e o sistema.

5.2.1 *Aplicar o Jogo do Jornal*

Este caso de uso descreve o procedimento adotado pelo professor e o papel do Sistema de análise durante o transcorrer de uma sessão do Jogo do Jornal. O professor está autenticado no sistema. Um aluno está autenticado pelo sistema e pronto a participar do jogo. O cadastro está associado ao registro do aluno no sistema e tem como objetivo garantir que o usuário tenha uma identificação única. No momento que o usuário aluno realiza seu registro no sistema, é solicitado que ele escreva textos, de gêneros diferentes, em caixas de texto vazias.

É importante alertar que nos diagramas onde se lê paciente, na verdade se refere ao aluno e aplicador se refere ao professor, alterações em relação a essas denominações foram feitas posterior à tarefa de modelagem do sistema, e como não afetam de forma significativa o modelo, resolvemos manter as denominações antigas e alertar ao leitor nessa nota sobre sua alteração.

O cadastro do aluno está associado ao registro do usuário:professor no sistema e tem como objetivo garantir que o usuário: professor tenha uma identificação única. No momento em que o usuário:professor faz seu registro, ele deve fornecer o nome dos alunos que irão jogar o jogo, junto com o ano e a série de escolaridade. As etapas são: usuário:professor libera jornal para edição, o aluno envia o texto para o sistema. Este envia-o para o Sistema de Análise, que o analisa e apresenta relatório para o usuário:professor. A figura 5.3 apresenta o diagrama de caso de uso. As linhas retas do Diagrama 5.3 não representam a conexão entre as etapas realizadas tanto pelo aluno como pelo professor.

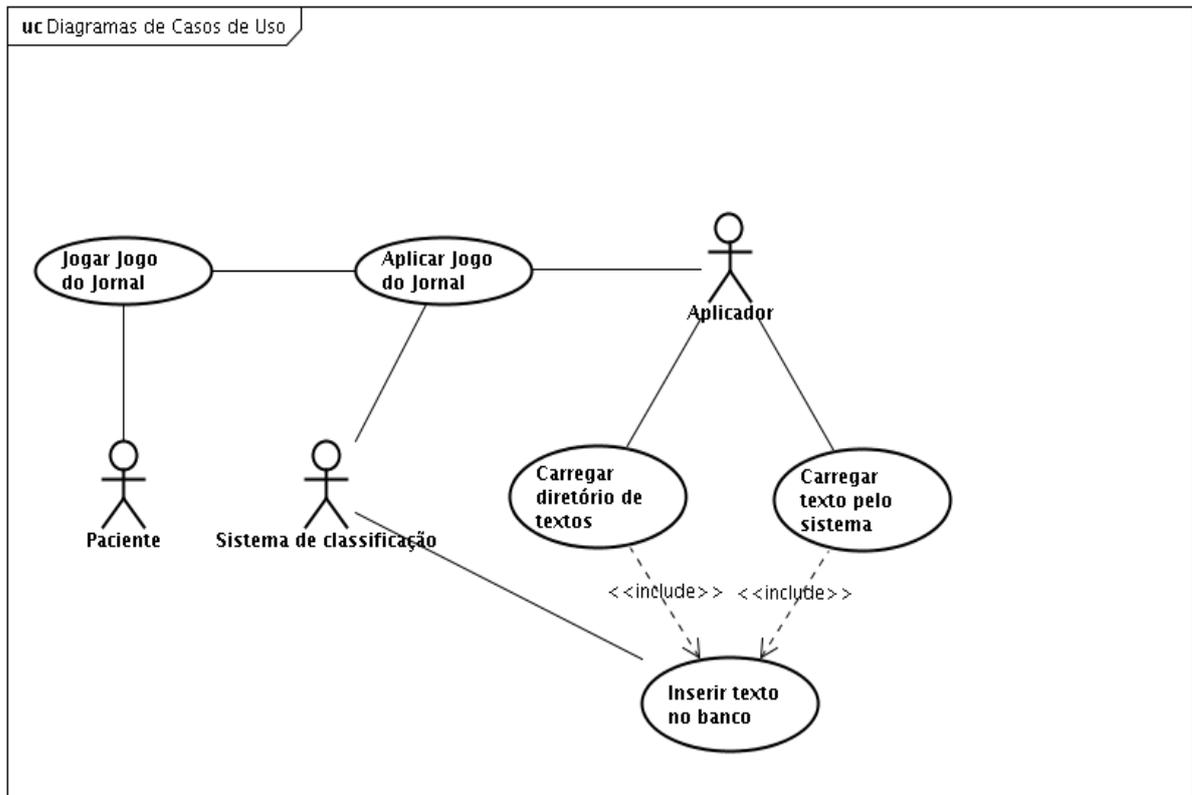


Figura 5.3: Diagrama UML de caso de uso relacionado ao mecanismo de cadastro.

5.2.3 Carregar diretório de texto

O usuário professor pode guardar uma coleção de textos armazenada num diretório de seu computador. Deste modo, o sistema pode carregá-los todos de uma só vez. Primeiro, o sistema apresenta tela de seleção de diretórios ao usuário:professor que escolhe o diretório. Feito isso, o sistema varre o diretório recursivamente, procurando textos. Para cada texto válido encontrado, o sistema executa caso de uso Inserir texto no banco.

5.2.4 Inserir texto no banco

O sistema do Jogo do Jornal deve ser capaz de inserir um texto associado a um gênero no seu banco de dados. Assim, o sistema insere texto e gênero no banco.

5.2.5 Jogar

Este caso de uso descreve as ações de um aluno durante uma sessão do Jogo do Jornal. O usuário aluno está autenticado no sistema. O usuário:professor () está autenticado no sistema e pronto para aplicar o jogo. O sistema apresenta o jornal, o aluno edita textos e submete.

5.3 Diagrama de caso de Classe

Um diagrama de classe tem como propósito descrever os vários tipos de objetos e o relacionamento existente entre eles. Veja Figura 5.4

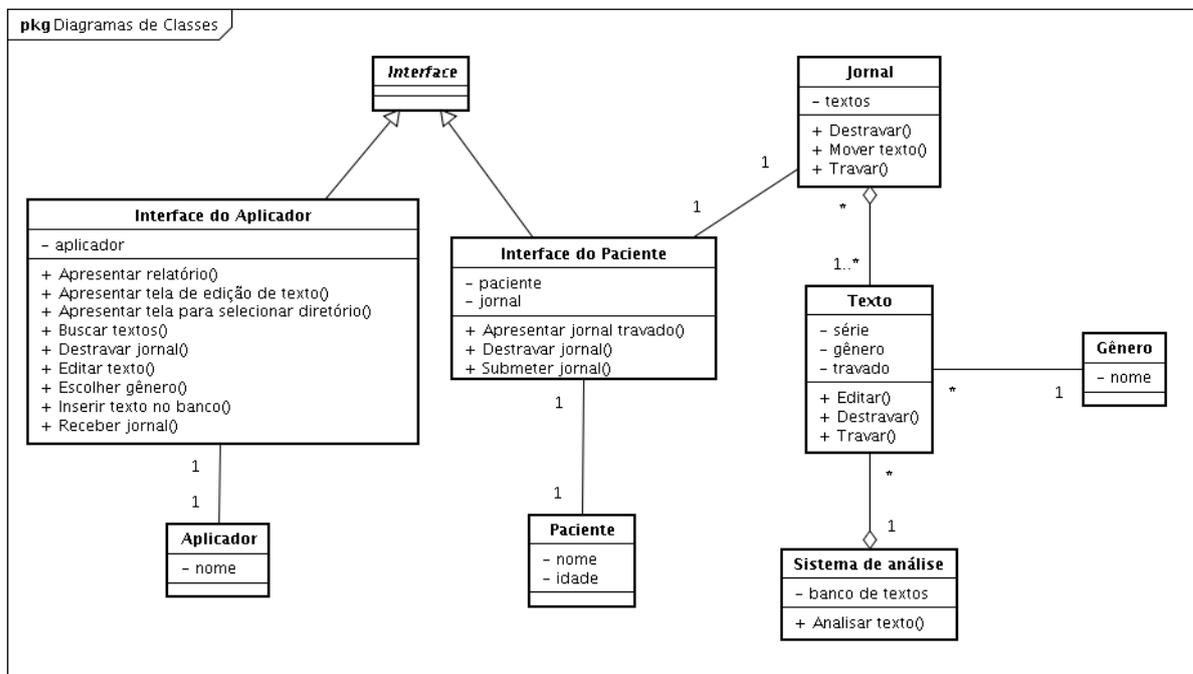


Figura 5.4: Diagrama de classes Conceitual do protótipo

5.4 Especificações do Protótipo

Esta aplicação é o cliente final. Três tipos de usuários têm acesso a ele: • O Administrador, o Professor e o Aluno.

- O Professor, que administra seus alunos e analisa os seus textos e resultados.
- O Aluno está aqui para jogar e escrever seus textos e enviá-los ao sistema. O Sistema de Will It Will analisa o texto de entrada e identifica o gênero do texto. Em seguida, identifica os critérios, quando não respeitados, notifica o usuário a esta lista, além de ajudá-lo a resolver esse erro com algumas dicas. Se o usuário passa de um limite de

tentativas no gênero, o sistema irá pedir-lhe para pedir a ajuda do professor. Todas as tentativas são registradas, e o professor pode ter acesso à elas.

5.4.1 Cadastro de Administrador

O administrador possui *login* de acesso ao *website* e faz qualquer ação; podendo sair a qualquer momento e cadastrar novas escolas com seu nome. Isto irá gerar automaticamente a eles uma senha e uma senha para os seus alunos (senha única para todos os estudantes da escola); tem acesso a lista de todas as escolas e suas senhas; podendo excluir escolas e ver os resultados do texto de cada aluno de cada escola. Os resultados são o número de tentativas, o texto produzido e comentários do professor sobre este texto;

O administrador tem acesso a pesquisa de cada escola; podendo ler o *log* do servidor, reiniciar e atualizar o servidor com o código mais recente. Como ilustra a Figura 5.5, ou para melhor visualização Anexo. 2.

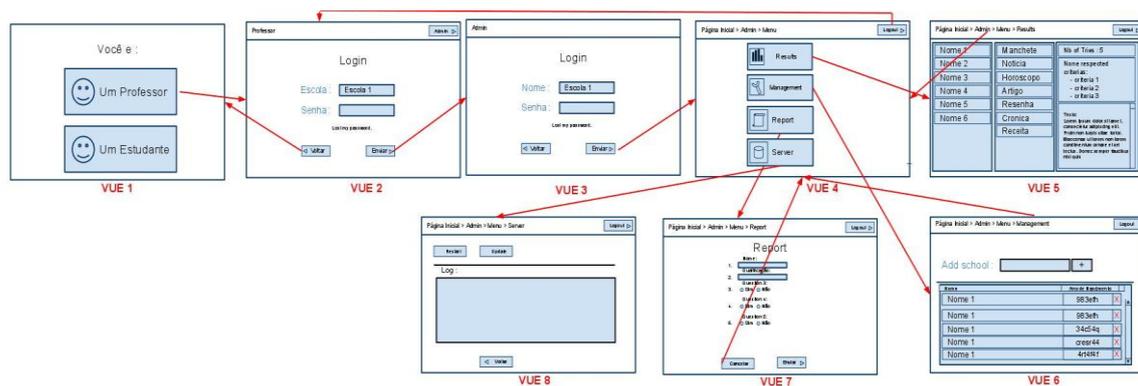


Figura 5.5 Cadastro do administrador

5.5.3 Cadastro de Aluno

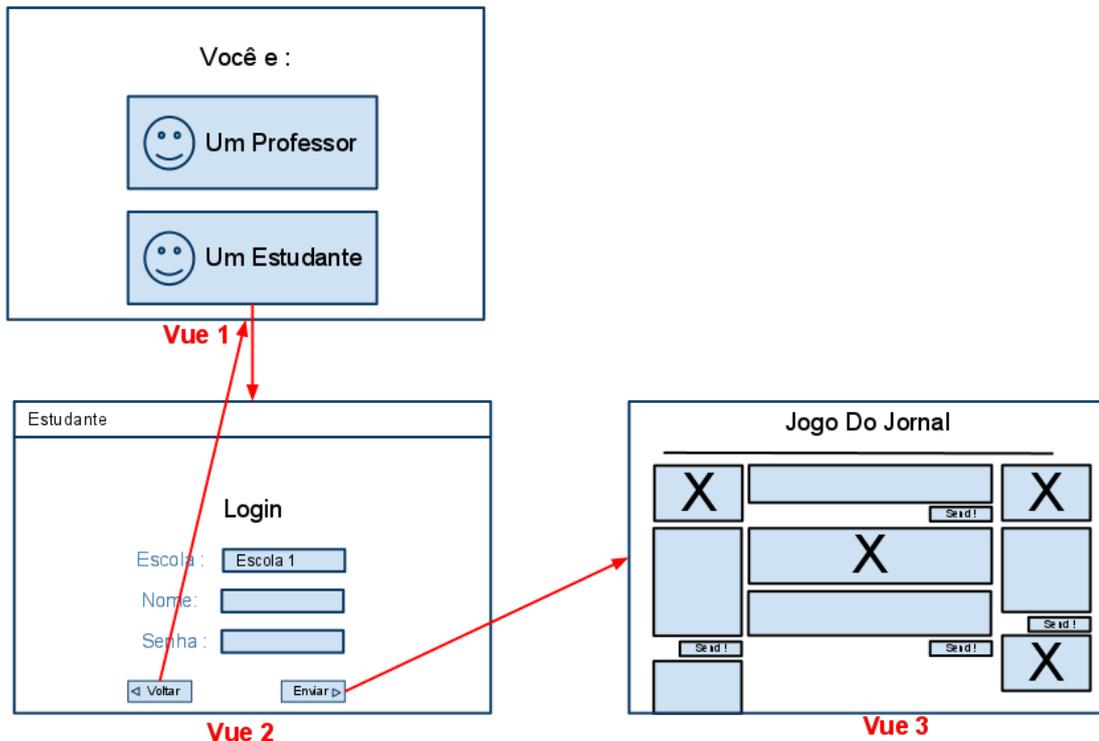


Figura 5.6: Cadastro do Aluno

O estudante faz *login* de acesso a sua página e ao *website* e pode fazer qualquer ação. Como estudante, pode sair a qualquer momento, acessar o jogo onde escreve cada gênero separado em caixas de texto, que são validados separadamente. Tendo um texto, ajudando-o a validar um texto. Esta lista os critérios para este gênero, e avisa como poderiam ser corrigidos. O estudante, tem um texto que diz em cada tentativa o que os sistemas espera dele (cerca de 3 tentativas de critérios para cada gênero).

5.4.3 Cadastro de Professor

O professor tem login de acesso à parte do *website* onde pode fazer qualquer ação e sair a qualquer momento; criar novo aluno com seu nome e grau; ver a lista de todos os alunos (nomes + graus); excluir aluno; ver a senha e fazer novas senhas, ver os resultados do texto de cada aluno. Os resultados são o número de tentativas. O professor pode comentar a produção dos alunos e responder a pesquisa. Veja Figura 5.6 ou Anexo 3.

