

UNIVERSIDADE ESTADUAL DE MARINGÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

VINÍCIUS MOURÃO ALVES DE SOUZA

Análise automática de coerência semântica em resumos acadêmicos
escritos em português

Maringá

2011

VINÍCIUS MOURÃO ALVES DE SOUZA

Análise automática de coerência semântica em resumos acadêmicos
escritos em português

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Departamento de Informática, Centro de Tecnologia da Universidade Estadual de Maringá, como requisito parcial para obtenção do título de Mestre em Ciência da Computação.

Orientadora: Profa. Dra. Valéria
Delisandra Feltrim

Maringá
2011

Dados Internacionais de Catalogação-na-Publicação (CIP)
(Biblioteca Central - UEM, Maringá – PR., Brasil)

S729a Souza, Vinícius Mourão Alves de
Análise automática de coerência semântica em
resumos acadêmicos escritos em português / Vinícius
Mourão Alves de Souza. -- Maringá, 2011.
132 f. : il. col., figs., tabs.

Orientadora: Prof.^a Dr.^a Valéria Delisandra
Feltrim.

Dissertação (mestrado) - Universidade Estadual de
Maringá, Centro de Tecnologia, Departamento de
Informática, Programa de Pós-Graduação em Ciência da
Computação, 2011.

1. Resumo acadêmico - Ciência da computação. 2.
Coerência semântica - Resumo acadêmico. 3. Análise
automática - Coerência semântica. 4. Análise de
Semântica Latente - Análise automática. 5. Latent
Semantic Analysis. 6. SciPo (Scientific Portuguese)-
Ambiente de auxílio à escrita. 7. Natural Language
Processing. I. Feltrim, Valéria Delisandra, orient.
II. Universidade Estadual de Maringá. Centro de
Tecnologia. Departamento de Informática. Programa de
Pós-Graduação em Ciência da Computação. III. Título.

CDD 21.ed. 006.35

FOLHA DE APROVAÇÃO

VINICIUS MOURÃO ALVES DE SOUZA

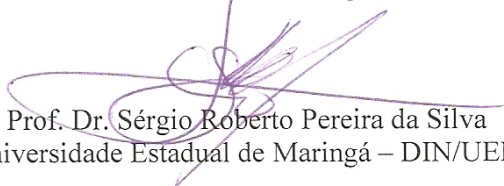
Análise automática de coerência semântica em resumos acadêmicos escritos em português

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação do Departamento de Informática, Centro de Tecnologia da Universidade Estadual de Maringá, como requisito parcial para obtenção do título de Mestre em Ciência da Computação pela Banca Examinadora composta pelos membros:

BANCA EXAMINADORA



Prof.ª Dra. Valéria Delisandra Feltrim
Universidade Estadual de Maringá – DIN/UEM



Prof. Dr. Sérgio Roberto Pereira da Silva
Universidade Estadual de Maringá – DIN/UEM



Prof. Dr. Thiago Alexandre Salgueiro Pardo
Universidade de São Paulo – ICMC/USP

Aprovada em: 27 de julho de 2011.

Local da defesa: Sala 101, Bloco C56, *campus* da Universidade Estadual de Maringá

AGRADECIMENTOS

Agradeço àqueles que contribuíram de forma direta para a realização deste trabalho, em especial:

Aos meus pais e irmãos pelo apoio e a compreensão de que um curso de pós-graduação não é sinônimo de desemprego, mas sim um aperfeiçoamento profissional, científico e pessoal.

Thainara Granero de Melo, não só pelo amor e carinho concedidos, mas pelo constante apoio do primeiro ao último dia no mestrado e pela compreensão dos altos e baixos vividos enquanto estudante. Você sempre soube o momento e a maneira correta de usar palavras motivadoras, sejam elas diretas ao ponto ou de ternura.

À minha orientadora Professora Doutora Valéria Delisandra Feltrim, pela confiança depositada em mim para o desenvolvimento deste projeto.

Aos colegas do café (Gustavo Sato, Everton Barros, César, Borth, Huff, ...) pelo compartilhamento de ideias, reclamações, experiências, conquistas, derrotas e planejamentos futuros. Em especial, os frequentadores do LSII (201) e do Laboratório do Mestrado.

Aos colegas Alison Rafael e Danilo Machado Junior que contribuíram respectivamente nas tarefas de coleta e anotação de corpúsculos. Também agradeço àqueles que voluntariamente participaram do experimento para avaliação do protótipo.

Ao pesquisador e desenvolvedor Radim Řehůřek, por sua prestatividade mesmo de longe e pelas barreiras idiomáticas impostas.

Maria Inês Davanço, por exercer o seu papel profissional com muita competência e nunca se negar a ajudar qualquer aluno.

Ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) por parte do apoio financeiro concedido a este trabalho.

Análise automática de coerência semântica em resumos acadêmicos escritos em português

RESUMO

O resumo pode ser considerado uma das seções mais importantes de um trabalho acadêmico, dado que, em conjunto com o título, é utilizado por pesquisadores na divulgação de suas pesquisas no meio científico. Nesse sentido, com o objetivo de auxiliar o processo de escrita das seções Resumo e Introdução de teses e dissertações da área de Ciência da Computação em português, Feltrim (2004) propôs um ambiente de auxílio à escrita denominado SciPo. O auxílio fornecido pelo SciPo é dado por meio de críticas e sugestões apresentadas ao usuário em relação ao uso da estrutura retórica (ou esquemática) identificada no texto submetido a análise. Embora o ambiente forneça *feedback* indicando quais pontos do texto devem ser melhorados, não analisa características relacionadas à semântica do texto, como a coerência, fundamental para a legibilidade e interpretabilidade do texto. Dessa maneira, o principal objetivo deste trabalho foi desenvolver recursos computacionais para a detecção automática de aspectos semânticos da seção Resumo de trabalhos acadêmicos escritos em português, de modo que eles possam ser utilizados pelo SciPo para o retorno de sugestões relacionadas à coerência. Mais especificamente, foram desenvolvidos classificadores que, com base em um certo conjunto de características extraídas automaticamente da superfície do texto e a partir da técnica LSA – *Latent Semantic Analysis* e de algoritmos de aprendizagem de máquina supervisionados, apresentam indicações sobre aspectos semânticos que contribuem para que o resumo seja considerado coerente. Tanto as avaliações intrínsecas dos classificadores quanto a avaliação do protótipo em um contexto de uso com usuários reais demonstraram o potencial dos classificadores no auxílio à escrita de resumos acadêmicos com um nível maior de coerência por meio da apresentação das novas sugestões.