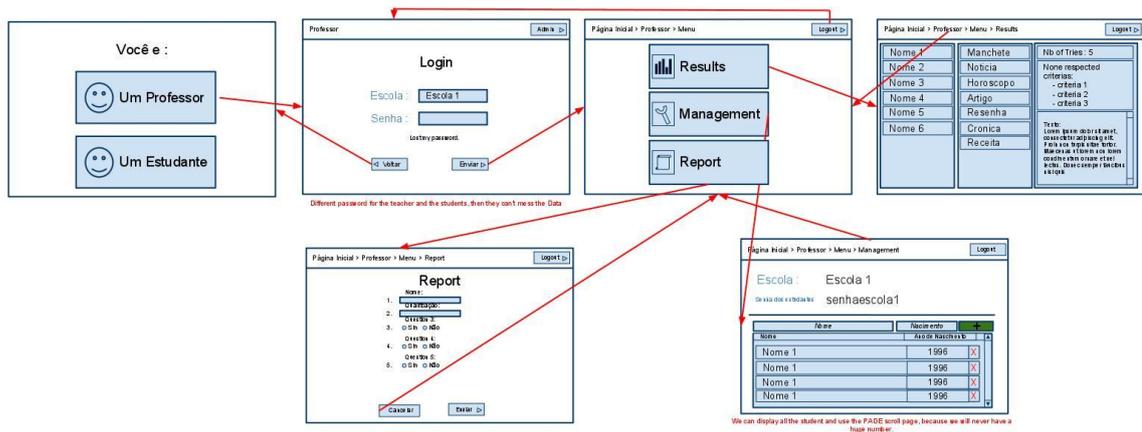


Figura 147: Cadastro do professor

5.4.4 Algoritmo

Em uma lista de *sites* ou fluxos RSS, de cerca de 2 ou 3 *sites* por gênero, o rastreador inspeciona cada nova página. O *software* verifica se ele já não tem o texto no banco de dados. Se ele não tem, então começa o texto do artigo e o salva no banco de dados na tabela do gênero. Esperam-se 12 horas para reiniciar o item 1.

Tabelas representam os gêneros e para cada texto nessas tabelas, analisa-se o texto com um conjunto de funções que representam os critérios para cada gênero da análise.

Usamos a classificação bayesiana de forma tratada para classificar o texto dado como o gênero ITS e Hadith. Para encontrar os atributos não respeitados do gênero utilizamos média bayesiana. Os atributos são, neste caso, equivalentes aos critérios de um gênero. Ou seja, textos são classificados como pertencentes a determinado gênero de acordo com a conformidade de certos atributos que contém esses atributos por sua vez podem ser critérios derivacionais, estruturais, lexicais, etc como descrito no Capítulo 3.

Há 32 atributos ou critérios 31 deles são expressões regulares que procuram alguns advérbios, alguns verbos, alguns adjetivos. Estes critérios são os listados em um arquivo que pode ser facilmente modificado pelos testes e não requer nenhuma programação. Veja Quadro 3.8.

Alguns critérios são mais frequentes do que outros para alguns gêneros, para esses um peso hadith é adicionado.

Todos os critérios são computados. Na verdade, um texto longo pode ter "poucos" critérios específicos que o determina. E um texto curto apresentar apenas critérios relevantes.

5.4.5 O rastreador ou *Crawler*

A fim de criar um grande banco de dados, feito a partir de textos que já sabemos os gêneros, criamos um rastreador. Um rastreador é um programa que visita *sites* de Web, lê as suas páginas e cria uma entrada para cada texto novo coletado no banco de dados.

Foram selecionados cerca de 2 ou 3 *sites* ou *blogs* por gênero e configuramos o rastreador para visitar e salvar as mensagens duas vezes por dia. Uma vez que já sabíamos o gênero de cada página visitada, salvamos diretamente o texto no banco de dados. A lista de referência a esses sites pode ser vista no Anexo 5.1

Mais tarde, esses registros foram utilizados para gerar a tabela exemplo de nossa classificação bayesiana.

5.5 Ferramentas utilizadas

Nessa seção faremos um resumo sobre as ferramentas utilizadas afim de preservar a esse texto do arcabouço técnico das definições e exemplos a cerca de cada ferramenta, assim optamos por fornecer maiores detalhes e explicações em Anexo 5.7.

A linguagem python foi escolhida por ser uma linguagem computacional completa e adequada ao desenvolvimento de aplicações baseadas na Web. Foram utilizadas também as Bibliotecas *Feedparser* e *BeautifulSoup* para obter o texto do artigo a partir de um fluxo de *RSS*.

Para analisar a página *HTML*, utilizamos um *Soup* nome python biblioteca *beautiful*. Já a classificação Bayesiana foi utilizada para classificar o texto dado como o gênero *ITS* e *Hadith* como descrito na Seção 5.4.4. Por fim o *Tornado web server* + *nginx* para lidar com o arquivo estático do *website* e o *Orange* para construção das tabelas com os atributos de cada gênero textual para o processo de calibração do classificador. Pois utilizamos alguns dos critérios lingüísticos, através de expressões regulares, para inferir melhor nos resultados e amarrar-los segundo sua classificação teórica.

5.6 Visão Geral do protótipo

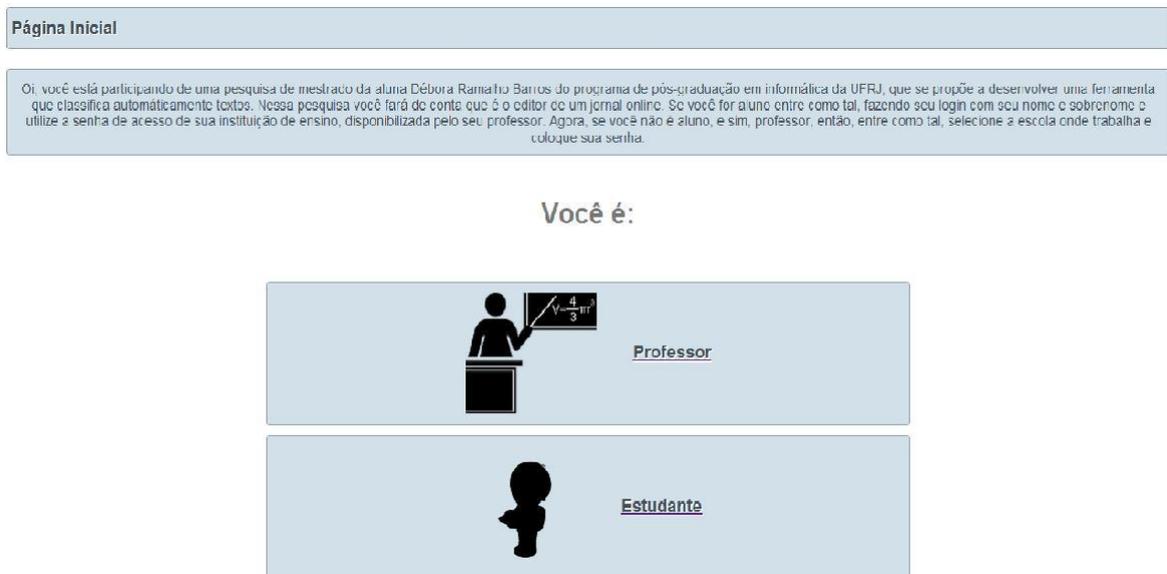


Figura 5.8: Tela inicial do Jogo do Jornal.

Nessa seção iremos descrever a interface do jogo, desenvolvida para a realização do estudo de caso dessa pesquisa, que pode ser acessado pelo endereço <http://146.164.248.76/>

A página inicial da aplicação Jogo do Jornal (conforme mostra a figura.5.9, para melhor visualização ver Anexo 5.9) apresenta uma explicação sobre o projeto e uma introdução da tarefa a ser realizada durante o jogo e instruções de acesso. Após a leitura os usuários, professor e aluno, escolhem sua forma de acesso.

5.7 Cadastro dos Usuários

O usuário professor somente terá acesso à ferramenta inicialmente após autorização pelo administrador, responsável por fazer seu cadastro no jogo, ao ser cadastrado o professor recebe uma senha de acesso e pode utilizar a ferramenta.

O administrador tem acesso ao jogo através da página de *login* do professor, onde fica localizada na parte superior da tela do lado direito um caminho para o administrador fazer seu *login*, como pode verificar na figura 5.9

Página Inicial > Professor Logout

Login:

Escola:

Senha:

Figura 5.9: Tela de login do professor.

O professor recebe do administrador uma senha e com ela tem acesso ao jogo, na tela inicial do professor existem três tarefas de *management*, resultado e questionário. Em *management*, o professor faz o cadastro dos alunos que participarão do jogo, o nome cadastrado será o *login* desses alunos. Devem ser cadastrados também pelo professor o ano e a série de escolaridade desses alunos. Veja figura 5.10.

Página Inicial > Professor > Menu > Management Logout

Olá professor, nessa página você deve colocar o nome dos alunos que irão participar da pesquisa e o ano letivo desses alunos.

Escola: I.E Roberto Silveira (E.F)

Senha dos Estudantes:

Adicionar um estudante :

Nome Grade +

Os estudantes :

Nome	Grade	Delete
Paula Freitas	8anoEF	X
Isabella	8ano	X
Jéssica Netto	8ano	X
Milena	8ano	X
Ketlin	8ano	X
Beatriz	8anoFF	X

Figura 5.10: Tela da tarefa de management.

O aluno faz seu *login* entrando como estudante, escolhe sua escola e coloca seu nome conforme cadastrado pelo seu professor e uma senha também fornecida pelo seu professor.

Login:

Escola :

Nome :

Senha :

Figura 5.11: Tela de login de aluno.

5.8 Jogar o Jogo do Jornal

Após realizar o *login*, o aluno terá acesso a uma página com ilustrações que sugerem gêneros de textos a serem escritos dentro de caixas de texto vazias com somente a função de digitar texto e um botão de enviar seus textos ao final de sua realização. Como os alunos são levados a imaginar o contexto de faz de conta que são editores de um jornal *online*, para cada pergunta de mediação eles receberão uma mensagem de seus chefes indicando o que deverá ser feito a seguir.

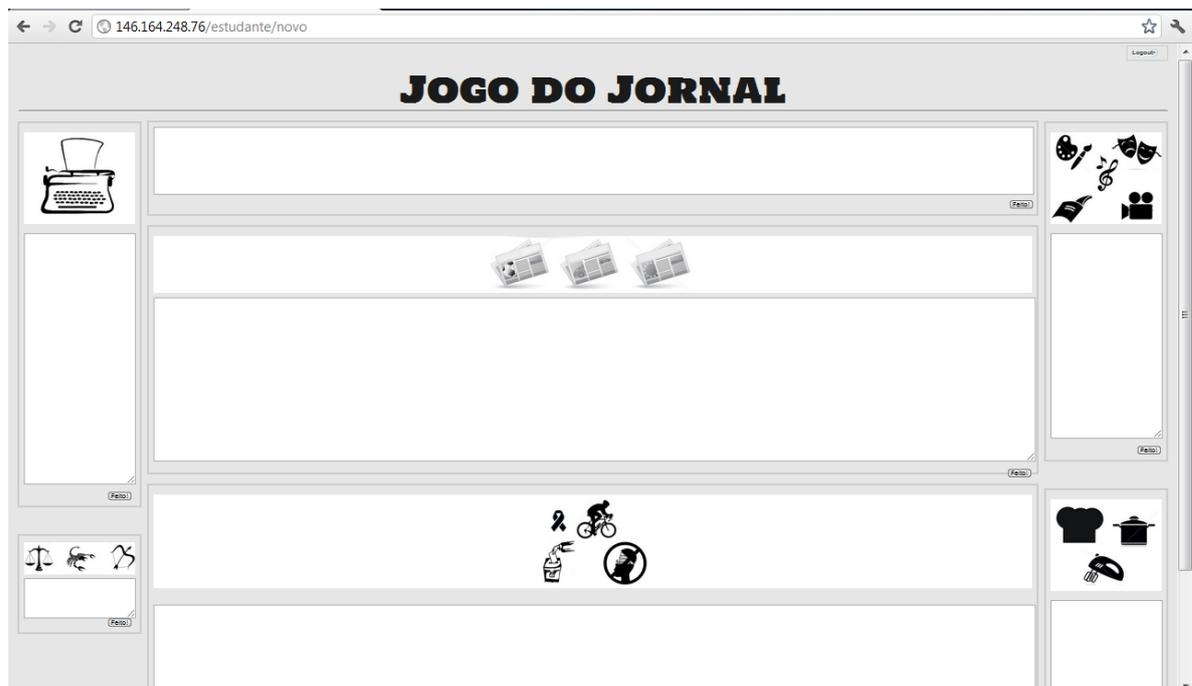


Figura 5.12: Tela de login de aluno.

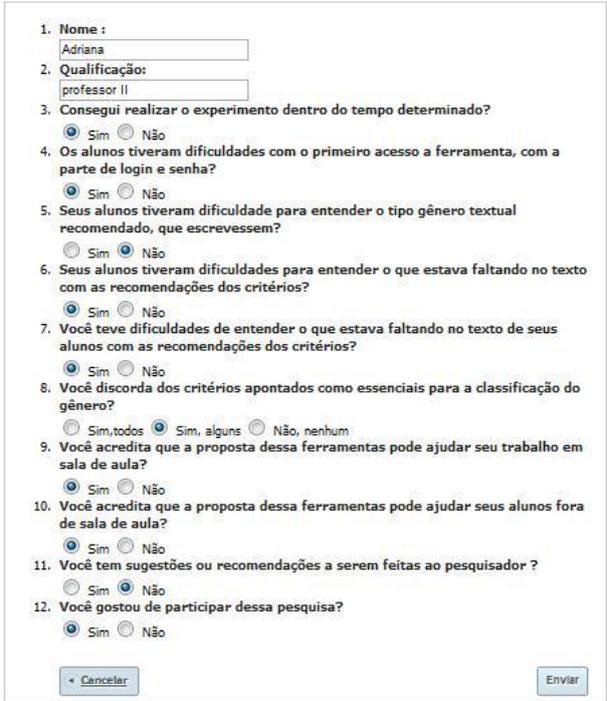
5.9 Tarefas do Professor

Além de cadastrar os alunos, o professor também terá acesso a duas outras tarefas, chamadas de *results* e *report*, que são: a primeira, acesso ao resultado da *desempenho* e textos produzidos por seus alunos, como também o número de tentativas e recomendações utilizadas, a classificação automática do texto do aluno pelo classificador, os critérios que ficaram faltando e um espaço para comentar a produção do aluno. Veja figura 5.13

Figura 5.13: Tela da tarefa de Resultados.

Ao final do trabalho com os alunos, o professor pode responder a um questionário (*report*) que visa verificar os pontos positivos e negativos da ferramenta que ainda precisam de ajustes.

Para a avaliação da proposta, foi implementado um questionário formado por doze perguntas de múltipla escolha para verificar a opinião do professor sobre o jogo, o seu objetivo foi aferir a contribuição da ferramenta para a aprendizagem dos alunos. Verificar na figura 5.13



1. Nome :
Adriana

2. Qualificação:
professor II

3. Consegui realizar o experimento dentro do tempo determinado?
 Sim Não

4. Os alunos tiveram dificuldades com o primeiro acesso a ferramenta, com a parte de login e senha?
 Sim Não

5. Seus alunos tiveram dificuldade para entender o tipo gênero textual recomendado, que escrevessem?
 Sim Não

6. Seus alunos tiveram dificuldades para entender o que estava faltando no texto com as recomendações dos critérios?
 Sim Não

7. Você teve dificuldades de entender o que estava faltando no texto de seus alunos com as recomendações dos critérios?
 Sim Não

8. Você discorda dos critérios apontados como essenciais para a classificação do gênero?
 Sim, todos Sim, alguns Não, nenhum

9. Você acredita que a proposta dessa ferramentas pode ajudar seu trabalho em sala de aula?
 Sim Não

10. Você acredita que a proposta dessa ferramentas pode ajudar seus alunos fora de sala de aula?
 Sim Não

11. Você tem sugestões ou recomendações a serem feitas ao pesquisador ?
 Sim Não

12. Você gostou de participar dessa pesquisa?
 Sim Não

Figura 5.14: Tela do relatório.