Palvaras-chave: resumo acadêmico, coerência semântica, análise automática de coerência, Análise de Semântica Latente

Automatic analysis of semantic coherence in academic abstracts written in portuguese

ABSTRACT

The abstract can be considered one of the most important sections of an academic work. Along with the title, it is used by researchers to disseminate their research in scientific circles. In this context, Feltrim (2004) proposed an environment to support the writing of Abstract and Introduction sections called SciPo. This environment provides writing support by means of criticism and suggestions presented to the user with respect to the rhetorical (or schematic) structure identified in text submitted for analysis. Although the SciPo provides feedback indicating which parts of the text should be improved, it does not analyze features related to semantics, such as coherence, which is essential to the readability and interpretability of the text. Therefore, the main goal of this research was to develop computational resources to the automatic detection of semantic aspects of the Abstract section. We use these resources for the return of new suggestions related to coherence in the SciPo enviroment. In particular, we develop classifiers based on a set of features extracted automatically from the surface of the text and from the LSA technique - Latent Semantic Analysis and machine learning algorithms. Thus, the classifiers provide indications on the semantic aspects that contribute to the abstract is considered coherent. Both the intrinsic assessments of the classifiers as the evaluation of the prototype in a context of use with real users demonstrated the potential of classifiers to aid writing academics abstracts with higher lever of coherence through new suggestions.

Keywords: *academic abstract, semantic coherece, automatic analysis of coherence, Latent Semantic Analysis*

LISTA DE FIGURAS

Figura 3.1	Representação da matriz de entidades (Barzilay e Lapata, 2008) . .	29
Figura 3.2	Exemplo de vetor de características com todas as transições envolvendo duas entidades e as categorias sintáticas {S, O, X, -} (Barzilay e Lapata, 2008)	30
Figura 3.3	Exemplo de estrutura retórica RST (Pardo, 2005)	33
Figura 3.4	Exemplo de estrutura retórica com as relações <i>Concession</i> e <i>Justify</i> (Pardo, 2005)	34
Figura 3.5	Exemplo de estrutura retórica com a relação <i>Sequence</i> (Pardo, 2005)	34
Figura 3.6	Matriz original X como um produto de três matrizes (Deerwester et al., 1990)	37
Figura 3.7	Processo de redução de dimensões da matriz criada pela SVD (Deerwester et al., 1990)	37
Figura 3.8	Comparação LSA entre três palavras (Kintsch, 2002)	38
Figura 4.1	Anotação realizada por um anotador humano em relação às dimensões Dim_P e Dim_T (Higgins et al., 2004)	42
Figura 4.2	Exemplo de <i>feedback</i> apresentado pelo <i>CriterionTM</i> em relação aos aspectos de avaliação de coerência (Burstein et al., 2003a)	43
Figura 4.3	Exemplo de <i>feedback</i> apresentado pela <i>Intelligent Essay AssessorTM</i>	45
Figura 4.4	Exemplo de <i>feedback</i> apresentado pela <i>Summary StreetTM</i>	46
Figura 4.5	Parte das métricas apresentadas pela ferramenta <i>Coh-Metrix</i> (Graesser et al., 2004)	48
Figura 5.1	Visão geral do processo de apoio à estruturação do texto (Feltrim, 2004)	51
Figura 5.2	Críticas e sugestões fornecidas pelo SciPo em relação a estrutura construída pelo usuário	52
Figura 5.3	Processo <i>bottom-up</i> realizado pelo SciPo (Feltrim, 2004)	53
Figura 5.4	Resumo com sua estrutura detectada automaticamente pelo SciPo .	54
Figura 6.1	Hierarquia da aprendizagem indutiva (Rezende, 2003)	60
Figura 6.2	Arquivo no formato ARFF (Witten e Frank, 2005)	66
Figura 6.3	Funcionalidade <i>Classify</i> da ferramenta <i>Explorer</i> do Weka	67
Figura 7.1	Exemplo de resumo preliminarmente anotado	71

Figura 7.2	Relacionamento das sentenças com o Título de acordo com as categorias retóricas	76
Figura 7.3	Relacionamento das sentenças com o(s) propósito(s) do resumo, de acordo com as categorias retóricas	78
Figura 7.4	Relacionamento das sentenças de <i>Gap</i> com as sentenças de <i>Back-ground</i>	80
Figura 7.5	Distribuição das categorias retóricas em relação a Dimensão Quebra de Linearidade	82
Figura 7.6	Captura de tela da ferramenta <i>online</i> para anotação do corpus . . .	83
Figura 8.1	Exemplo de resumo com anotação final	88
Figura 8.2	Gráficos que correspondem às respostas referentes às opiniões sobre a relevância das sugestões, nível de coerência após a reescrita e dificuldade de adequação das sugestões	110
Figura 8.3	Captura de tela do protótipo desenvolvido	113