5.10. O Administrador

O administrador funciona como uma espécie de responsável pelo andamento e desenvolvimento do jogo *online*. Funciona como o gerente das tarefas realizadas no jogo, assim como também na implementação de mudanças para seu melhor desempenho. Ele tem acesso a tudo o que acontece no ambiente do jogo, é responsável por cadastrar os professores, tem acesso às tarefas do professor e aos resultados do desempenho e textos de cada aluno de cada professor, às respostas do questionário respondido por eles, e ao servidor, a fim de evitar que o jogo saia do ar garantindo seu melhor desempenho. As tarefas do administrador podem ser verificadas na figura 5.15:

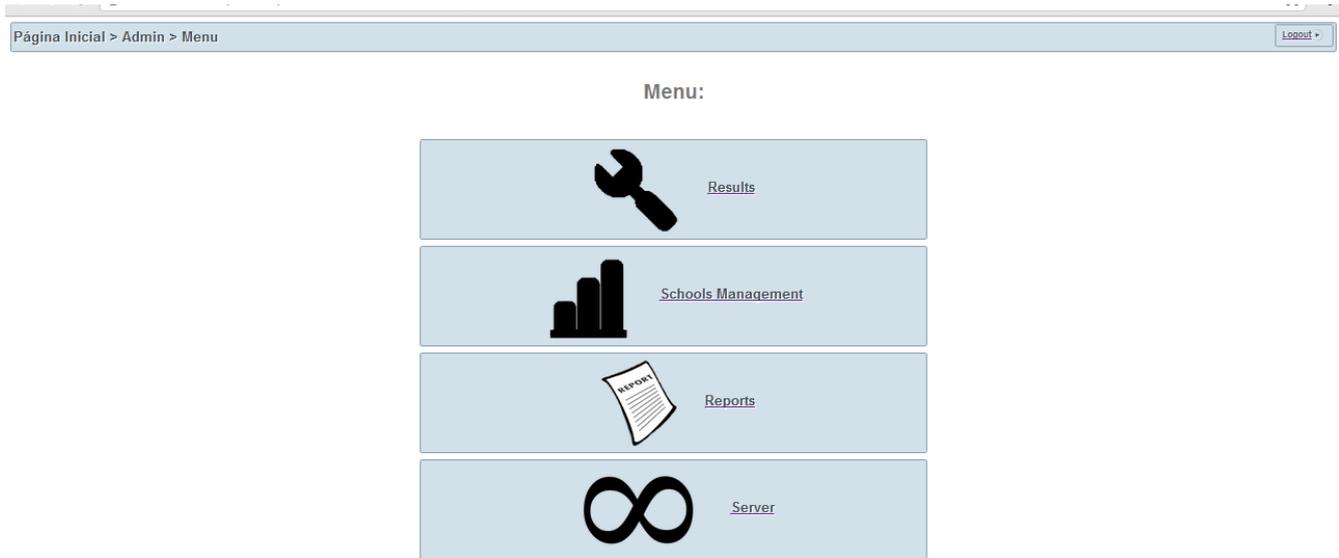


Figura 5.14: Tela das tarefas do administrador.

6. ESTUDO DE CASO

Neste Capítulo são apresentados dois estudos de casos, que utilizam as funcionalidades desenvolvidas no ambiente de classificação automática de texto. Pretende-se com isso mostrar que a ferramenta é capaz de apresentar resultados satisfatórios para seu uso em sala de aula. Para a realização do segundo experimento foi utilizado o *software* descrito no Capítulo 4.

6.1 Estudo de caso piloto

Nossa pesquisa se propõe a utilizar a técnica psicopedagógica da elaboração dirigida (Seminário, 1997). A possibilidade de se trabalhar com a diversidade textual e a psicopedagogia foi construída a partir de um estudo de caso piloto, realizado no primeiro ano dessa pesquisa, com dois grupos heterogêneos de amostras de 9 alunos do ensino Fundamental, cinco meninos entre 10 e 14 anos e 4 meninas entre 10 e 13 anos, cujo objetivo era verificar a possibilidade de intervenção dialogada do professor em relação à produção textual desses alunos. Sob a hipótese de que, se possível, então a construção de uma ferramenta a partir do crivo do especialista poderia ser feito. Assim, automatizando o trabalho com a produção textual dentro de sala de aula de maneira eficiente e extensiva para todos os alunos.

O experimento foi realizado em dentro de sala de aula do curso livre Cultura Inglesa, com autorização da instituição de ensino e dos responsáveis pelos alunos. Durante (1) uma semana os alunos tiveram que montar em uma cartolina um Jornal, a professora disponibilizou no quadro uma lista de 14 gêneros textuais que eles poderiam utilizar para compor seus jornais, já que na época o escopo de pesquisa era maior.



Figura 6.1: Estudo de Caso Piloto.

Os Textos dos alunos foram avaliados segundo um crivo levantado à luz das teorias de categorização textual (Travaglia, 2007) e aqueles que se distanciaram das classificações foram trabalhados individualmente com a professora, através da técnica dialogada da elaboração dirigida.

Portanto, essa foi à maneira como os textos dos alunos foram avaliados, para cada jornal produzido, o professor leu os textos e verificou nas tabelas dos Quadros 3.1, 3.2, 3.3, 3.4 e 3.5 do capítulo 3 se os critérios haviam sido respeitados, para a produção de cada gênero. Pois, os alunos escolhiam previamente o gênero que seria produzido por eles, assim a verificação ocorreu de maneira pontual, verificou-se se o texto escrito era ou não correspondente com o gênero escolhido e senão quais desses gêneros ele se encaixava melhor, segundo o número de critérios encontrados.

A Figura 6.2 é a fotografia de um jornal produzido por um dos alunos da amostra. Nela é verificado o título do jornal: Jornal Legal e os gêneros escolhidos: manchete, artigo e notícia. Para melhor visualização ver Anexo 6.2.

A figura 6.3 é a fotografia do texto desse aluno antes da intervenção dialogada que se propusera ser um artigo e foi classificado pelo crivo como uma notícia, pois constavam mais características de notícia do que de artigo. A aferição desse resultado foi feita através de perguntas de mediação e preenchidos tabelas do Quadro 6.8, veja Seção 6.1.3 para maiores detalhes.

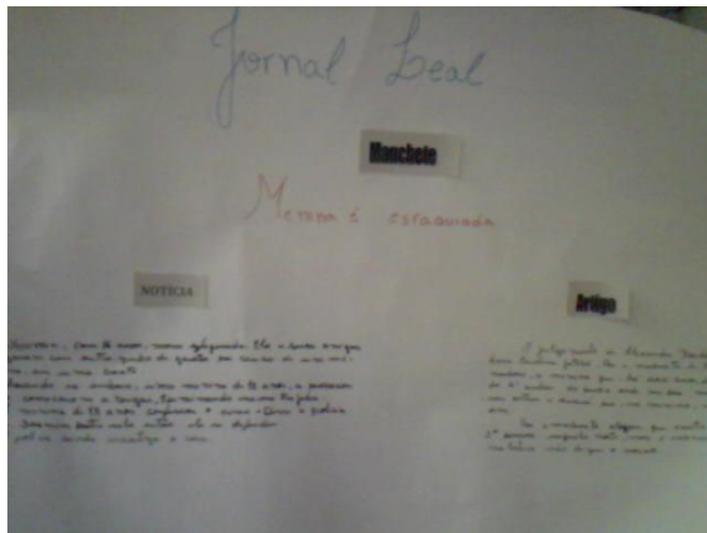


Figura 6.2: Jornal Legal, da aluna, 9º ano instituição particular.

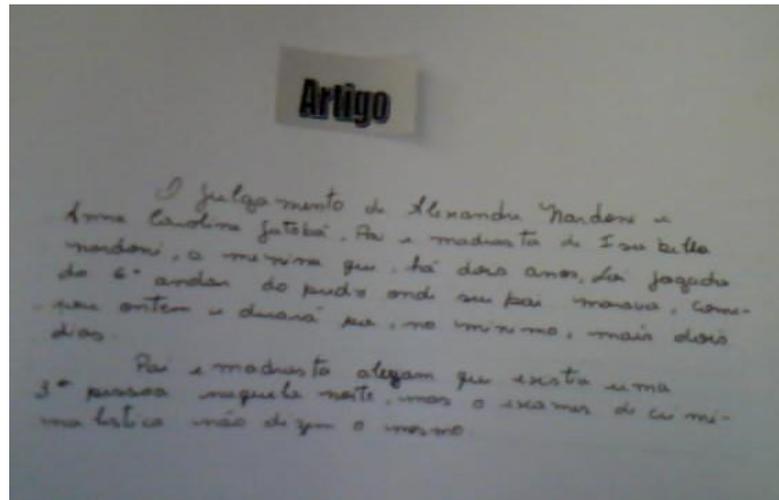


Figura 6.3: Texto antes da intervenção dialogada

6.1.1 Resultados

O resultado desse estudo piloto foi medido semi-qualitativamente, pois do total de oito projetos executados foram reavaliados: a questão do escopo, as perguntas de intervenção e os critérios selecionados para composição do crivo e do aprendizado.

6.1.2 Escopo

A intenção de se trabalhar com um escopo de 14 gêneros foi revista com esse estudo piloto. Do escopo de 14 gêneros, a média de produção foi de 3 gêneros por projeto, veja no Quadro 6.6: 1) manchete 2) artigo 3) entrevista, 4) resenha (crítica) 5) quadrinhos/tirinhas 6) horóscopo 7) *slogans* 8) anúncios 9) receitas 10) cartas do leitor 11) crônica 12) editorial 13) reportagem 14) notícia.

Amostra	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Aluno1	■					■			■					■
Aluno2	■			■		■			■		■			■
Aluno3						■			■					■
Aluno4	■								■					■
Aluno5	■					■			■					■
Aluno6					■						■			■
Aluno7		■												■
Aluno8	■	■												■
Aluno9	■								■					■

Quadro 6.6: Produção de textos por alunos e gêneros

A partir desse resultado, repensamos nosso escopo e a dificuldade de se automatizar critérios para 14 gêneros. Reduzimos nosso escopo levando em consideração a escolha de gênero feita pelos participantes, facilidade de coleta e armazenamento desses para a composição do banco

de textos da ferramenta de classificação automática. Chegamos então ao escopo de sete gêneros textuais, os gêneros (1), (2), (4), (6), (9), (11) e (14) do Quadro 6.6.

6.1.3 Perguntas de intervenção

Cada critério do crivo corresponde a uma pergunta, que servirá de parâmetro para a intervenção na fase 4 (Elaboração Dirigida), ainda separadas por tópicos e não por gêneros. A seguir, a API utilizada na classificação do texto e trabalho de intervenção da aluna do 9º ano de instituição Federal.

CONTEÚDO TEMÁTICO

	Sim	Não
1. Traz o testemunho direto de fatos e situações?		X
2. Trata de uma perspectiva atual?	X	
3. Apresenta a possibilidade de o autor expressar-se?		X
4. É uma matéria escrita sobre algum tema específico		X
5. Traz dados bibliográficos?		X
6. Faz alusão a uma notícia?	X	
7. É uma matéria escrita sobre algum tema específico?		X
8. Traz o posicionamento da empresa de comunicação sobre determinado assunto?		X
9. Traz argumentos?		X
10. Um texto em forma de síntese que expressa à opinião do autor sobre um determinado fato cultural, que pode ser um livro, um filme, peças teatrais, exposições, <i>shows</i> etc?		X
11. Faz uma crítica direta?		X
12. O texto não é espontâneo nem natural		
13. Uma obra em que o autor cria os diálogos e as situações que envolvem os falantes, previamente.		X
14. Divide-se em estágios na maioria dos casos uma verdade, uma crença arraigada na cultura popular, que ajuda a fazer o leitor recuperar um intertexto, evidentemente, pretendido pelo autor do texto.		X
15. A relação entre texto, contexto, conhecimento de mundo e coerência	X	
16. Deve ser memorável (capacidade de a frase ser lembrada sem auxílio)		X
17. A grande ideia deve ser transmitida no anúncio		X
18. Comida e Lista de ingredientes que serão utilizados, com a respectiva quantidade		X
19. Dar uma opinião, uma sugestão ou debater um tema		X
20. O tema é o cotidiano e que leva o leitor a refletir sobre a vida comentando assuntos do dia a dia		X
21. Divulgação de um acontecimento.	X	

Quadro 6.7: intervenção dialogada do conteúdo temático

ESTRUTURA COMPOSICIONAL

	Sim	Não
1. É um título no alto da primeira página de jornal ou revista?		X
2. Possui 4 parágrafos?		X
3. Traz informações novas e inéditas?	X	
4. Possui tema específico e um autor?	X	
5. Faz um posicionamento crítico?		X
6. Traz justificativas?	X	
7. Traz exemplos concretos e dados factuais?	X	

8. Tem uma conotação negativa?	X	
9. Apresenta um consenso de várias opiniões sobre determinado assunto?	X	
10. Apresenta opiniões de terceiros?		X
11. Traz o testemunho direto de fatos?		X
12. Apresenta argumentos?		X
13. O texto traz dados bibliográficos?		X
14. Faz uma divisão em Capítulos ou seções?		X
15. Utiliza de 3 a 5 parágrafos?		X
16. Utiliza o requadro, o balão, e as legendas?		X
17. Faz metadiscorso, através de marcadores pessoais, <i>hedges</i> e enfatizadores?		X
18. Evita a linguagem atitudinal normalmente associada ao significado interpessoal?		X
19. Inclui um benefício-chave, deve diferenciar a marca, lembrar o nome da marca fazer rima com ele?		X
20. Transmite sentimentos positivos sobre a marca?		X
21. Faz uma propaganda, um pedido, ou outros tipos de comunicação?		X
22. Traz uma sequência ordenada de procedimentos que deverão ser realizados		X
23. Informação sobre o tempo de preparo a quantidade que será produzida?		X
24. Especifica o assunto e é breve; trace previamente o objetivo da carta (opinar, sugerir, debater)?		X
25. Narra fatos históricos em ordem cronológica. Trata de temas da atualidade		X
26. Trazem as pessoas comuns como personagens, sem nome ou com nomes genéricos sem aprofundamento psicológico?		X

Quadro 6.8: intervenção dialogada Estrutura Composicional

OBJETIVOS E FUNÇÕES SOCIOCOMUNICATIVAS

	Sim	Não
1. Critica alguma coisa?		X
2. Chama atenção para um determinado assunto?	X	
3. Faz a exposição de algum assunto de interesse mútuo?	X	
4. Aprofunda mais em uma notícia ou determinado tema?	X	
5. Veículo pelo qual o autor expressa sua opinião sobre determinado assunto ou o resultado de estudos e pesquisas de determinado assunto?		X
6. Traz o posicionamento da empresa de comunicação sobre determinado assunto que direciona as matérias publicadas por ele?	X	
7. Apresenta o formato de divulgação de um acontecimento excepcional, anormal e de grande impacto social?	X	
8. Guia o leitor pelo emaranhado da produção cultural que cresce a cada dia e que tende a confundir até os mais familiarizados com todo esse conteúdo?		X
9. Expressa a opinião do autor sobre um determinado fato cultural, que pode ser um livro, um filme, peças teatrais, exposições, <i>shows</i> etc?		X
10. Analisa de forma crítica: sua opinião?		X
11. Recomenda uma obra?		X
12. Entretém, fazendo ou não críticas?		X
13. Tem como finalidade entretenimento e aconselhamento?		X
14. É memorável (capacidade de a frase ser lembrada sem auxílio)?		X
15. Faz uma propaganda, um pedido, ou outros tipos de comunicação que interesse ao público em geral, necessitando ou não de uma resposta para o autor?		X
16. Faz com que o destinatário possa fazer comidas a partir da leitura da receita?		X
17. Opina, sugere ou debate um determinado tema?		X
18. Analisa os fatos que acontecem e dá um colorido emocional, mostrando aos olhos do leitor uma situação comum, vista por outro ângulo, singular?		X
19. Envolve e emociona o leitor?		X
20. Fazer o posicionamento da empresa de comunicação sobre determinado assunto?		X

Quadro 6.9: intervenção dialogada Objetivos e funções sócias comunicativas

CARACTERÍSTICAS DA SUPERFÍCIE LINGÜÍSTICA

	Sim	Não
1. Faz a omissão das palavras desnecessárias (adjuntos adnominais e adverbiais)?		X
2. Usa abr ATEH (para até), IMPAGO (não pago) e outras?		X
3. Introduce um argumento apresentado como acréscimo, um argumento a favor de uma determinada conclusão, além disso, ainda mais, além de, e mais, e, também, nem, nem mesmo, ademais, não apenas... Mas, não apenas... Mas também, e não só por que... Mas também, não só... Mas também, sobretudo, e também, além, mais uma vez, mais ainda, aliás, ainda, e ainda?		X
4. Assinala uma oposição, mas, mais ainda, mas também, entretanto, porém, contudo, todavia, do contrário, no entanto, agora, ao contrário. - embora, mesmo que, apesar de, ainda que?	X	
5. Utiliza conectores para os mais diferentes tipos de relações? Conjunção, contração ou oposição, disjunção ou alternância, causa /consequência (explicativas, justificativas, causais, condicionais, finais, consecutivas, conclusivas), comparação (comparativas, (conformativas), comprovação, especificação (incluindo a exemplificação) entre outras?	X	
6. Usa de adjetivos e gírias?		X
7. Faz descrições?		X
8. Utiliza argumentação?	X	
9. Apresenta de 3 a 5 parágrafos?		X
10. Discurso direto, onomatopéia, expressões populares?		X
11. Prolongamento e intensificação de sons?		X
12. Usa linguagem persuasiva?		X
13. Faz construções condicionais com conectivo explícito, isto é, com se ou implícito?		X
14. Lembra o nome da marca e o ideal é que o nome da marca seja incluído na frase?		X
15. Utiliza linguagem popular como um chavão ou aliteração relevante e palavras cunhadas ou compostas?		X
16. Utiliza trocadilhos e rimas?		X
17. Faz propaganda, um pedido, ou outros tipos de comunicação?		X
18. Escreve texto impessoal, usando de 3ª pessoa (do ponto de vista morfológico) com valor de 2ª pessoa (do ponto de vista discursivo)?		X
19. Usa o imperativo?		X
20. Usa Advérbio e locuções de modo (lentamente, levemente, bem devagar etc.) e, eventualmente, de tempo (depois, seguida, 20 minutos etc.)		X
21. Apresentam nomes de alimentos e temperos, adjetivos e locuções adjetivas específicas (brando, fresca, média, fervente, fria, quente etc.), e verbos específicos de atividades culinárias (cortar, picar, lavar, misturar, bater, despejar, colocar, arrumar, descascar, cozinhar, preparar, juntar, escorrer, ferveretc.)?		X
22. Escreva em uma linguagem clara, precisa e nunca faça uso de palavras de baixo calão?		X
23. Indica o texto fazendo referência ao assunto, tais como: "Com relação a...", "Em atenção ao anúncio publicado..."?		X
24. Usa oralidade na escrita: linguagem coloquial?		X
25. Utiliza uma dose de lirismo?		X
26. Utiliza humor.?		X

Quadro 6.10: intervenção dialogada Características da superfície lingüística.

CONDIÇÕES DE PRODUÇÃO (quem, para quê, quando, onde)

	Sim	Não
1. Escreve como se fosse o Jornalista, num jornal na 1ª página?	X	
2. A informação demonstra ser importante e significativa?	X	

3. Escreve para o suporte JORNAL?	X	
4. Escreve para o suporte REVISTA?		X
5. Escreve para os leitores em geral?	X	
6. Escreve para os jornalistas que trabalham em um determinado jornal ou revista?		X
7. Escreve um texto em forma de síntese que expressa a opinião do autor sobre um determinado fato cultural, que pode ser um livro, um filme, peças teatrais, exposições, shows etc.?	X	X
8. Escreve para um público seletivo?		X
9. Faz uma crítica direta?		X
10. Identifica-se como autor, assinando?		X
11. O espaço e o tempo em que os fatos ocorrem são produtos de um planejamento prévio tanto do tema quanto do aspecto lingüístico-discursivo, sujeito a correções?		X
12. Parece ter sido escrito por astrólogo para pessoas guiadas pela força do gênero que buscam explicações e previsões?		X
13. Faz-se de vendedor para divulgação ou venda de algo ou conceito para o público os leitores?		X
14. Em jornais impressos e revistas, além de <i>websites</i> , os anúncios podem ser publicados junto ao conteúdo editorial ou em seções separadas, denominado classificado?		X
15. Algumas vezes não é identificado. Outras vezes é alguém especialista no assunto que escreve para jornais ou revistas ou livros. Em todos os casos, é alguém que sabe fazer uma determinada comida e pretende ensinar como fazer?		X
16. Escreve como um <i>chefe</i> de cozinha para leitores de jornal, revistas e livros de culinária. Alguém que pretende fazer uma determinada comida?		X
17. Faz-se de leitor para opinar, sugerir ou debater sobre determinado assunto no jornal?		X
18. Demonstra estar dentro de um espaço sempre igual e com a mesma localização?		X

Quadro 6.11: intervenção dialogada Condição de Produção

6.1.4 Critérios selecionados para composição do crivo

A lista de critérios encontra-se extensa demais e devido a isso a apuração dos resultados manualmente demandou muito tempo e esforço do pesquisador, visto que a organização do formato de tabela escolhido tem o objetivo de gerar uma classificação para o texto analisado entre um dos 14 gêneros do universo estudado. Para então, apontar a eficiência dos critérios.

Como pode observar a análise é feita a partir do resultado da correção do texto pelo professor que utiliza as perguntas acima como um verificador do crivo. Para cada resposta positiva (sim), deve-se localizar na tabela do crivo, Quadros 6.7 -6.11, o gênero apontado. Assim espera-se que o gênero mais apontado seja a classificação do texto da aluna do 9º ano.

As tabelas dos Quadros 6.12 - 6.16 trazem o resultado da análise do texto, a primeira coluna corresponde ao número do critério positivo, e a segunda, ao gênero que ele se refere. Por exemplo, na tabela correspondente a superfície lingüística, pergunta 4 corresponde à pergunta: *Assinala uma oposição, mas, mais ainda, mas também, entretanto, porém, contudo, todavia, do contrário, no entanto, agora, ao contrário. - embora, mesmo que, apesar de, ainda que?* A resposta foi sim, se verificarmos o critério 4 do Quadro 6.4 iremos perceber que ele corresponde ao gênero artigo (2). Como a classificação de gêneros não é tão simples e nem pode ser realizada de

maneira ingênua, respeitamos a questão de mobilidade dos critérios que podem aparecer em um ou mais gêneros específicos, não sendo únicos em um determinado gênero. O resultado está baseado na frequência de aparição de determinado gênero no resultado. Por exemplo, no texto demonstrado aqui encontramos oito (08) possibilidades de gênero em sua composição, entre eles reportagem (03), manchete (03), *slogan* (01), anúncio (01), crônica (01), notícia (08), artigo (08), editorial (03). Assim o resultado encontrado é uma interseção entre notícia e artigo.

O professor compara o resultado com a intenção de gênero a ser escolhida pela aluna e a partir do gênero que fora escolhido para ser produzido, oferece a intervenção dialogada para que esse gênero seja alcançado. Esta intervenção é feita a partir das perguntas do crivo de classificação, com objetivo de levar o aluno à reflexão sobre o que gostaria de colocar e o que de fato colocou em seu texto e a definição do gênero segundo sua definição e o que é realmente.

Superfície Linguística	
04	Artigo
05	Notícia Reportagem
08	Resenha

Quadro 6.12: Resultado da Superfície Linguística

Estrutura Composicional	
03	Notícia
04	Artigo
06	Editorial
07	Artigo
08	Notícia
09	Artigo/ Editorial/ Notícia

Quadro 6.13: Resultado da Estrutura Composicional

Conteúdo Temático	
2	Reportagem
6	Manchete
16	<i>Slogan</i>
22	Anúncio/ Crônica/ Notícia.

Quadro 6.14: Resultado do Conteúdo Temático

Condições de Produção	
01	Manchete
02	Artigo/notícia
03	Artigo
05	Artigo/Editorial/Reportagem/ Notícia
07	Editorial

Quadro 6.15: Resultado da Condição de Produção

Obj. E Funç. Sociocomunicativas.	
02	Manchete
03	Artigo
04	Notícia
06	Notícia
07	Notícia

Quadro 716: Resultado dos Objetivos e Funções sócio comunicativas

A intervenção foi feita a partir das perguntas sobre as características de artigo que não constavam no texto. Inicialmente, a professora ativou o conhecimento de mundo da aluna em relação ao conceito de artigo que ela tinha, e em seguida foi questionando a sua definição com as características de artigo que não estavam no texto, evitando dar respostas para o aluno. O resultado disso foi à adaptação da própria aluna de seu antigo texto, com as características que foram discutidas, assim ao passar no crivo novamente o texto ganhou características de artigo que o colocou na intercessão do gênero, podendo sim ser classificado como artigo. Veja Figura 6.8

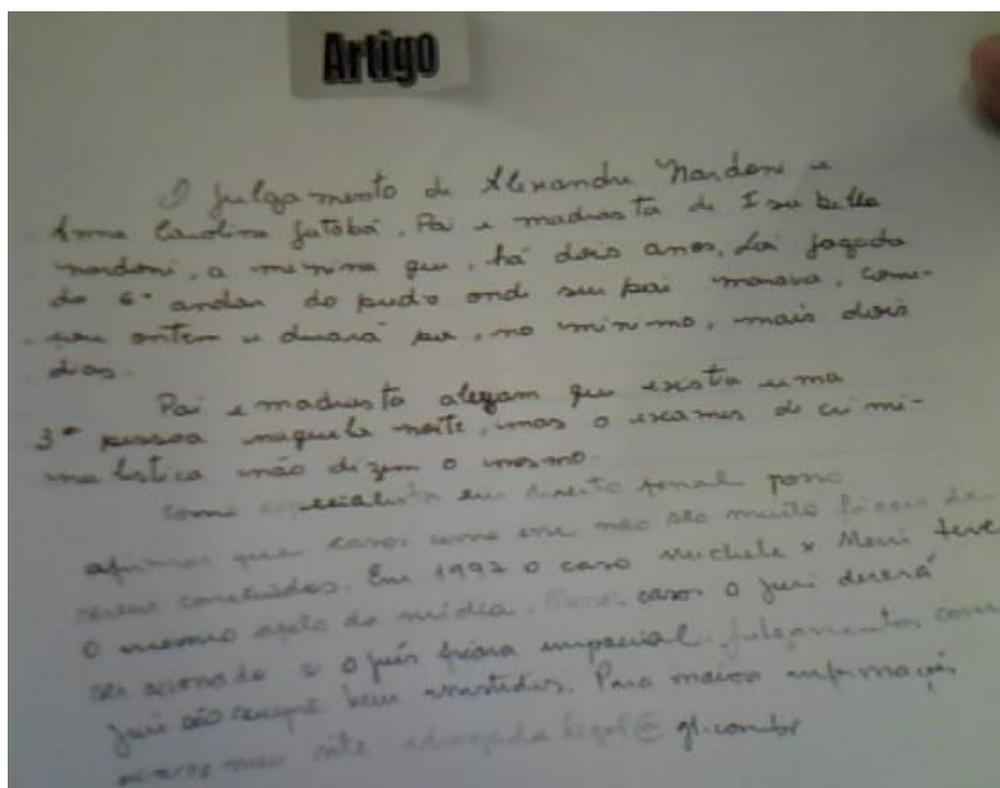


Figura 6.4 Texto depois da intervenção dialogada

6.1.5 Conclusão

O resultado desse estudo piloto foi considerado positivo, do total de oito projetos, foram produzidos 31 textos e dentre esses somente 10 precisaram de intervenção e todas elas com resultado positivo ao seu término. Outro aspecto interessante foi à eficiência dos resultados gerados pelo crivo de classificação dos textos, 20 textos foram classificados de acordo com a intenção de gênero produzido pelo aluno. Entre os 10 textos com resultado diferente da intenção do aluno, apenas 03 ficaram fora da interseção dos gêneros, mas sabendo a intenção de produção do aluno, facilitou o trabalho de intervenção.

Texto	Classificação	Intenção	Intervenção
1.	Manchete	Manchete	Não
2.	Notícia e Artigo	Artigo	positiva
3.	Resenha	Resenha	Não
4.	Resenha	Manchete	Não
5.	Receita	Horóscopo	positiva
6.	Notícia e Artigo	Notícia	positiva
7.	Receita	Receita	Não
8.	Horóscopo	Horóscopo	Não
9.	Receita e Horóscopo	Horóscopo	positiva
10.	Receita	Receita	Não
11.	Horóscopo	Horóscopo	Não
12.	Receita	Receita	Não
13.	Receita	Receita	Não
14.	Notícia e Reportagem	Crônica	positiva
15.	Receita	Receita	Não
16.	Quadrinhos	Quadrinhos	Não
17.	Notícia	Manchete	positiva
18.	Notícia e manchete	Manchete	positiva
19.	Notícia e manchete	Manchete	positiva
20.	Notícia	Notícia	Não
21.	Receita	Receita	Não
22.	Manchete	Manchete	Não
23.	Notícia	Notícia	Não
24.	Artigo	Crônica	positiva
25.	Notícia	Notícia	Não
26.	Notícia	Notícia	Não
27.	Notícia	Notícia	Não
28.	Notícia e Artigo	Artigo	positiva
29.	Notícia	Notícia	Não
30.	Notícia	Notícia	Não
31.	Notícia	Notícia	Não

Quadro 6.17: Resultado da Intervenção

Com esses resultados, foi possível notar a eficiência da intervenção e a curva de aprendizagem do aluno, 100 % de aprendizagem visto que todos os textos que foram reescritos atingiram o resultado esperado. Com isso, foi constatada a eficiência da proposta de se trabalhar com a diversidade textual através da intervenção dialogada e um crivo classificatório, impulsionando-nos ainda mais para a construção da ferramenta de classificação automática de textos e desenvolvimento

do Jogo do Jornal, a fim de apressar o processo de coleta, classificação e intervenção, possibilitando que o letramento se dê ao maior número possível de pessoas.

6.2 Estudo de caso Final

Esse segundo estudo de caso trata do experimento de pesquisa final cujos objetivos principais foram:

- u) verificar a eficácia da ferramenta em analisar automaticamente textos de alunos do ensino fundamental e médio;
- v) verificar a relevância dos critérios selecionados para essa função;
- w) verificar se há algum efeito na aprendizagem da elaboração desses gêneros por esses alunos;
- x) verificar a eficiência das recomendações automáticas que lhes foram fornecidas.

A fim de que os objetivos 1 e 2 fossem alcançados, foi observado pelo pesquisador o texto do aluno e o resultado da ferramenta, estando de acordo com as duas classificações, foi validada a funcionalidade da ferramenta; não estando de acordo com as duas classificações, foi feito um estudo da possibilidade de os critérios terem sido responsáveis pelo resultado negativo da classificação. Esses critérios foram destacados e novamente analisados. Para a verificação do objetivo 4, foram estudados pelo pesquisador os textos dos alunos que foram ou não tiveram foram classificados corretamente após várias tentativas, pois assim verificamos se as recomendações eram ou não eficientes no que se propunham. Para a validação do objetivo 3, os professores de português responderam a um questionário após o experimento, sobre suas impressões e expectativas da possibilidade e dos efeitos de poder contar com uma ferramenta como esta em suas práticas.