LISTA DE TABELAS

Tabela 3.1	Relações de equivalência entre os conjuntos C_b e C_p e as transições da Teoria de <i>Centering</i>	25
Tabela 3.2	Entidades e transição identificada no enunciado (U_1)	26
Tabela 3.3	Entidades e transição identificada no enunciado (U_2)	26
Tabela 3.4	Entidades e transição identificada no enunciado (U_3)	26
Tabela 3.5	Entidades e transição identificada no enunciado (U_{4a})	27
Tabela 3.6	Entidades e transição identificada no enunciado (U_{4b})	27
Tabela 3.7	Relações e nuclearidade da RST (Mann e Thompson, 1987)	32
Tabela 3.8	Exemplo de matriz de co-ocorrência de termos	35
Tabela 3.9	Valores atribuídos pela LSA na comparação entre três sentenças	39
Tabela 5.1	Modelo de um resumo típico, proposto por Weissberg e Buker (1990)	55
Tabela 5.2	Modelo de resumo proposto por Feltrim et al. (2003)	55
Tabela 5.3	Descrição e exemplo dos papéis retóricos identificados pelo AZPort	57
Tabela 5.4	Conjunto de atributos utilizados pelo AZPort	57
Tabela 6.1	Conjunto de exemplos no formato atributo-valor (Rezende, 2003)	59
Tabela 6.2	Matriz de confusão genérica (Rezende, 2003)	62
Tabela 6.3	Tipos de erros de classificação em uma matriz com duas classes	62
Tabela 6.4	Níveis de concordância dos valores <i>Kappa</i> (Landis e Koch, 1977)	64
Tabela 7.1	Distribuição e totais de palavras do corpus por área de conhecimento	70
Tabela 7.2	Frequência dos componentes observada no corpus	72
Tabela 7.3	Distribuição dos componentes que compõem a estrutura retórica dos resumos analisados	72
Tabela 7.4	Quantidade média de sentenças por resumo, distribuído por componentes da estrutura retórica	73
Tabela 7.5	Matriz de confusão: análise manual <i>vs.</i> análise automática	74
Tabela 7.6	Valores da concordância <i>Kappa</i> entre dois anotadores humanos para a Dimensão Título	77
Tabela 7.7	Valores da concordância <i>Kappa</i> entre dois anotadores humanos para a Dimensão Propósito	79
Tabela 8.1	Conjunto de atributos extraídos do corpus	89
Tabela 8.2	Concordância <i>Kappa</i> dos classificadores da Dimensão Título com diferentes conjuntos de atributos	92

Tabela 8.3	Matriz de confusão do classificador final da Dimensão Título	93
Tabela 8.4	Desempenho do classificador final da Dimensão Título	94
Tabela 8.5	Concordância <i>Kappa</i> dos classificadores da Dimensão Propósito (envolvendo o componente <i>Methodology</i>) com diferentes conjuntos de atributos	95
Tabela 8.6	Matriz de confusão do classificador final da Dimensão Propósito (componente <i>Methodology</i>)	96
Tabela 8.7	Desempenho do classificador final da Dimensão Propósito (compo- nente <i>Methodology</i>)	96
Tabela 8.8	Concordância <i>Kappa</i> dos classificadores da Dimensão Propósito (envolvendo o componente <i>Result</i>) com diferentes conjuntos de atributos	98
Tabela 8.9	Matriz de confusão do classificador final da Dimensão Propósito (componente <i>Result</i>)	99
Tabela 8.10	Desempenho do classificador final da Dimensão Propósito (compo- nente <i>Result</i>)	99
Tabela 8.11	Concordância <i>Kappa</i> dos classificadores da Dimensão Propósito (envolvendo o componente <i>Conclusion</i>) com diferentes conjuntos de atributos	101
Tabela 8.12	Matriz de confusão do classificador final da Dimensão Propósito (componente <i>Conclusion</i>)	102
Tabela 8.13	Desempenho do classificador final da Dimensão Propósito (compo- nente <i>Conclusion</i>)	102
Tabela 8.14	Concordância <i>Kappa</i> dos classificadores da Dimensão Lacuna-Contexto com diferentes conjuntos de atributos	104
Tabela 8.15	Matriz de confusão do classificador final da Dimensão Lacuna-Contexto	104
Tabela 8.16	Desempenho do classificador final da Dimensão Lacuna-Contexto .	105
Tabela 8.17	Versão inicial de um resumo acadêmico	111
Tabela 8.18	Versão final de um resumo acadêmico após as sugestões de coerência	112
Tabela A.1	Lista de <i>stopwords</i>	127
Tabela A.2	Lista de possíveis elementos anafóricos	128
Tabela A.3	Lista de palavras que podem representar uma transição	129

LISTA DE SIGLAS E ABREVIATURAS

AES: *Automated Essay Scoring*

AMADEUS: *Amiable Article Development for User Support*

AZ: *Argumentative Zoning*

AZPort: *Argumentative Zoning for Portuguese*

DC-UEL: Departamento de Computação da Universidade Estadual de Londrina *

DIN-UEM: Departamento de Informática da Universidade Estadual de Maringá

ETS: *Education Testing Service*

GMAT: *Graduate Management Admission Test*

HTML: *Hypertext Markup Language*

IBL: *Instance Based Learning*

IDF: *Inverse Document Frequency*

IEA: *Intelligent Essay Assessor*

Inf-UFPel: Departamento de Informática do Instituto de Física e Matemática da Universidade Federal de Pelotas