Trata-se de uma pesquisa semiquantitativa: que envolveu manipulação de variáveis e tratamento experimental; que se deu a partir do estudo do número de tentativas e erro por parte daqueles que não alcançaram inicialmente a probabilidade esperada, pois para o texto do aluno ser classificado automaticamente dentro do gênero esperado, esse precisa conter um número significativo dos critérios previamente estabelecidos para a classificação desse gênero. Assim foi disponibilizado automaticamente um número de recomendações associadas aos critérios que não foram alcançados na primeira tentativa, porém as possibilidades de tentativas foram limitadas a 2x o número de critérios não alcançados, permitindo-se assim, que o aluno possa errar mais de uma vez.

A ferramenta do jogo: foi o instrumento principal na coleta dos dados e registro das tentativas. O pesquisador: foi o instrumento principal na análise dos dados. O objeto de estudo foi a classificação automática de textos de alunos do ensino fundamental e médio. A Amostra foram alunos do segundo ciclo do ensino Fundamental e Médio de escolas com classificação acima e abaixo de 25% da classificação do IDEB. Ou seja, duas escolas do segundo ciclo do ensino Fundamental e duas escolas do ensino Médio.

Os sujeitos foram alunos do segundo ciclo do ensino Fundamental e do ensino Médio e professores de Língua Portuguesa. A ferramenta utilizada foi o *software* livre *online* e questionário de

avaliação. As técnicas utilizadas foram a observação do participante e levantamento de dados arquivados. Os instrumentos utilizados para coletar os dados foram: o arquivo de log. de registro da ferramenta e o questionário.

A descrição das situações propostas para a coleta de dados.

- Primeiro e segundo dia: Os alunos do segundo ciclo do ensino Fundamental e médio logaram-se na ferramenta e produziram textos. Eles primeiro receberam informação sobre os gêneros correspondentes às caixas de texto do jornal *online* e depois escreveram um de cada vez, enviando-os para análise automática. Obtendo resultado positivo (verificação do objetivo 1 e 2), seguiram para a produção de outro gênero; quando contrário, seguiram as recomendações automáticas geradas em sua tela e reescreveram seus textos, acabados as correções, reenviaram para análise novamente até a obtenção do resultado esperado ou atingido o número limite de tentativas. (verificação do objetivo 4). O professor respondeu ao questionário *online* a partir de uma semana do experimento colocando suas impressões. (verificação do objetivo 3)
- Dias finais da pesquisa: foi feita uma entrevista semiestruturada com os professores de Língua Portuguesa do segundo ciclo do ensino Fundamental e Médio, que responderam ao questionário *online* a fim de analisar melhor suas impressões acerca do experimento.

6.2.1 Eficiência da Ferramenta

Os alunos do segundo ciclo do ensino Fundamental e médio se logaram na ferramenta e produziram textos. Eles primeiro receberam informação sobre os gêneros correspondentes às caixas de texto do jornal *online*, de maneira dialogada e negociada, pois a tentativa de trabalhar com o *design* metacognitivo, ou seja, levar os alunos à interpretação automática de quais gêneros estavam sendo pedidos que produzissem, única e exclusivamente pelo *design* e escolha de ícones da interface do jogo não se mostrou muito eficiente, pois os alunos muitas vezes demonstraram dúvida e até mesmo o não conhecimento dos gêneros cobrados. Assim, a alternativa encontrada pelo pesquisador presente foi a de acionar o conhecimento dos alunos sobre os gêneros requisitados, instruí-los sobre o local onde deveriam ser impressos, e até mesmo quando necessário explicar minimamente seu conteúdo. Os alunos escreveram um texto de cada vez, enviando-os para análise automática. Obtendo resultado positivo (verificação do objetivo 1 e 2), seguiram para a produção de outro gênero quando contrário, seguiram as recomendações automáticas geradas em sua tela e reescreveram seus textos.

Foram duas escolas de Ensino médio, uma com IDEB2010 de 4.3 e outra escola com IDEB2010 de 7.5 um total de 16 alunos 8 de uma e 8 de outra. A quantidade de alunos foi escolhida por circunstâncias da capacidade do laboratório de informática das escolas, pois o experimento foi realizado em 50 minutos com um computador por aluno. Durante o experimento foi verificado a ineficiência dos critérios do gênero manchete, pois as recomendações eram acionadas independente de os textos escritos estarem de acordo com as características de estrutura lingüística desse gênero.

Assim de 16 textos escritos as manchetes de todos os alunos foram reprovadas, o mesmo aconteceu com o gênero artigo. Sendo que quando analisados manualmente isso não era verdade. Vale lembrar ainda que em 60 minutos nem todos os alunos conseguiram terminar todos os textos. Num universo de sete gêneros uma média de 5 gêneros no total foram escritos, uma maior frequência para a notícia e a manchete.

Já para o ensino Fundamental I, 8 série foram selecionadas duas escolas uma co IDEB2010 4.3 e outra com IDEB2010 4.7. Para o ensino fundamental II, 4/5 série foram selecionados escolas com IDEB2010 3.9 e IDEB2010 4.4.

COL BRIGADEIRO NEWTON BRAGA	manchete	Noticia	Artigo	Resenha	Receita	Crônica	horoscopo
Aluno 1							
Aluno 2							
Aluno 3							
Aluno 4							
Aluno 5							
Aluno 6							
Aluno 7							

Quadro 6.17: Gêneros produzidos pelo E.M.

I E GOV ROBERTO SILVEIRA	manchete	Noticia	Artigo	Resenha	Receita	Crônica	horoscopo
Aluno 1							
Aluno 2							
Aluno 3							
Aluno 4							
Aluno 5							
Aluno 6							
Aluno 7							

Quadro 6.18: Gêneros produzidos pelo E.M.

I E GOV ROBERTO SILVEIRA	manchete	Noticia	Artigo	Resenha	Receita	Crônica	horoscopo
Aluno 1							
Aluno 2							
Aluno 3							
Aluno 4							
Aluno 5							
Aluno 6							
Aluno 7							

Quadro 6.19: Gêneros produzidos pelo E.F.I.

E M MAURO DE CASTRO	manchete	Noticia	Artigo	Resenha	Receita	Crônica	horoscopo
Aluno 1							
Aluno 2							
Aluno 3							

Aluno 4							
Aluno 5							
Aluno 6							
Aluno 7							

Quadro 6.20: Gêneros produzidos pelo E.F.I.

I E GOV ROBERTO SILVEIRA	manchete	Noticia	Artigo	Resenha	Receita	Crônica	horoscopo
Aluno 1							
Aluno 2							
Aluno 3							
Aluno 4							
Aluno 5							
Aluno 6							
Aluno 7							

Quadro 6.21: Gêneros produzidos pelo E.F.II

E M MAURO DE CASTRO	manchete	Noticia	Artigo	Resenha	Receita	Crônica	horoscopo
Aluno 1							
Aluno 2							
Aluno 3							
Aluno 4							
Aluno 5							
Aluno 6							
Aluno 7							

Quadro 6.22: Gêneros produzidos pelo E.F. II

6.2.2 Relevância dos critérios

Os alunos escreveram um texto de cada vez, enviando-os para análise automática. Obtendo resultado positivo (verificação do objetivo 1e 2), seguiram para a produção de outro gênero; quando contrário, seguiram as recomendações automáticas geradas em sua tela e reescreveram seus textos. Como se pode verificar nos quadros acima, a obtenção do resultado positivo foi verificado em somente alguns textos, imediatamente, possibilitando que os alunos seguissem para a produção dos outros, porém foi também verificado que a não aprovação automática do gênero manchete encontra-se equivocada, pois os critérios estão muito rígidos e não deixam textos corretos serem aprovados. Alguns alunos não conseguiram a aprovação imediata de muitos gêneros pela falta de compreensão dos critérios pedidos, visto que eles foram escamoteados por perguntas de reflexão (que podem ser verificadas no Capítulo 3) em vez de itens que deveriam ser inclusos do texto, de forma clara e objetiva.

Nas tabelas abaixo, o resultado do experimento foi tabelado da seguinte maneira: quando o aluno tem seu texto aprovado, é informado o número de tentativas feitas para esta aprovação (T); do contrário, é assinalada a reprovação do texto do aluno, através de um (x) ao lado do número de tentativas feitas (T). Sendo que 0T, significa ter acertado de primeira.

COL BRIGADEIRO NEWTON BRAGA	Manchete	Notícia	Artigo	Resenha	Receita	Crônica	Horóscopo
Aluno 1	X 6T	1T	X 8T	5T	2T	X 8T	4T
Aluno 2	X 6T	3T	7T	5T	1T	X 8T	
Aluno 3	X 6T	1T			0T		3T
Aluno 4	X 6T	3T		X 8T	2T		4T
Aluno 5	X 6T	1T		X 8T	2T		4T
Aluno 6	X 6T	1T	X 8T	7T	2T	7T	1T
Aluno 7	X 6T	1T	X 8T			0T	1T

Quadro 6.23: Gêneros produzidos pelo E.M.

I E GOV ROBERTO SILVEIRA	Manchete	Notícia	Artigo	Resenha	Receita	Crônica	Horóscopo
Aluno 1	X 6T	1T	7T	X 8T	1T	2T	X 6T
Aluno 2	X 6T	1T	7T	X 8T	2T	5T	X 6T
Aluno 3	X 6T	1T			4T		4T
Aluno 4	X 6T	3T	X 8T	5T	0T	0T	4T
Aluno 5	X 6T	3T		5T	2T		5T
Aluno 6	X 6T	3T	X 8T	7T	1T	X 8T	2T
Aluno 7	X 6T	1T	X 8T	3T	1T	X 8T	3T

Quadro 6.24: Gêneros produzidos pelo E.M.

I E GOV ROBERTO SILVEIRA	Manchete	Notícia	Artigo	Resenha	Receita	Crônica	Horóscopo
Aluno 1	X 6T	3T	X 8T	7T	2T	4T	5T
Aluno 2	X 6T	6T	X 8T	X 8T	2T	X 8T	2T
Aluno 3	X 6T	4T	X 8T	X 8T	2T	X 8T	5T
Aluno 4	X 6T	1T	7T	X 8T	4T	7T	0T
Aluno 5	X 6T	1T	X 8T	X 8T	2T	7T	0T
Aluno 6	X 6T	3T	X 8T	X 8T	2T	7T	X 6T
Aluno 7	X 6T	X 6T	X 8T	X 8T	4T	X 8T	X 6T

Quadro 6.25: Tabelas com os gêneros produzidos pelo E.F. I..

E M MAURO DE CASTRO	Manchete	Notícia	Artigo	Resenha	Receita	Crônica	Horóscopo
Aluno 1	X 6T	6T	X 8T	X 8T	3T	4T	5T
Aluno 2	X 6T	4T	X 8T	6T	4T	4T	4T
Aluno 3	X 6T	1T			1T		0T
Aluno 4	X 6T	1T	X 8T	X 8T	2T	X 8T	2T
Aluno 5	X 6T	1T		X 8T	4T		X 6T
Aluno 6	X 6T	6T	7T	X 8T	1T	X 8T	X 6T
Aluno 7	X 6T	X 6T	7T	6T	4T	X 8T	X 6T

Quadro 6.26: gêneros produzidos pelo E.F.I.

I E GOV ROBERTO SILVEIRA	Manchete	Notícia	Artigo	Resenha	Receita	Crônica	Horóscopo
Aluno 1	X 6T	1T	7T	X 8T	4T	X 8T	X 6T
Aluno 2	X 6T	4T	7T	7T	2T	X 8T	X 6T
Aluno 3	X 6T	4T			1T		X 6T

Aluno 4	X 6T	X 6T	X 8T	7T	3T	X 8T	0T
Aluno 5	X 6T	4T		X 8T	2T		4T
Aluno 6	X 6T	1T			X 6T	X 8T	3T
Aluno 7	X 6T	X 6T			4T	4T	2T

Quadro 6.27: gêneros produzidos pelo E.F. II

E M MAURO DE CASTRO	Manchete	Notícia	Artigo	Resenha	Receita	Crônica	Horóscopo
Aluno 1	X 6T	4T	X 8T	X 8T	4T	X 8T	2T
Aluno 2	X 6T	4T	X 8T	7T	X 6T	X 8T	X 6T
Aluno 3	X 6T	X 6T			X 6T		X 6T
Aluno 4	X 6T	X 6T	X 8T	X 8T	4T	X 8T	2T
Aluno 5	X 6T	X 6T		X 8T	2T		3T
Aluno 6	X 6T	5T	X 8T		2T	X 8T	3T
Aluno 7	X 6T	X 6T	X 8T	X 8T	4T	X 8T	X 6T

Quadro 6.28: gêneros produzidos pelo E.F. II

A tabela aponta o número de tentativas dos alunos que não alcançaram inicialmente a probabilidade esperada, pois para o texto do aluno ser classificado automaticamente dentro do gênero esperado, esse precisa conter os critérios previamente estabelecidos para a classificação desse gênero. Assim foi disponibilizado automaticamente um número de recomendações associadas aos critérios que não forem alcançados na primeira tentativa, porém as possibilidades de tentativas estão limitadas a 2x o número de critérios não alcançados, permitindo-se assim que o aluno possa errar mais de uma vez. Quando na tabela o número vem acompanhado de um X significa que o total de tentativas possíveis foi alcançado e por isso o texto do aluno deverá ser analisado para verificar a eficiência do analisador automático em não classificar o texto do aluno como o gênero em questão.

A análise dos resultados reforça uma das conclusões obtidas pelo pesquisador ao longo da pesquisa, sobre a rigidez dos critérios do gênero manchete, como também indica a ineficiência dos critérios dos gêneros crônica e artigo. Como pode ser observado no Quadro 6.30.

	Total de texto	Aprovados	Reprovados	Aprovados por releitura (entre os reprovados)	Margem de erro
Manchete	42	0	42	40	95%
Notícia	42	34	08	02	25%
Artigo	29	08	21	12	57%
Resenha	33	13	20	05	25%
Receita	41	38	03	0	0
Crônica	31	12	19	09	47%
Horóscopo	41	28	13	03	23%

Quadro 6.29: Gêneros produzidos pelo E.F.I.

6.2.3 Efeito na Aprendizagem

Para a validação do objetivo 3, o professor respondeu ao questionário *online* a partir de uma semana do experimento, colocando suas impressões. O questionário respondido pelos professores

procura verificar algum ponto sobre o experimento e a importância da ferramenta que tenha ficado a desejar. Contudo, gostaríamos de deixar claro que os resultados apurados aqui não são estatisticamente significativos, e os argumentos apresentados são somente interpretativos afim de descrever uma tentativa de qualificação da ferramenta.

Nome:

Qualificação:

Responda as perguntas abaixo.

1) Consegui realizar o experimento dentro do tempo determinado?

sim

não

2) Os alunos tiveram dificuldades com o primeiro acesso a ferramenta, como a parte de login e senha?

sim

não

3) Seus alunos tiveram dificuldade para entender o gênero textual recomendado que escrevessem?

sim

não

4) Seus alunos tiveram dificuldades para entender o que estava faltando no texto com as recomendações dos critérios?

sim

não

5) Você teve dificuldades de entender o que estava faltando no texto de seus alunos com as recomendações dos critérios?

sim

não

6) Você discorda dos critérios apontados como essenciais para a classificação do gênero?

sim, todos

sim, alguns

não, nenhum.

7) Você acredita que a proposta dessa ferramentas pode ajudar seu trabalho em sala de aula?

sim

não

8) Você acredita que a proposta dessa ferramentas pode ajudar seus alunos fora de sala de aula?

sim

não

9) Você tem sugestões ou recomendações a serem feitas ao pesquisador ?

sim

não

10) Você gostou de participar dessa pesquisa?

sim

não

Figura 6.9: Questionário de avaliação do experimento.