KAT: *Knowledge Analysis Technologies*

LDA: *Latent Dirichlet Allocation*

LSA: *Latent Semantic Analysis*

LSI: *Latent Semantic Indexing*

MR: Meta-regra

PLN: Processamento de Linguagem Natural

RP: *Random Projections*

RST: *Rhetorical Structure Theory*

SciPo: *Scientific Portuguese*

SGML: *Standard Generalized Markup Language*

SMO: *Sequential Minimal Optimization*

SVD: *Singular Value Decomposition*

TCC: Trabalho de Conclusão de Curso

TF: *Term Frequency*

VSM: *Vector Space Model*

XML: *Extensible Markup Language*

W3C: *World Wide Consortium*

WEKA: *Waikato Environment for Knowledge Analysis*

SUMÁRIO

1	Introdução	14
2	Coesão e Coerência Textual	18
2.1	Coesão	18
2.2	Coerência	19
2.2.1	Coerência Local e Coerência Global	21
2.2.2	Tipos de Coerência	21
3	Teorias e Métodos Relacionados à Coerência Textual	23
3.1	Teoria de Centering	23
3.2	Modelo de Entidades	27
3.3	Teoria da Estrutura Retórica	31
3.4	Análise de Semântica Latente	35
4	Análise Automática de Textos	40
4.1	Criterion TM	40
4.2	Knowledge Analysis Technologies TM	44
4.3	IntelliMetric TM	47
4.4	Coh-Metrix	48
5	Ambiente SciPo	50
5.1	Estruturação de Resumos Acadêmicos	54
5.2	Argumentative Zoning for Portuguese - AZPort	56
6	Aprendizagem de Máquina	58
6.1	Conceitos de Aprendizagem de Máquina	59
6.1.1	Métricas de Avaliação	61
6.2	Waikato Environment for Knowledge Analysis	66
7	Coleta, Anotação e Análise de Corpus	69
7.1	Coleta	69
7.2	Anotação	70
7.3	Análise	72
7.4	Segunda Fase de Anotação e Análise	75
7.4.1	Ferramenta para Anotação	82

8	Análise Automática de Coerência	84
8.1	Pré-processamento	85
8.2	Construção do Espaço Semântico	86
8.3	Comparação Semântica entre Sentenças	87
8.4	Extração de Atributos	87
8.5	Seleção de Atributos	90
8.6	Construção dos Modelos de Classificação	91
8.6.1	Dimensão Título	91
8.6.2	Dimensão Propósito (componente Methodology)	94
8.6.3	Dimensão Propósito (componente Result)	97
8.6.4	Dimensão Propósito (componente Conclusion)	100
8.6.5	Dimensão Lacuna-Contexto	103
8.7	Sugestões Apresentadas no Protótipo	105
8.7.1	Dimensão Título	106
8.7.2	Dimensão Propósito	106
8.7.3	Dimensão Lacuna-Contexto	107
8.8	Avaliação do Protótipo	107
9	Conclusões e Trabalhos Futuros	114
	Referências	118
	Apêndice A Listas de Palavras Utilizadas nos Atributos dos Classificadores	127
	Apêndice B Questionário Utilizado na Avaliação do Protótipo	131

Introdução

O resumo pode ser considerado uma das seções mais importantes de um trabalho acadêmico, dado que, em conjunto com o título, é utilizado por pesquisadores na divulgação de suas pesquisas no meio científico. Segundo Feltrim (2004), um resumo acadêmico deve ser redigido com muito cuidado, de forma a ser completo (em termos de informações necessárias), interessante e informativo, dispensando a consulta ao restante do texto para que o leitor tenha ideia do que trata o trabalho e, ao mesmo tempo, estimular o interesse pela leitura do texto completo.

Assim como os trabalhos científicos possuem uma estrutura bem definida, que pode ser enunciada como: Introdução – Desenvolvimento – Conclusão, sendo que Desenvolvimento pode desdobrar-se nas seções Materiais e Métodos e Resultados ou ainda Materiais e Métodos, Resultados e Discussão, a seção destinada ao resumo desses trabalhos também possui um esquema estrutural bem definido e que abrange aspectos essenciais da comunicação científica, denominado de estrutura esquemática ou retórica.

Como nos resumos os tipos de informação e a ordem em que aparecem são muito convencionais, são propostos modelos que objetivam guiar o escritor quanto aos tipos de informações necessárias, opcionais e a ordem de apresentação. Baseado nos modelos de Aluísio e Oliveira Jr (1996); Swales (1990) e Weissberg e Buker (1990), Feltrim et al. (2003) propôs um modelo composto por seis componentes esquemáticos dispostos na seguinte ordem: Contexto, Lacuna, Propósito, Metodologia, Resultado e Conclusão, que juntos formam a estrutura esquemática ou retórica da seção resumo. Destacamos que cada

um desses componentes é composto por sentenças que nomeiam e descrevem os tipos de informações esperadas no resumo.

Com base em modelos que levam em conta a estrutura retórica dos textos, diferentes ferramentas computacionais foram propostas ao longo dos anos para auxiliar a escrita do primeiro rascunho de um texto e/ou a sua revisão. Entre os trabalhos voltados à língua inglesa, citamos o AMADEUS – *Amiable Article Development for User Support* (Aluísio et al., 2001), SciPo-Farmácia (Aluísio et al., 2005), Writer’s Assistant (Sharples et al., 1994; Sharples e Pemberton, 1992), Writing Environment (Lansman, 1993), Composer (Pemberton et al., 1996; Shurville et al., 1997), Academic Writer (Broady e Shurville, 2000), Abstract Helper (Narita, 2000a,b) e Mover (Anthony e Lashkia, 2003). Tais ferramentas auxiliam estudantes não-nativos na escrita de textos acadêmicos. Para o português, destacamos o ambiente de auxílio à escrita SciPo – *Scientific Portuguese* (Feltrim, 2004; Feltrim et al., 2006), que fornece suporte à escrita das seções Resumo e Introdução de teses e dissertações em Ciência da Computação. O SciPo tem por objetivo apoiar a estruturação de textos e a sua realização linguística de forma flexível, deixando o usuário livre para escolher entre dois modos de trabalho, a saber:

- i) um processo *top-down*, que parte do planejamento estrutural para a escrita propriamente dita, herdado do projeto AMADEUS; ou
- ii) um processo *bottom-up*, em que se submete um texto já escrito à análise (detecção e crítica) automática da estrutura.

Na verdade, trata-se de pontos de partida distintos para um mesmo processo cíclico de refinamento, já que a estrutura detectada e criticada em (ii) pode ser aprimorada por meio dos recursos disponíveis em (i).

O auxílio fornecido pelo SciPo é dado por meio de críticas e sugestões que são apresentadas ao usuário em relação ao uso da estrutura retórica identificada no texto submetido à análise. Embora o ambiente forneça *feedback* indicando quais pontos do texto devem ser melhorados, não analisa características relacionadas à semântica do texto, como a coerência, fundamental para a legibilidade e interpretabilidade do texto.