Um total de 6 professores responderam o questionário, e a proporção das perguntas respondidas foi a seguinte:

Professores	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
01EM	sim	Não	Não	Não	não	não, nenhum	sim	sim	Não	sim
02EM	sim	Não	Não	Não	Não	não, nenhum	sim	sim	Não	sim
03EF II	sim	Não	Não	Não	Não	não, nenhum	sim	sim	Não	sim
04EF II	sim	sim	sim	Não	Não	sim, alguns	sim	sim	Não	sim
05EF I	Não	Não	sim	sim	sim	Sim, todos	sim	sim	Não	sim
06EF I	Não	Sim	sim	Não	Não	não, nenhum	sim	sim	Não	sim

Quadro 6.30: respostas do questionário.

Como as respostas para questão 09 foram sim unânime, então foi decidido não realizar a entrevista para verificar o efeito da aprendizagem, visto que é também unânime a opinião sobre a relevância da ferramenta dentro e fora de sala de aula. A resposta positiva dos alunos em relação ao experimento e o resultado positivo da classificação de gêneros que nunca haviam sido escritos anteriormente também é um forte indício da aprendizagem ocorrendo durante o experimento.

6.2.4 Eficiência das recomendações automáticas

Nossa hipótese de que se perante a mediação apropriada o aluno seria capaz de escrever seu texto de maneira correta, prova-se sensível a partir do experimento, como, por exemplo, a rigidez dos critérios do gênero manchete, que teve um total de 42 textos produzidos e 100% destes textos reprovados pelo sistema, sendo somente 2 reprovados realmente por releitura; os outros todos de acordo com a avaliação de manchete pelo professor. A margem de erro do sistema é de 95% para 4.75% de acerto. Para o gênero artigo foram produzidos um total de 29 textos, sendo que somente 8 foram aprovados pelo sistema e dos 21 artigos reprovados, 12 foram reprovados pela releitura do professor. A crônica teve um total de 31 textos produzidos, 19 aprovados e 12 reprovados, sendo que, destes, 9 foram reprovados por releitura. Para os outros gêneros, essa margem de erro foi bem reduzida. Conclui-se que a rigidez dos critérios influencia diretamente na ineficácia das recomendações, ou seja, é preciso repensar os critérios e consequentemente as perguntas de mediação, desses três gêneros.

6.2.5 Conclusão do estudo de caso Final

O resultado não foi como esperado. Todavia, o estudo de caso valida a ferramenta do jogo para alguns dos gêneros trabalhados e abre caminho para uma nova perspectiva de se trabalhar como

classificação automática de textos segundo o gênero, não somente pela contribuição estatística pura e simplesmente, mas pela forte influência da estrutura linguística desses gêneros. É preciso refazer as perguntas de mediação a fim de torná-las mais precisas, assim como também os critérios. O aprendizado pode ser verificado segundo o relato dos professores participantes da pesquisa sobre a intenção e vontade dos alunos em realizar o experimento, muitos alunos que nunca haviam entrado em contato com a tarefa de produzir texto antes do experimento, outros não conheciam todos os gêneros e, ao experimentarem, aprenderam algo, pois quase todos os alunos experimentaram produzir todos os gêneros. Foi muito gratificante ouvir desses alunos comentários como: "nossa, que legal isso!" "Vou escrever sobre nossa escola", "Posso usar o jogo em casa depois também?". "Acho que entendi o que é uma crônica." "Vocês podem voltar amanhã também."

Conclui-se o estudo de caso ciente das limitações da ferramenta, em relação à rigidez dos critérios e da ineficiência de algumas perguntas de mediação, entendeu-se também que esse resultado não inviabiliza a utilização da ferramenta só para os critérios que tiveram erro pequeno e de maneira supervisionada pelo professor, este que pode ser uma ferramenta fundamental para o desenvolvimento do jogo. Se o jogo for disponibilizado para professores utilizarem com seus alunos, registros novos poderão ser gravados e utilizados para melhorar o desempenho do Jogo, do melhor jeito colaborativo que uma ferramenta web pode oferecer.

7. CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

7.1 Resumo do Trabalho

No presente trabalho, foi apresentado um conjunto de técnicas e algoritmos para a solução de problemas relacionados com a classificação automática de textos segundo seus gêneros textuais. Foi modelada uma ferramenta que se propõe à auxiliar o trabalho do professor de português dentro e fora de sala de aula, buscou-se também um levantamento bibliográfico sobre o estado da arte da classificação automática de texto, das teorias linguísticas a cerca da temática da diversidade textual e o letramento.

A ferramenta recebeu a temática de jogo do jornal motivada pela escolha da técnica de abordagem psico-pedagógica metaprocessual da elaboração dirigida, representada através da intervenção automática da produção do aluno.

Para a construção do Objeto de Medida foram estudados diversos processos utilizados para descobrir conhecimento útil em uma coleção de documentos, como, por exemplo, o processo de Mineração de Texto, uma área multidisciplinar que incorpora técnicas de diversas áreas, como Recuperação de Informação, Aprendizado de Máquina, Estatística, Linguística Computacional, Extração de Informação, Visualização e especialmente Mineração de Dados.

O ambiente computacional configurado para a utilização do *software* produzido para essa dissertação envolve ferramentas de código aberto, de maneira a permitir futuras modificações e evoluções que se fizerem necessárias, *software* foi modelado por *extreme programming*.

Foram feitos dois estudos de caso. O primeiro se propõe a utilizar a técnica psicopedagógica da elaboração dirigida (Seminário, 1997). A possibilidade de se trabalhar com a diversidade textual e a psicopedagogia, realizado no primeiro ano dessa pesquisa, com dois grupos heterogêneos de amostras de alunos do ensino Fundamental, cujo objetivo era verificar a possibilidade de intervenção dialogada do professor em relação à produção textual desses alunos. Sob a hipótese de que, se possível, então, a construção de uma ferramenta a partir do crivo do especialista poderia ser feito.

O resultado desse estudo piloto foi considerado positivo, do total de oito projetos, foram produzidos 31 textos e dentre esses somente 10 precisaram de intervenção e todas elas com resultado positivo ao seu término.

O segundo estudo de caso trata do experimento de pesquisa final cujos objetivos principais foram:

- y) verificar a eficácia da ferramenta em analisar automaticamente textos de alunos do ensino fundamental e médio;
- z) verificar a relevância dos critérios selecionados para essa função;
- aa) verificar se há algum efeito na aprendizagem da elaboração desses gêneros por esses alunos;
- bb) verificar a eficiência das recomendações automáticas que lhes foram fornecidas.

A análise dos resultados reforça uma das conclusões obtidas pelo pesquisador ao longo da pesquisa, sobre a rigidez dos critérios do gênero manchete, como também indica a ineficiência dos critérios dos gêneros crônica e artigo.

O estudo de caso valida a ferramenta do jogo para alguns dos gêneros trabalhados e abre caminho para uma nova perspectiva de se trabalhar como classificação automática de textos segundo o gênero, não somente pela contribuição estatística pura e simplesmente, mas pela forte influência da estrutura linguística desses gêneros. É preciso refazer as perguntas de mediação a fim de torná-las mais precisas, assim como também os critérios.

Para a validação do objetivo 3, o professor respondeu ao questionário *online* a partir de uma semana do experimento, colocando suas impressões. O questionário respondido pelos professores procura verificar algum ponto sobre o experimento e a importância da ferramenta que tenha ficado a desejar. Assim, havendo ou não a necessidade de uma entrevista semiestruturada após o experimento com os professores de português envolvidos, busca suas impressões e expectativas da possibilidade de poder contar com uma ferramenta como esta em suas práticas. Assim, o aprendizado pode ser verificado segundo o relato dos professores participantes da pesquisa sobre a intenção e vontade dos alunos em realizar o experimento, muitos alunos que nunca haviam entrado em contato com a tarefa de produzir texto antes do experimento.

Concluimos esse trabalho cientes das limitações da ferramenta, em relação à rigidez dos critérios e da ineficiência de algumas perguntas de mediação, entendemos também que esse resultado não inviabiliza a utilização da ferramenta de maneira supervisionada pelo professor.

7.2. Contribuições da Dissertação

A metodologia aplicada atendeu razoavelmente aos objetivos propostos, pois o segundo estudo de caso poderia ter sido realizado em partes. Apesar de os testes e as avaliações apresentados não terem mostrado a eficiência e aplicabilidade da solução proposta para o problema de classificação automática de textos segundo seus gêneros textuais. A visualização dos resultados se mostrou bastante informativa e de grande valia para a interpretação dos mesmos.

8.3. Problemas Encontrados

Um dos principais problemas encontrados durante a fase de testes e tratado dos textos foi encontrar fontes fiéis aos gêneros estudados aqui advindos da internet e com formato compatível com o esperado.

Para a etapa de pré-processamento, desenvolveu-se uma ferramenta, com o objetivo de sistematizar todo o processo de preparação dos dados, desde o momento da coleta dos textos até a saída do arquivo pronto para ser executado pelos algoritmos de mineração de dados, muitas dificuldades foram vencidas em relação a programação desses estágios.

Na etapa de exploração dos dados foram constatadas evidências de critérios muitos fracos para classificação de determinados gêneros que em teoria foram considerados importantes como por exemplo o tempo verbal do futuro do indicativo para o gênero horóscopo.

Na etapa de classificação, os classificadores Orange e Naive Bayes tiveram um desempenho ruim, apresentando percentuais acima de 50% para a eficiência na classificação. O objetivo desta etapa foi constatar a distribuição dos textos pelos gêneros correspondentes, de acordo com a definição do especialista. No tratamento dos textos inicial foi observado que apresentaram um percentual maior de erro nas classificações com o classificador *Naive Bayes*. Assim, foi estudada uma melhor maneira de tratar os textos com o auxílio dos critérios linguísticos, através do *orange*.

O processo de calibração dos resultados, através dos critérios linguísticos descritos em expressões regulares, se mostrou eficiente somente após certos critérios serem definidos como mais frequentes para determinados gêneros, e receberem um peso Haidth maior a fim de se destacarem em meio aos outros. Para processar os documentos, optou-se pelo uso do *orange* e *naive bayes*.

As dificuldades na implementação do modelo, concentram-se nas fases de recomendação automática para a intervenção da máquina na produção do aluno. Tivemos dificuldade em implementar as perguntas livres dos critérios, comprometendo e restringindo assim sua funcionalidade. Pois como as perguntas de intervenção estão anexadas aos criterios que correspondem, podemos somente acioná-las quando o critério é omitido no texto.

Outro problema foi reduzir o número de critérios para cada gênero, visto que a interseção entre eles escamoteia sua classificação. Ao reduzirmos os critérios mais frequentes, resolvemos esse problema, porém perdemos muita informação sobre a diversidade textual relevantes para o efeito da aprendizagem.

8.4 Trabalhos futuros

A seguir são apresentadas as perspectivas para o prosseguimento dessa pesquisa para trabalhos futuros, tanto em nível teórico, como de desenvolvimento de novas aplicações e realização de novos estudos de caso

Para trabalhos futuros, propõem-se experimentos com novos critérios complementando os já existentes, e estendendo para além das características da superfície linguística do texto. Nesse trabalho nos restringimos a trabalhar somente com um dos parâmetros de caracterização tipológica sugerido por Travaglia (2007). Seria interessante tentar programar os critérios correspondentes aos outros parâmetros, visto que os critérios já foram organizados para cada um desses parâmetros como pode ser verificado no Quadro 3.4, do capítulo 3.

Por fim, aumentar o número de estudos de casos, com a inclusão de outros gêneros; melhorar a interface do sistema para agilizar o processo de entrada dos documentos; aumentar a

stoplist, para abranger mais termos da língua portuguesa por gênero, adicionar suporte a algoritmos de classificação e agrupamento pela ferramenta proposta e desenvolver uma ontologia dos gêneros.

REFERÊNCIAS

ADAM, J-M, (1993). *Les Textes: types et prototypes*. Paris: Nathan

AL-SHALABI,R.K,GHASSAN,G.,“Arabic Text Categorization Using Knn Algorithm. *Manaf*2004 - Disponível:<www.ijicis.net/Vol6_No1%20No_1.pdf >. Acessado em: 20 Out 2009.

BANDURA, A.; Azzi, R. G.; POLYDORO, S. *Teoria Social Cognitiva: conceitos básicos*. Porto Alegre: ArtMed, 2008, p.176.

BANDURA, A. Influence of models' reinforcement contingencies on the acquisition of imitative responses. *Journal of Personality and Social Psychology*, 1, 589-595, 1965, p.15

BANDURA, A. *Teoria Social Cognitiva*. SP. ARTMED.2008 p.78-89

BARROS, D. Motta, C.L.R.; Oliveira, C.E.T. *O Jogo do Jornal: construindo novas estratégias de Letramento – Anais do XX Simpósio Brasileiro de Informática na Educação*. Setembro de 2010.

BAKHTIN, M.; VOLOCHINOV, V.N.. *Marxismo e filosofia de linguagem*. São Paulo: Hucitec. [1929]1998, p.34-56

BAKHTIN, M, Os gêneros do discurso. In Bakhtin, M. *Estética da criação verbal*. São Paulo: Martins Fontes, [1979]1992, p.277-326

BEZERRA, M. A. (Org.) *O livro didático de português: múltiplos olhares*. Rio de Janeiro: Lucerna, 2001. p. 33-45.

BIBER, D. *Variation Across Speech and Writing*. Cambridge: Cambridge University Press,1998 p.2-5

_____, *Spoken and written textual dimensions in English: Resolving the contradictory findings*. *Language*, 1986, 62(2):384–413.

_____, *The multidimensional approach to linguistic analyses of genre variation: An overview of methodology and finding*. *Computers in the Humanities*, 1992, 26(5–6):331–347.

_____, *Dimensions of Register Variation: A Cross-Linguistic Comparison*. Cambridge University Press, Cambridge, England. 1995, p.58-90

BRONCKART, J-P. (1999). *Atividade de linguagem, textos e discursos. Por um interacionismo sócio-discursivo*. Trad. A. R Machado e P. Cunha. São Paulo: EDUC

BRENTAN et al. "Web-Specific Genre Visualization", *Proc. of the 30th Hawaii International Conference on System Science*, Jan 1997.

CASTRO, L Estratégias para Letramento: Ensino Médio, lingüística aplicada. Dissertação de Mestrado, Pontifícia Universidade Católica do Rio de Janeiro. Defendida em maio de 2009

COLELLO, S. M. Alfabetização e letramento: repensando o ensino da língua escrita. Videtur, Porto/Portugal, v. 29, 2004, p. 43-52, 2004. Disponível em:
<http://www.hottopos.com/videtur29/silvia.htm> (p. 1-15). Acesso em: out. 2009.

COLE, R. M. *Clustering with Genetic Algorithms*, M. Sc., Department of Computer Science, University of Western Australia, Australia. 1998, p.50-56.

DOLZ, J. & SCHNEUWLY, B. *Pour un enseignement de l'oral. Initiation au genres formels à l'école*. Paris: ESF éditeur, 1998, p.45-67.

_____. Gêneros orais e escritos na escola. Campinas: Mercado de Letras, 2004, p45-90.

DÖRRE, J., P. GERSTL, & R. SEIFFERT. Text mining: finding nuggets in mountains of textual data. In: *KDD '99: Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, NY, USA, 1999, p.4-6.

LEWIS, D; RINGUETTE, M. "Comparison of two learning algorithms for text categorization," *Proc. Of the 3rd Annual Symposium on Document Analysis and Information Retrieval*, 1994. p.1-3.

DORNAN, A. "Lesson 188: Bayesian Spam Filtering". *Network Magazine*, Manhasset, NY, Mar.2004, p.64.

DUMAIS, S., PLATT, J., HECKERMAN, D., and SAHAMI, M., "Inductive learning algorithms and representations for text categorization". In: *Proceedings 83 of the 7th 1998 ACM international conference on information and knowledge management*, 1998, p.148-155,.

DEWE et al. "Assembling a Balanced Corpus from the Internet", 11th Nordic Conference of Computational Linguistics, pages 100-107, Copenhagen, 1998.