A coerência em conjunto com a coesão são elementos responsáveis para que um grupo de palavras ou frases faça sentido entre si. Neste trabalho assumimos que a coerência diz respeito a possibilidade de se estabelecer um sentido lógico entre diferentes sentenças de um texto. Logo, é um princípio de interpretabilidade ligado à situação de comunicação e à capacidade do receptor em calcular o sentido desse texto e, por isso, está vinculada ao texto, mas não depende somente dele (Koch e Travaglia, 2003). Por outro lado, para que

seja possível o estabelecimento de sentido, se faz necessário o uso de elementos textuais responsáveis pela conexão entre palavras ou frases, pertinentes a coesão (de Beaugrande e Dressler, 1981). Por isso, a coerência e a coesão estão intimamente relacionadas e em boa parte dos casos são abordadas em conjunto.

Desse modo, visando complementar as funcionalidades do ambiente SciPo, em especial na escrita da seção Resumo, este trabalho teve como objetivo propor a análise de diferentes aspectos semânticos do texto e o desenvolvimento de recursos computacionais para a detecção automática de tais aspectos, de modo que eles possam ser utilizados pelo SciPo para o retorno de sugestões relacionadas à coerência.

A análise proposta se deu pela elaboração de quatro tipos diferentes de relacionamentos entre as sentenças do resumo, denominados de dimensões, em que: (i) Dimensão Título, verificou-se o relacionamento semântico do componente Propósito com o Título do resumo; (ii) Dimensão Propósito, verificou-se o relacionamento semântico dos componentes Metodologia, Resultado e Conclusão com o componente Propósito; (iii) Dimensão Lacuna-Contexto, verificou-se o relacionamento semântico entre os componentes Lacuna e Contexto; e (iv) Dimensão Quebra de Linearidade, verificou-se a quebra de sentido lógico entre sentenças adjacentes do resumo. Exceto na Dimensão Quebra de Linearidade, considerou-se a estrutura retórica do resumo na análise do relacionamento semântico esperado entre os componentes.

Para a verificação de tais relacionamentos entre os componentes, foi realizada a anotação e análise de um corpus de resumos de Trabalhos de Conclusão de Curso em Ciência da Computação. Os resultados dessa análise (Souza e Feltrim, 2011) embasaram a proposta das dimensões e o corpus anotado serviu de material de treinamento e teste na indução de classificadores responsáveis pela automatização das tarefas realizadas pelas dimensões.

Os classificadores foram desenvolvidos com base em um certo conjunto de características extraídas automaticamente do texto, com o objetivo de apresentar indicações sobre aspectos semânticos que contribuem para que um texto seja considerado coerente. Entre as características extraídas automaticamente do texto, as de maior valor preditivo foram calculadas a partir da LSA – *Latent Semantic Analysis* (Landauer et al., 1998), um método estatístico para a extração e representação de conhecimento aplicado em corpus. A ideia básica da LSA é formar um espaço semântico em que a semelhança entre os termos se dá pela ocorrência em contextos comuns. Apesar de ter surgido no contexto de recuperação de informação, o método tem sido utilizado por diversas aplicações, como ferramentas de *Automated Essay Scoring* (AES) na avaliação de redações, em que, por exemplo, se compara o texto avaliado com um texto base escrito por um profissional e

a nota emitida leva em consideração a similaridade semântica entre os textos. Segundo Kukich (2000), ferramentas de AES têm apresentado bons resultados quando aplicadas à redações, como as exigidas em testes de proficiência e admissão, sendo destaques nessa categoria as ferramentas *E-rater*TM(Attali e Burstein, 2006) e *Intellimetric*TM(Rudner et al., 2006), o que contribui para justificar o uso do método no contexto deste trabalho, embora nosso objetivo não seja retornar uma nota em relação ao texto analisado.

Foram desenvolvidos e avaliados classificadores para cada uma das dimensões propostas, com exceção da Dimensão Quebra de Linearidade, em que o baixo número de exemplos identificados no processo de anotação impossibilitou o desenvolvimento de um classificador para desempenhar sua tarefa. Os classificadores das outras três dimensões obtiveram um bom desempenho, o que possibilitou a sua utilização em um contexto real de uso do ambiente SciPo. Tal fato também é comprovado pela avaliação positiva realizada com usuários reais, em que um protótipo do ambiente SciPo com novas sugestões relacionadas à coerência é avaliado.

Acreditamos que este trabalho possa beneficiar não só o ambiente de auxílio à escrita SciPo, mas também outras ferramentas de Processamento de Linguagem Natural (PLN), como sumarizadores automáticos.

Esta dissertação está dividida em nove capítulos, sendo este o primeiro deles. No Capítulo 2 são introduzidos os conceitos de coesão e coerência textual; no Capítulo 3 são apresentadas teorias e métodos empregados na análise da coerência textual, tanto de base linguística quanto estatística, e que têm sido utilizados em diferentes sistemas computacionais; no Capítulo 4 são apresentadas ferramentas computacionais que realizam a análise automática de textos e que possuem algum critério de análise que julgamos relacionado à coerência; no Capítulo 5 é abordado o ambiente de auxílio à escrita SciPo, no qual este trabalho está diretamente inserido; no Capítulo 6 são introduzidos conceitos de aprendizagem de máquina, necessários para o desenvolvimento e avaliação dos classificadores propostos; no Capítulo 7, os processos de coleta e anotação e os resultados da análise do corpus utilizado no desenvolvimento do trabalho são apresentados; no Capítulo 8 são abordadas as fases necessárias para o desenvolvimento dos classificadores propostos e apresentada a avaliação intrínseca dos mesmos, bem como a avaliação do protótipo com usuário reais; no Capítulo 9 são apresentadas as conclusões do trabalho. Também faz parte da dissertação dois anexos: um contendo listas de palavras utilizadas no cálculo de atributos dos classificadores desenvolvidos e outro contendo o questionário utilizado na avaliação do protótipo desenvolvido.