ELLE, revista Ano 2, nº 10. São Paulo: Ed. Abril, outubro 1989:p.209

FELDMAN, R. & H. HIRSH. Exploiting background information in knowledge discovery from text. *Journal of Intelligence Information Systems*, 1997, p.5-7

GIBBS, M , Math to fight spam. *Network World*, Southborough. MA, 2003, p.30.

HERRERA, J. R. A Educação e exclusão na América Latina. *Reformas em tempos de globalização*. Brasília: Universa, 2000, p.50-56

H. J. Oh, MYAENG S. H., and LEE, M., "A practical hypertext categorization method using links and incrementally available class information", *Proc. of the 23rd ACM SIGIR Conference*, , Athens, Greece, 2000, pages 264-271.

IBGE (2001), IBGE, Censo Demográfico - 2001.

IPEA – TD (2003) *Texto para discussão número 1000: Edição especial – Disponível em:* <
http://www.ipea.gov.br/portal/images/stories/PDFs/TDs/td_1000.pdf, acessado em Nov 2010.

JAIN, A.K e DUBES, R.C. *Algorithms for Clustering Data*, Prentice Hall, 1988, p.56.

JOLIBERT, J. *Formando Crianças Produtoras de Textos*. Porto Alegre: Artes Médicas, 1994, p.45-78

LOPES, M. C. S. *Mineração de dados textuais utilizando técnicas de clustering para o idioma português*. Tese de D.Sc., COPPE/UFRJ, Rio de Janeiro, defendido em 2004.

KARLGREN, J., “*Stylistic Variation in an Information Retrieval Experiment*”, Proc. of the 2nd International Conference on New Methods in Language Processing-NeMLaP.1996, p.45-67

_____, J., DEWE, J., BRETAN, I, DEWE, J. HALLBERG, A, WOLKERT, W, (1998) “*Iterative Information Retrieval Using Fast Clustering and Usage-Specific Genres*”, 8th DELOS Workshop on User Interfaces in Digital Libraries, , 1998, p. 85-92

_____, J., D. CUTTING, “*Recognizing Text Genres with Simple Metrics Using Discriminant Analysis*”, Proc. of COLING94, Kyoto, 1994, p..5-9.

_____, J., DEWE, J., BRETAN, I. “*Assembling a Balanced Corpus from the Internet*”, 11th Nordic Conference of Computational Linguistics, Copenhagen, 1998, p. 100-107.

KERN, R.. *Literacy and language teaching*. Oxford: Oxford University Press, 2000, p. 16.

KESSLER, B, NUNBERG, G. SCHOTZE, H, “*Automatic Detection of Text Genre*”, ACL’1997, pages 32-38,.

KOCH, I, *Desvendando os segredos do texto*. São Paulo: Cortez, 2002, p.56-88.

KOCH, I. & L. FÁVERO.. *Contribuição a uma tipologia textual*. Letras e Letras, *Uberlândia*, 3(1): 1987, p.3-10.

_____. *Linguística textual: introdução*. São Paulo: Cortez, 1988, p.56-79.

MARTINS, C.A. *Uma abordagem para pré-processamento de dados textuais em algoritmos de aprendizado*. Tese de Doutorado (ICMC-USP), São Paulo, 2003, p.10-25.

MARCUSCHI, L.A.. *Produção Textual, análise de gêneros e compreensão*. São Paulo: Parábola Editorial. 2008, p.12-24.

_____, L A, “*Produção textual, análise de gênero e compreensão*.” Parábola, 2008, p.34-45.

MARQUES, C. V. M; Motta, C.L. R; Oliveira, C. E. T.; Vrabl, S. D. P.; Lapolli, F., Pereira, A. P. M; Daflon, L. (2009a). *Avaliação de Crianças Deficientes Visuais através de jogos neuropedagógicos*. SCA 2009 - *Simpósio de Computação Aplicada*, Passo Fundo, Rio Grande do Sul.

_____, C.V.M, C.V. ; Oliveira, C. E. T. ; Motta, C. (Org.). *A revolução cognitiva; um estudo sobre a teoria de Franco Lo Presti* Seminário. Rio de Janeiro: PPGI/IM/NCE, (2009b). (Relatório Técnico, 04/09)

_____, C. V. M. (2009). *Laboratório de neuropsicologia cognitiva-projeto geral: avaliação de crianças deficientes visuais*. Rio de Janeiro: NCE/UFRJ, 10 p. (Relatório Técnico, 02/09).

MEURER, J.L.; BONINI, A. & MOTTA-ROTH, D. (org.) (2005). *Gêneros: teorias, métodos, debates*. São Paulo: Parábola MORGADO, J.C.Jr.; *Modelo Computacional para mineração de texto e análise de questões de concurso*, UFRJ, 2008.

MORAES, E. (2006) A Formação Cidadã do Jornalista no Brasil: um estudo de caso da formação do jornalista na USP, *Universidade de São Paulo, USP, Brasil*.

MOORE, R., Lopes, J., 1999. Paper templates. In *TEMPLATE'06, 1st International Conference on Template Production*. SciTePress.

ORANGE website <<http://orange.biolab.si/>> acessado em fevereiro de 2010.

PARÂMETROS CURRICULARES NACIONAIS (1998). – 3º. e 4º. ciclos do ensino fundamental: língua portuguesa Brasília, Secretaria de Educação Fundamental MEC.

PARÂMETROS CURRICULARES NACIONAIS (1999). – ensino médio: linguagens Brasília, Secretaria de Educação Fundamental MEC.

PCN, - *Parâmetros curriculares nacionais: língua portuguesa*. Brasília: MEC/SEF. 1997,p.41

PARKER, D & J: *O Grande Livro da Astrologia*, Círculo do Livro S.A, São Paulo, 1971, (p. 16).

Portal MEC.gov “*Campanha de incentive ao letramento*”.> Acessado em outubro 2009.<

STAMATATOS E., FAKOTAKIS, N., KOKKINAKIS, G., “*Text Genre Detection Using Common Word Frequencies*”, Proc. of the 18th International Conference on COLING2000, 2000.

SEBASTIAN, “A Tutorial on Automated Text Categorization”. In: Proceedings of ASAI-99, 1st Argentinian Symposium on Artificial Intelligence, pp.7-35, Buenos Aires, AR.

SILVA, C. F (2008), 2004. *Uso de Informações Lingüísticas na etapa de pré-processamento em Mineração de Textos*. Dissertação de M.Sc. Universidade do Vale do Rio dos Sinos, Programa Interdisciplinar de pós-graduação em Computação aplicada, São Leopoldo – Rio Grande do Sul.

SEBASTIANI, F. “Machine learning in automated text categorization”. *ACM Computing Surveys*, vol. 34, no. 1, , 2002. p.1-47

Seminário, F. L. P. Novos rumos na Psicologia e na Pedagogia: metacognição: uma nova opção. *Arquivos Brasileiros de Psicologia*, Rio de Janeiro, v. 49, n. 3, 2007, p. 5-22, jul./set.

SEBASTIANI, F. “Machine learning in automated text categorization”. *ACM Computing Surveys*, vol. 34, no. 1, pp.1-47, 2002.

SMITH, J. *The book*, The publishing company. London, 2nd edition.1998,p.34-37.

SOARES, Magda. *Letramento: um tema em três gêneros*. 2ed. Belo Horizonte: Autêntica,2000.

SWALES, J. M. . *Genre Analysis. English in Academic and Research Settings*. Cambridge: Cambridge University Press 1990, p.33.

TAN, A., "Text mining: The state of the art and the challenges". In: Proceedings of the Pacific Asia Conf on Knowledge Discovery and Data Mining PAKDD'99 workshop on Knowledge Discovery from Advanced Databases, 1999, p. 65–70.

Travaglia, L.C. A caracterização de categorias de textos: tipos, gêneros e espécies. Alfa: Revista de Lingüística, v. 51, 2007, p. 39-79, 2007. ISSN/ISBN: 19815794.

_____. Um estudo textual-discursivo do verbo no português. 1991. 330, 124f. Tese (Doutorado em Lingüística) – Instituto de Estudos da Linguagem, Universidade Estadual de Campinas, Campinas, 1991.

_____. Gramática: ensino plural. São Paulo: Cortez, 2003, p.34-45.

YANG, Y; PEDERSON, J, "A Comparative Study on Feature Selection in Text Categorization". In: *Proceedings of 14th International Conference on Machine Learning*, San Francisco, USA, 1997, p. 412-420.

Y. Yang and X. Liu, "A re-examination of text categorization methods," Proc. Of the 22nd ACM SIGIR Conference, 1999, p.4-10.

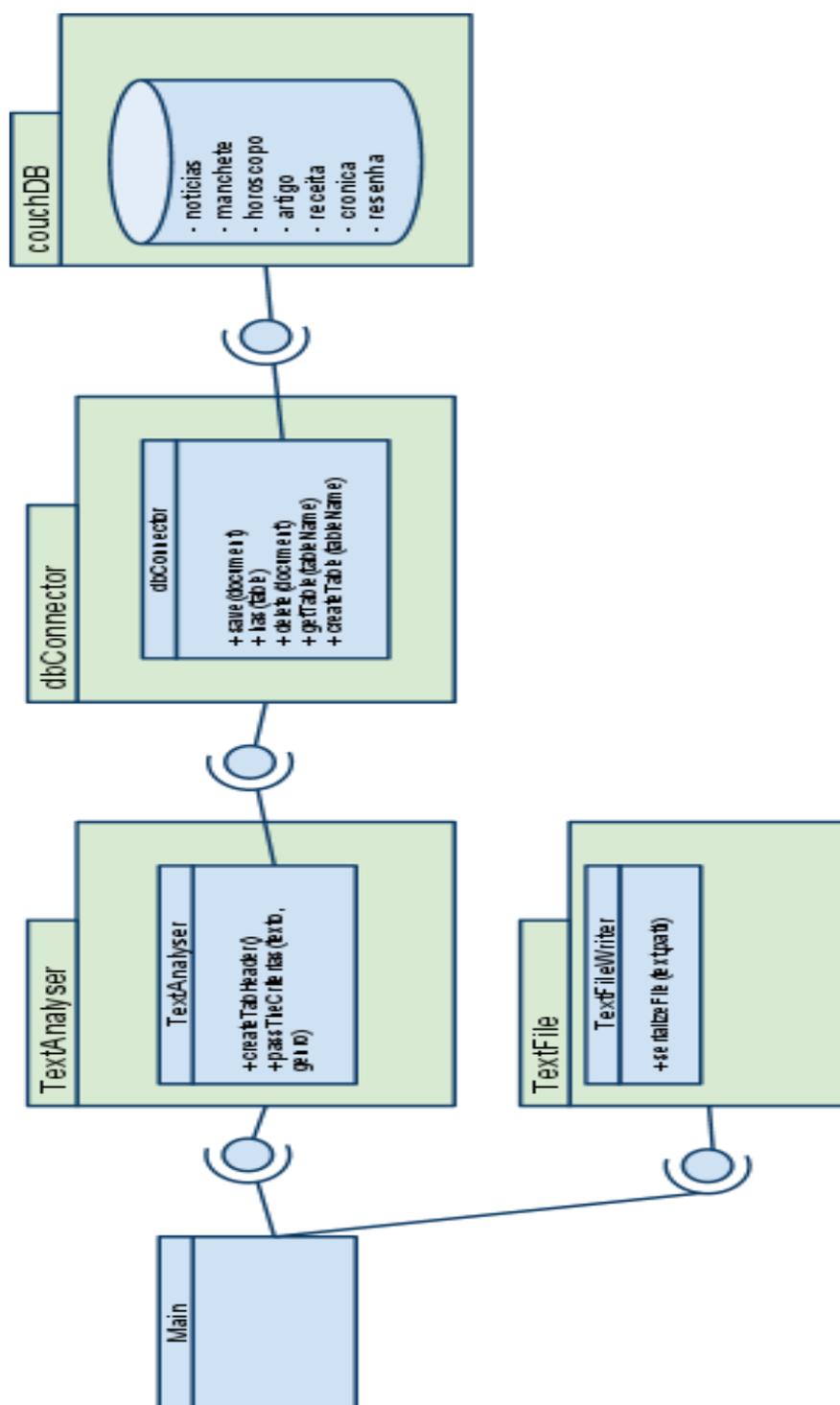
Yong-Bae Lee and Sung Hyon Myaeng, "Text Genre Classification with Genre-Revealing and Subject-Revealing Features," Proceedings of the 25th ACM SIGIR Conference, 2004, p.145-150, Tampere, Finland.

WIVES, L. K, Um estudo sobre Agrupamento de documentos textuais em Processamento de Informações não estruturadas usando técnicas de Clustering Dissertação de M.Sc., Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2002, p.3-12.

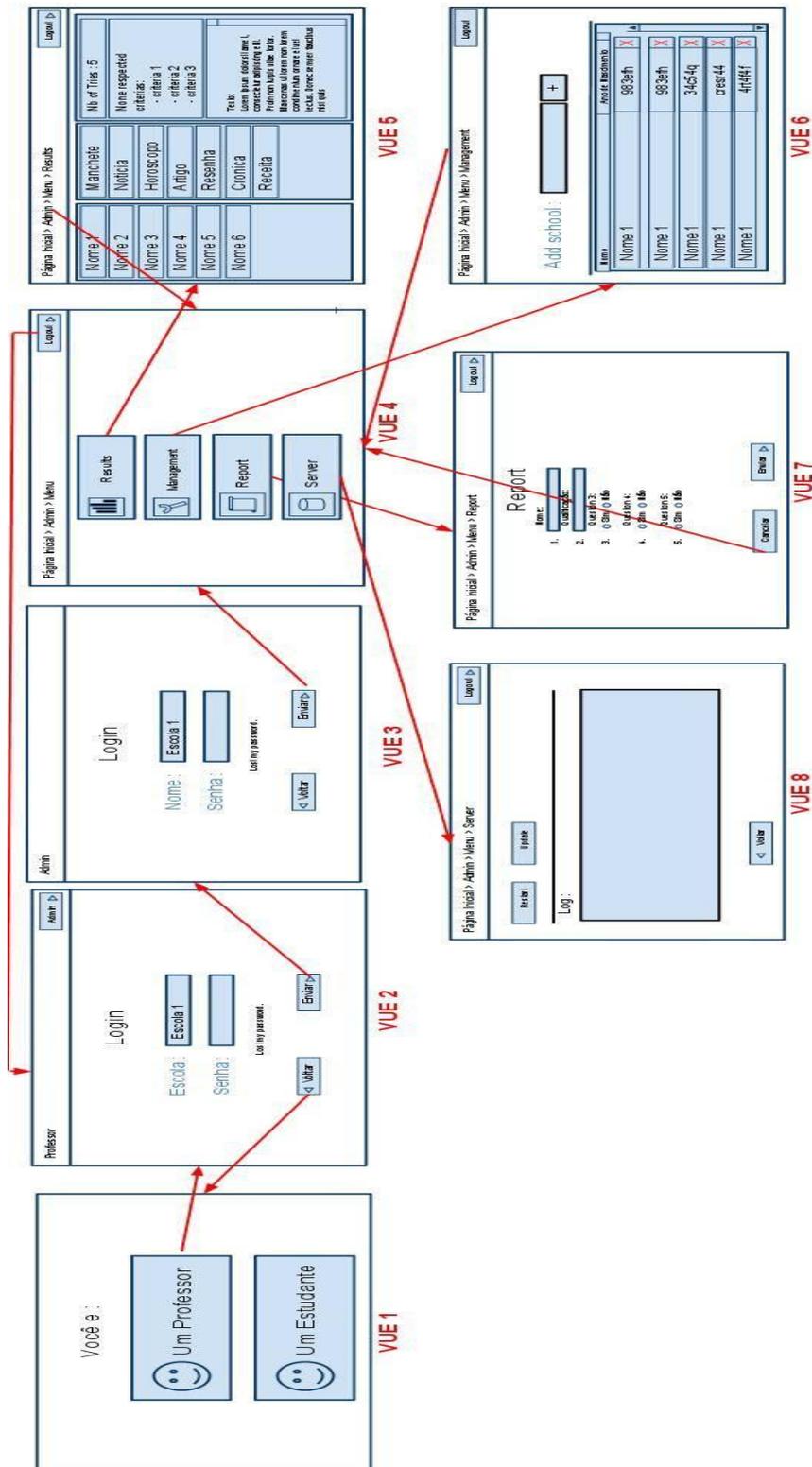
WOLKERT, Jussi Karlgren, "Web-Specific Genre Visualization", Proc. of the 30th Hawaii International Conference on System Science, Jan 1997.