Coesão e Coerência Textual

Na linguística não existe um consenso quanto às definições de coesão e coerência textual, sendo que diversas definições foram formuladas ao longo dos anos por teóricos como Halliday e Hasan (1976), de Beaugrande e Dressler (1981), Widdowson (1984), van Dijk e Kintsch (1983) e Charolles (1978).

Ainda que cada autor apresente um ponto de vista distinto, todos estão de acordo quanto ao fato de que coesão e coerência estão intimamente relacionadas, tanto no processo de produção, como na compreensão do texto e, por isso, são as principais responsáveis para se alcançar a textualidade, fazendo com que o texto não seja um amontoado de palavras ou frases sem sentido entre si (Koch e Travaglia, 1989).

Este capítulo aborda questões relacionadas à linguística textual que são de interesse para este trabalho, tais como a definição de coesão, abordada na Seção 2.1, e coerência, abordada na Seção 2.2. Ainda na Seção 2.2, apresentamos as Subseções 2.2.1 e 2.2.2, referentes às definições sobre coerência local/global e os diferentes tipos de coerência.

2.1 Coesão

Neste trabalho assumimos que a coesão trata o modo como os elementos textuais são relacionados em uma sequência formada por conexões entre palavras utilizando elementos formais como marcas linguísticas, que assinalam um vínculo entre os seus componentes (de Beaugrande e Dressler, 1981). Dentre os mecanismos gramaticais que auxiliam para o alcance da coesão, podemos citar o uso de uma classe de palavras invariáveis chamada

conjunções coordenativas e subordinativas, que estabelecem conexões entre diferentes termos ou orações, o emprego de pronomes, numerais e elipses. Dentre os mecanismos lexicais podemos citar o uso de repetição, substituição e associação.

Os elementos de coesão determinam a transição de ideias entre frases ou parágrafos, como pode ser observado nas palavras destacadas em negrito no exemplo de Jordão e Bellezi (2007):

*“Os sem-terra fizeram um protesto em Brasília contra a política agrária do país, **porque** consideram injusta a atual distribuição de terras. **Porém** o ministro da Agricultura considerou a manifestação um ato de rebeldia, **uma vez** que o projeto de Reforma Agrária pretende assentar milhares de sem-terra.”*

2.2 Coerência

A coerência, segundo Koch e Travaglia (2003), está diretamente ligada à possibilidade de estabelecer um sentido para o texto, sendo um princípio de interpretabilidade, ligada à inteligibilidade do texto em uma situação de comunicação e à capacidade que o receptor tem para calcular o sentido desse texto. É construída a partir da interação entre a situação de comunicação, os conceitos ativados pelo texto e os conhecimentos armazenados na memória dos participantes. Dessa maneira, a coerência está vinculada ao texto, mas não depende somente dele.

Charolles (1978) fez um estudo sobre coerência e coesão sem diferenciá-las. Para o autor, o conceito de coesão apresentado por outros autores é denominado de coerência microestrutural, enquanto o conceito de coerência é tratado como coerência macroestrutural. A coerência microestrutural corresponde a um nível local de organização do texto formada por relações estabelecidas entre palavras ou frases sucessivamente ordenadas, enquanto a coerência macroestrutural corresponde a um nível global de organização formada por sequências maiores do texto. Ainda segundo Charolles (1978):

“[...] a coerência de um enunciado deve ser conjuntamente determinada de um ponto de vista local e global, pois um texto pode muito bem ser microestruturalmente coerente sem o ser macroestruturalmente.”

Dessa forma, Charolles (1978) compartilha a visão de outros autores que consideram a possibilidade de um texto ser coeso, porém incoerente.

Para a formação de um texto coerente, Charolles (1978) propôs quatro meta-regras:

1. Meta-regra de repetição (MR-I): um texto coerente deve ter elementos repetidos.

Esta meta-regra está relacionada à recuperação de termos de frases anteriores por meio de pronomes, elipses, elementos lexicais ou substitutivos, o que constitui um processo de repetição ou recorrência. Podemos interpretá-la como a coesão textual definida por autores como de Beaugrande e Dressler (1981) e que esse elemento é uma condição necessária, mas não suficiente para que um texto seja coerente.

2. Meta-regra de progressão (MR-II): um texto coerente deve apresentar renovação do suporte semântico.

A segunda meta-regra completa a primeira, posto que estipula que um texto coerente não pode simplesmente repetir indefinidamente seu próprio assunto, evitando a circularidade do discurso e apresentando informações novas à medida que vai sendo escrito.

3. Meta-regra de não-contradição (MR-III): cada parte de um texto coerente deve fazer sentido com o que foi escrito anteriormente.

Com base na lógica, Charolles (1978) recorda que é inadmissível que uma mesma proposição seja conjuntamente verdadeira e não verdadeira, ou falsa e não falsa.

4. Meta-regra de relação (MR-IV): para que uma sequência ou um texto seja coerente, é preciso que os fatos que se denotam no mundo representado estejam diretamente relacionados.

Sobre a última meta-regra, Charolles (1978) relata que a regra traz em si intensas características pragmáticas. As relações entre os enunciados podem ocorrer tanto com a presença explícita de conectivos, como sem a presença explícita dos mesmos. Nesse último caso, a conexão entre os fatos ocorre somente mediante a fatores pragmáticos, como conhecimento de mundo e conhecimento compartilhado.

As quatro meta-regras de Charolles (1978) não satisfazem sozinhas todas as condições necessárias para um texto ser avaliado como coerente, fato reconhecido pelo próprio autor. Também não são pensadas como tendo caráter normativo, de modo a estabelecer regras que constituem a coerência textual, mas sim como diretrizes para a composição e avaliação de textos coerentes por meio de orientações referentes à continuidade temática, progressão semântica, possíveis contradições entre ideias e pertinência do conteúdo abordado.

2.2.1 Coerência Local e Coerência Global

Diferentemente de autores como Koch e Travaglia (2003), para os quais a coerência é uma propriedade inteiramente global ligada à composição do texto como um todo, van Dijk (1983) e van Dijk e Kintsch (1983) estabelecem dois níveis básicos de coerência que se complementam, sendo eles:

- **Coerência local (microestrutural):** diz respeito às partes do texto como sentenças ou proposições e suas conexões lineares;
- **Coerência global (macroestrutural):** referente à totalidade do texto ou a fragmentos maiores, em que se estabelece uma conexão com o tema, ideia, essência ou finalidade.

Embora seja feita essa distinção de níveis, os autores afirmam que existe uma interdependência entre os mesmos, pois um conjunto de sentenças linearmente coerentes contribui para se alcançar a coerência global e, da mesma forma, a coerência local pode ser construída a partir da coerência global.

2.2.2 Tipos de Coerência

Segundo van Dijk e Kintsch (1983), a coerência textual pode ser analisada sob quatro perspectivas: (1) coerência semântica, (2) coerência sintática, (3) coerência estilística e (4) coerência pragmática.

1. **Coerência semântica:** refere-se à relação entre os significados dos elementos que constituem as sentenças, que devem ser relacionados ou complementares. A coerência semântica também se refere à relação entre sentenças sucessivas do discurso, que devem ter seus conteúdos interligados. A incoerência semântica ocorre quando o significado dos elementos das sentenças ou os conteúdos de sentenças sucessivas são semânticamente distantes. A coerência semântica pode se manifestar tanto em nível local quanto global.
2. **Coerência sintática:** refere-se aos meios sintáticos usados para expressar a coerência semântica, como o uso de conectivos, pronomes, artigos, advérbios, entre outros meios. Equivale aos recursos coesivos da língua que auxiliam na construção da coerência.

3. **Coerência estilística:** refere-se ao estilo de escrita conforme o gênero textual, sem a mistura de registros linguísticos, como o uso de linguagem informal em um texto formal. A coerência estilística envolve o léxico e as estruturas frasais adequadas ao texto produzido. Segundo autores como Koch e Travaglia (2003), caso se faça necessário empregar gírias ou outras construções informais em determinados textos aos quais seriam inadequadas, devem ser empregadas expressões explicativas e de ressalvas que antecedam as coloquialidades.
4. **Coerência pragmática:** refere-se ao texto visto como uma sequência de atos de fala realizados de forma apropriada e com informações claras. Por exemplo, a resposta de uma pergunta pode se manifestar por meio de uma afirmação, de outra pergunta, de uma promessa, de uma negação e não de algo totalmente desconectado ao tema da pergunta. Nesse sentido, os fatores pragmáticos equivalem ao fato do produtor e o interlocutor reconhecerem o esquema textual empregado na situação comunicativa, considerando-se o contexto extralinguístico e o significado pretendido.

É importante destacar que uma incoerência pode ser caracterizada sob mais de uma perspectiva. Por exemplo, problemas estilísticos podem equivaler também à falhas lexicais, caracterizando tanto incoerência estilística quanto sintática.

Teorias e Métodos Relacionados à Coerência Textual

Neste capítulo são apresentados teorias e métodos empregados na análise da coerência textual, tanto de base linguística quanto estatística, e que têm sido utilizados em diferentes sistemas computacionais.

3.1 Teoria de Centering

A *Centering* (Grosz et al., 1995) é uma teoria discursiva capaz de medir a coerência local do discurso, partindo do princípio da compatibilidade entre os centros (ou focos) de atenção dos enunciados de um texto e da escolha de expressões de referenciamento.

Essa teoria tem sido utilizada em várias aplicações na área de Processamento de Linguagem Natural, tais como geração automática de texto (Karamanis, 2003), sumarização automática (Orăsan, 2003) e, mais recentemente, na avaliação automática de textos (Barzilay e Lapata, 2008; Hasler, 2004, 2008; Karamanis et al., 2004; Miltsakaki e Kukich, 2004).

Os autores da teoria propõem que o discurso é composto por segmentos textuais que, por sua vez, são compostos por uma sequência de enunciados e que tais enunciados possuem centros de atenção, que são as entidades mais salientes das frases que servem de ligação entre um enunciado e outro no segmento do discurso. Centros são objetos semânticos e não palavras, frases ou formas sintáticas.

Desse modo, a coerência global está relacionada aos diversos segmentos do texto, enquanto a coerência local está vinculada aos enunciados adjacentes de um mesmo segmento. Sendo assim, a teoria de *Centering* apresenta um modelo para o tratamento da coerência local, com um sistema de restrições e regras que governam as relações entre os centros de atenção e as formas escolhidas para construção dos enunciados.

Adaptamos o exemplo de Grosz et al. (1995), em que temos um segmento S com quatro enunciados (*utterances* ou U_i), sendo que o último enunciado do segmento pode ser formado por (U_{4a}) ou (U_{4b}):

- S : (U_1) João foi à sua loja de música favorita para comprar um piano.
 (U_2) Ele frequentava a loja há vários anos. (Ele = João)
 (U_3) Estava animado por finalmente poder comprar um piano. (Estava = João)
 (U_{4a}) Ele chegou justo quando a loja fechava. (Ele = João)
 (U_{4b}) Ela estava fechando assim que João chegou. (Ela = loja de música)

Segundo os autores, o segmento S é intuitivamente mais coerente se finalizado com o enunciado (U_{4a}) em vez de (U_{4b}), embora expressem exatamente a mesma informação. Isso se deve ao enunciado (U_{4a}) formar um segmento com um único foco de atenção, João, enquanto o uso do enunciado (U_{4b}) faria o foco de atenção oscilar entre João e loja de música. A teoria de *Centering* foi desenvolvida para capturar essas diferenças de continuidade.