ANEXOS

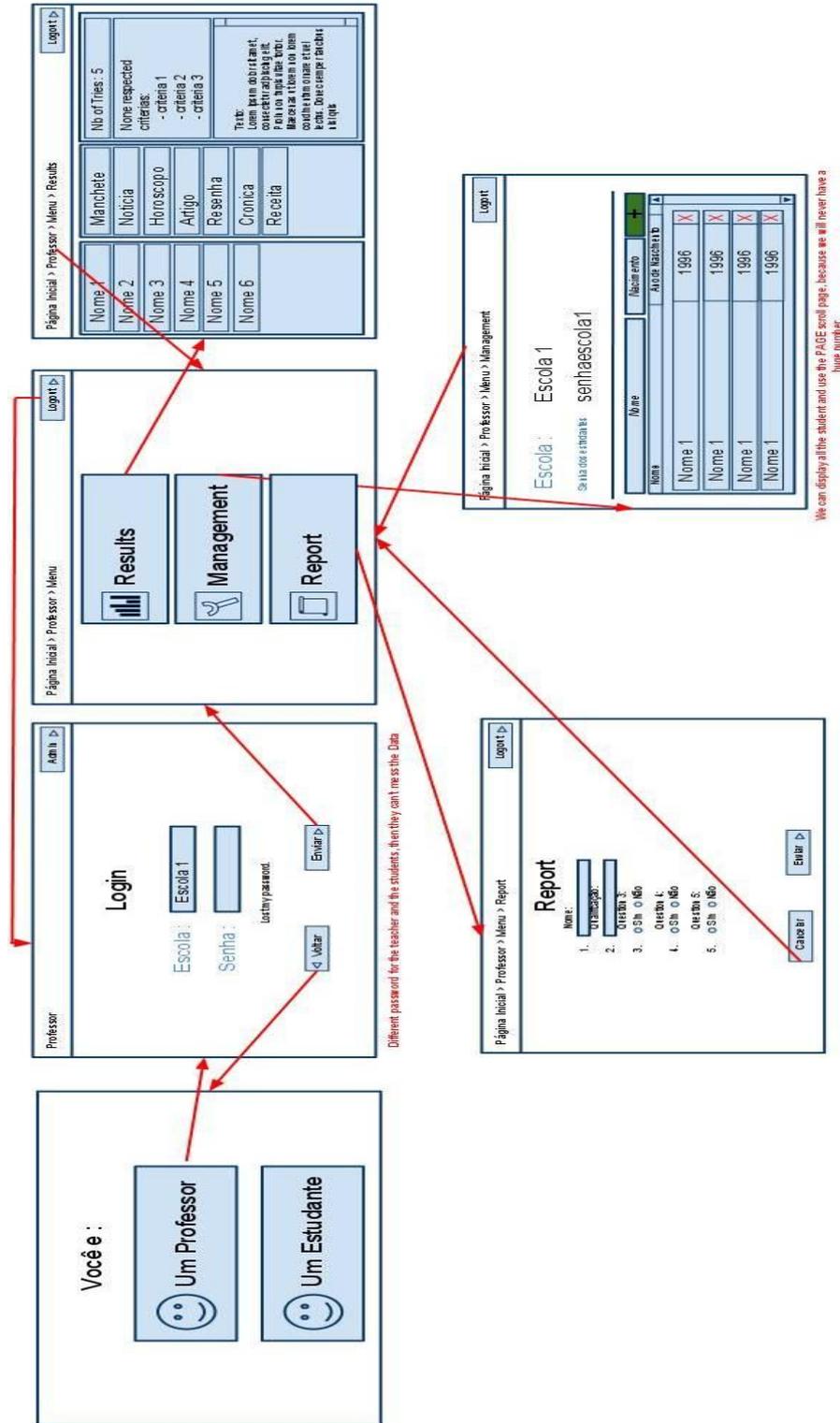
1. Figura 5.1:Textanalyser: Detalhe do processo de extração de texto



2. Cadastro do Administrador



3. Cadastro do professor



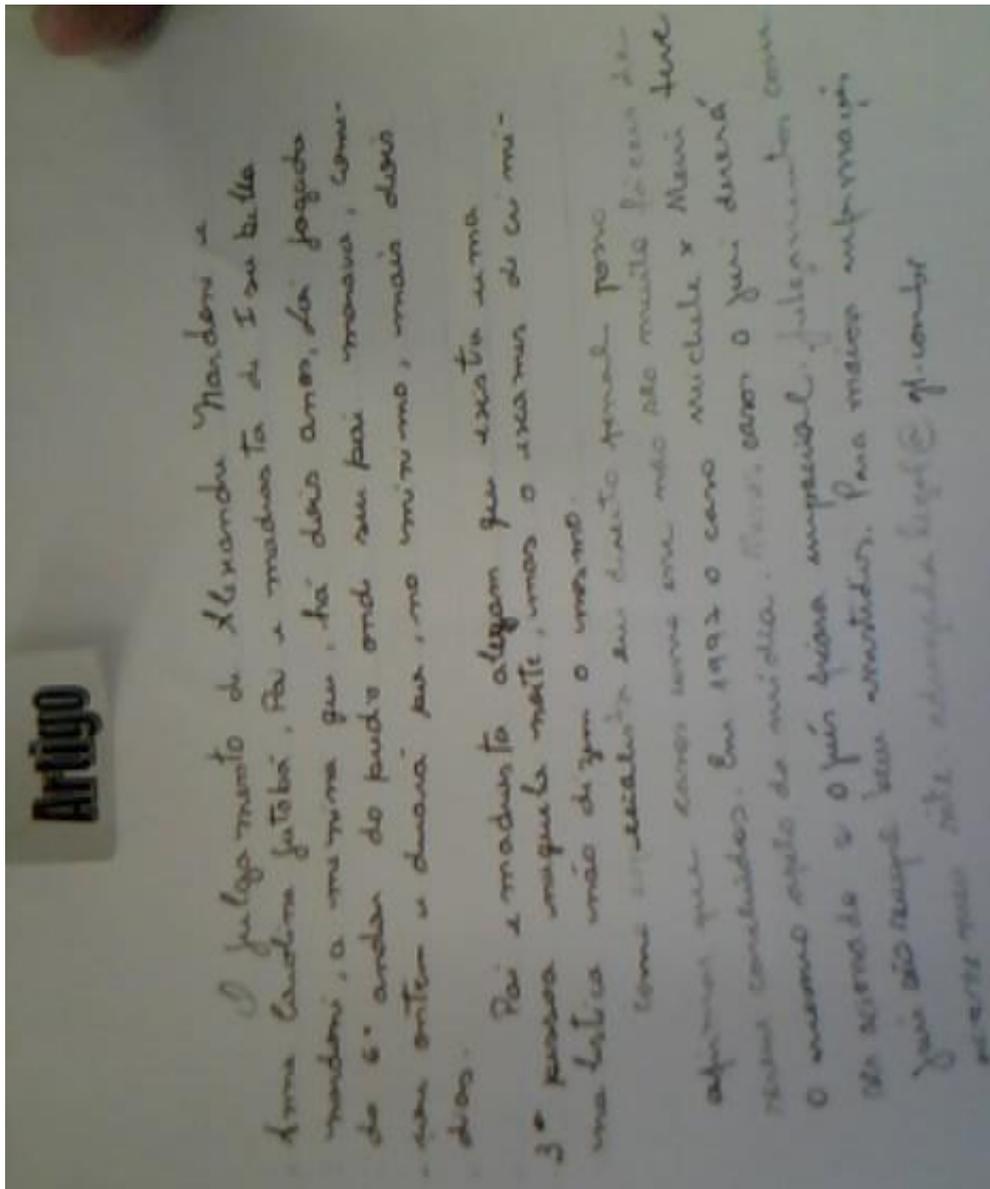
4. Figura 6.3: Texto antes da intervenção dialogada



“O julgamento de Alexandre Nardoni e Ana Carolina Jatobé. Pai e madastra de Isabella Nardoni, a menina que há dois anos foi jogada do 6º andar do prédio onde seu pai morava, começou ontem e durará no mínimo mais dois dias.

Pai e madastra alegaram que existiria uma terceira pessoa naquela noite, mas os exames de criminalista não dizem o mesmo.”

5. Figura 6.4: Texto depois da intervenção dialogada.



“O julgamento de Alexandre Nardoni e Ana Carolina Jatobé. Pai e madastra de Isabella Nardoni, a menina que há dois anos foi jogada do 6º andar do prédio onde seu pai morava, começou ontem e durará no mínimo mais dois dias.

Pai e madastra alegaram que existiria uma terceira pessoa naquela noite, mas os exames de criminalista não dizem o mesmo.

Como especialista em direito criminal posso afirmar que casos como esse não são muito fáceis de serem resolvidos. Em 1997 o caso Michelle x Maria teve o mesmo apelo da mídia. Nesses

casos o Juíz deve ser acionado e um juri popular. Julgamentos assim são sempre bem assistidos. Para maiores informações acesse meu site: advogadalegal@g1.com.br.”

6. Ferramentas utilizadas

6.1 Linguagem de programação *Python*

A linguagem python foi escolhida por ser uma linguagem computacional completa e adequada ao desenvolvimento de aplicações baseadas na Web, redes fechadas ou programas *stand-alone*. Python é poderosa em ambientes distribuídos complexos, como uma rede Internet. Porém, por sua versatilidade, oferece ao programador uma poderosa linguagem de programação de uso geral, com recursos para a construção de uma variedade de aplicativos, podendo ou não depender do uso de recursos de conectividade. Atualmente, a linguagem disponibiliza ao programador alguns avanços tecnológicos, fundamentais para o desenvolvimento de poderosas ferramentas, bem como uma estreita integração com diversos gerenciadores de bancos de dados.

7. Bibliotecas Feedparser e BeautifulSoup

Utilizamos o *feedparser* biblioteca para obter o texto do artigo a partir de um fluxo de RSS. De fato, um fluxo RSS é um XML padronizado. Em seguida, sabemos que vamos encontrar o texto do artigo para a marcação com o nome "<description>". Aqui é um método de extração de texto do fluxo do Artigos.com *site*:

```
def getTextsFromArtigosCom(self,url):
    if url != None and url != "":
        try:
            f=feedparser.parse(url)
        except:
            return {}
        textsList={}
        for entry in f.entries:
            textsList[entry.title] = self._html2text(entry.description)
        return textsList
    else:
        return {}
```

A fim de analisar a página HTML, utilizamos um Soup nome python biblioteca beautiful. É um parser HTML / XML para Python que pode transformar até mesmo marcação inválida em uma árvore de análise. Ele oferece maneiras simples e idiomático de navegar, pesquisar e modificar a árvore de análise. Aqui está um exemplo de análise de HTML, a partir do webartigos.com *site*:

```
soup = BeautifulSoup(html)
texts = soup.find("div", { "class" : "PostContent" }).contents[3].contents
```

Onde o `html` variável representa o HTML todo- da página. A função de encontrar a partir da instância da classe `BeautifulSoup` retornar a lista de todas as marcações `<div>` presente na página com o atributo `class = "PostContent"`.. Cada página web foi implementada com um HTML diferente *layout* / código e, em seguida, os fluxos RSS foram os preferidos, a fim de economizar tempo na implementação.

7.1 Naive Bayesian classification

A teoria de classificação bayesiana é amplamente utilizada para classificar os indivíduos em grupos não sobrepostos. É um ramo da lógica aplicada à tomada de decisão e estatística inferencial que lida com a probabilidade de inferência: usa o conhecimento dos eventos antes de prever acontecimentos futuros

7.2 Tornado web server + nginx

Nginx é uma fonte livre, aberta, de alto desempenho do servidor HTTP e proxy reverso, bem como um servidor proxy IMAP/POP3. Igor Sysoev iniciou o desenvolvimento de Nginx em 2002, com o primeiro lançamento público em 2004. Nginx agora abriga cerca de 7,65% (22,8 milhões) de todos os domínios em todo o mundo.

Nginx é conhecida por seu alto desempenho, estabilidade, rico conjunto de recursos, configuração simples, e baixo consumo de recursos.

Nginx é um de um punhado de servidores escritos para resolver o problema C10K. Ao contrário dos servidores tradicionais, Nginx não depende de *threads* para manipular as solicitações. Em vez disso, utiliza uma muito mais escalável event-driven architecture (assíncrono). Esta arquitetura utiliza memória pequena de valor previsível durante a carga.

Para lidar com milhares de solicitações simultâneas, nos beneficiamos de Nginx de memória de alto desempenho em todas as direções: desde a menor VPS até a clusters de servidores. Powers Nginx vários locais de alta visibilidade, como o Wordpress, Hulu, Github, Ohloh, SourceForge, Páginas Brancas e reator torrent. #

Usamos nginx balanceador de carga rápida para o problema C10K. Ele redireciona os usuários em diferentes instâncias de servidores do tornado que correm o *website*. Mais, nginx pode servir diretamente o arquivo estático do *website*.

O **Tornado** é uma versão open source do escalável servidor Web, non-blocking e ferramentas FriendFeed. A aplicação FriendFeed é escrita usando um *framework web* que parece um pouco com `web.py` ou `webapp` do Google, mas com ferramentas adicionais e otimizações para aproveitar a infraestrutura non-blocking subjacente.

O quadro é distinto da maioria dos *frameworks* servidor *web* tradicionais (e com certeza a maioria dos quadros Python) porque é razoavelmente firme o bloqueio. Porque com non-blocking e os usos *epoll* ou *kqueue*, ele pode lidar com milhares de conexões simultâneas em pé, o que significa que é ideal para serviços em tempo real *web*. Nós construímos o servidor *web* especificamente para lidar com tempo real do FriendFeed recursos - todos os usuários ativos do FriendFeed e uma conexão aberta para os servidores FriendFeed. (Para mais informações sobre servidores de escala para suportar milhares de clientes, consulte O problema C10K.)

7.3. Library Orange the input file

O Laranja é um componente baseado em mineração de dados e máquina suíte de software de aprendizagem, apresentando amigável mas poderosa e flexível programação visual front-end para análise exploratória de dados e visualização, e bindings Python e bibliotecas para execução de scripts. Ele inclui conjunto completo de componentes para o pré-processamento de dados de recursos, de pontuação e de filtragem, modelagem, a avaliação do modelo e técnicas de exploração. Ele é implementado em C++ (velocidade) e Python (flexibilidade).

A biblioteca de laranja requer como entrada um arquivo com um formato CSV separado com uma tabulação. O objetivo deste módulo é hadith a criação deste arquivo.

A primeira linha do arquivo de listas de nomes de atributos e de classe (para nós os atributos são os critérios). Segunda linha é dado o tipo do atributo (ou "discretas" ou "contínuo"). A terceira linha adiciona uma descrição adicional para cada coluna, a última coluna é indicada como "classe". Então, o resíduo da tabela é dado os dados.

Aqui está um pequeno exemplo da nossa tabela de exemplo gerado:

verbos irregulares	marcadores de tempo	[...]	text_size	type	
continous	continous		[...]	continous	discrete
class					
25.641025641	0.0		[...]	6	
Manchete					
3.87239535313	0.368799557441		[...]	897	artigo
3.61663652803	2.41109101869		[...]	284	crônica