Os enunciados que compõem um segmento textual são representados pela sequência $U_i - U_n$. Para cada enunciado U_i é associado um conjunto ordenado de entidades de contexto do discurso, denominados *Forward-Looking Centers* ou centros prospectivos e representados por $C_f(U_i)$. Esse conjunto representa os potenciais centros de atenção dos próximos enunciados. Por exemplo, no enunciado (U_1) do segmento S apresentado anteriormente, identificamos o seguinte conjunto de entidades $C_f(U_1) = \{\text{João, loja de música, piano}\}$. A ordenação do conjunto de entidades é de vital importância para a teoria. Um critério usual é a função gramatical subcategorizada pelo verbo principal: sujeito > objeto direto > objeto indireto > outras sub-categorizações > adjuntos. Vale ressaltar que esse critério de ordenação é aplicável em línguas de ordem quase fixa como inglês, francês e português, não tendo resultados satisfatórios em línguas como alemão e tcheco. Problemas na ordenação podem levar a escolha incorreta dos elementos de outros conjuntos.

O primeiro elemento do conjunto $C_f(U_i)$ é o mais altamente classificado e denominado *Preferred Center* ou próximo centro preferencial, sendo representado por $C_p(U_i)$. De acordo com a teoria, em um segmento com maior nível de coerência, o elemento C_p do

enunciado (U_i) deve ser o centro de atenção do enunciado (U_{i+1}). No exemplo do segmento S , $C_p(U_1) = \{\text{João}\}$, já que é o elemento mais altamente classificado em $C_f(U_i)$.

Outro elemento pertencente ao conjunto $C_f(U_i)$ é o *Backward-Looking Center* ou centro retrospectivo, que estabelece uma relação coerente com o enunciado imediatamente anterior e é representado por $C_b(U_i)$. Podemos dizer que $C_b(U_i)$ é uma referência a um elemento de $C_f(U_{i-1})$, desde que $i > 1$. De acordo com a teoria, em um segmento com maior nível de coerência, o elemento $C_b(U_i)$ é o mesmo de $C_p(U_{i-1})$. No exemplo do segmento S , $C_b(U_2) = \{\text{João}\}$, já que no enunciado (U_2) o pronome “Ele” faz uma referência à João, que se encontra no conjunto $C_f(U_{i-1})$ como próximo centro preferencial.

Segundo Miltsakaki e Kukich (2004), espera-se que $C_b(U_i)$ e $C_p(U_{i-1})$ sejam idênticos em uma sequência de enunciados, não sendo essa uma condição obrigatória, porém satisfatória e que contribui para o grau de coerência local do discurso. Ainda segundo os autores, a relação entre esses conjuntos em enunciados subsequentes caracteriza sempre uma de quatro transições definidas na teoria: *Continue*, *Retain*, *Smooth-Shift* e *Rough-Shift*. Essas transições se dão pelas relações de equivalência entre os conjuntos C_b e C_p , como pode ser observado na Tabela 3.1.

Tabela 3.1: Relações de equivalência entre os conjuntos C_b e C_p e as transições da Teoria de *Centering*

	$C_b(U_i) = C_p(U_{i-1})$	$C_b(U_i) \neq C_p(U_{i-1})$
$C_b(U_i) = C_p(U_i)$	<i>Continue</i>	<i>Smooth-Shift</i>
$C_b(U_i) \neq C_p(U_i)$	<i>Retain</i>	<i>Rough-Shift</i>

A teoria considera que essas transições definem o grau de coerência do discurso, sendo que a transição *Continue* é preferível em relação à *Retain*, que por sua vez é preferível à *Smooth-Shift*, que é preferível em relação à transição *Rough-Shift*. No entanto, cabe ressaltar que embora um segmento formado somente por transições *Continue* seja considerado coerente pela teoria, sua leitura não é agradável, já que não há progressão.

Além das transições, a teoria de *Centering* apresenta um conjunto de restrições e regras quanto a forma como os centros de atenção podem ser utilizados para a composição de um texto coerente, sendo que algumas dessas informações já foram apresentadas anteriormente:

Restrições:

1. Cada enunciado (U_i) possui somente um C_b ;
2. Todos os elementos de $C_f(U_i)$ são realizados em (U_i) ; e
3. O $C_b(U_i)$ é o elemento melhor classificado em $C_f(U_{i-1})$ e que é realizado no enunciado (U_i) .

Regras:

1. Se um elemento de $C_f(U_{i-1})$ é realizado por um pronome em (U_i) , então $C_b(U_i)$ também deve ser realizado por um pronome; e
2. As transições possuem uma ordem de preferência:

$$Continue > Retain > Smooth-Shift > Rough-Shift.$$

Dadas as definições da teoria, podemos investigar a afirmação feita no início desta seção em relação ao segmento S . A seguir, nas Tabelas 3.2, 3.3, 3.4, 3.5 e 3.6, apresentamos as entidades identificadas em cada enunciado do segmento e suas respectivas transições. Vale lembrar que as transições ocorrem entre pares de enunciados adjacentes ao verificar a relação entre os conjuntos C_b e C_p dos mesmos, de acordo com a Tabela 3.1.

Tabela 3.2: Entidades e transição identificada no enunciado (U_1)

João foi à sua loja de música favorita para comprar um piano.		
$C_f(U_1) = \{\text{João, loja de música, piano}\}$	$\parallel$	$C_p(U_1) = \{\text{João}\} \parallel C_b(U_1) = \emptyset$
Transição = $\emptyset$		

Tabela 3.3: Entidades e transição identificada no enunciado (U_2)

Ele frequentava a loja há vários anos. (Ele = João)		
$C_f(U_2) = \{\text{João, loja de música}\}$	$\parallel$	$C_p(U_2) = \{\text{João}\} \parallel C_b(U_2) = \{\text{João}\}$
Transição($U_1 - U_2$) = <i>Continue</i>		

Tabela 3.4: Entidades e transição identificada no enunciado (U_3)

Estava animado por finalmente poder comprar um piano. (Estava = João)		
$C_f(U_3) = \{\text{João, piano}\}$	$\parallel$	$C_p(U_3) = \{\text{João}\} \parallel C_b(U_3) = \{\text{João}\}$
Transição($U_2 - U_3$) = <i>Continue</i>		

Tabela 3.5: Entidades e transição identificada no enunciado (U_{4a})

Ele chegou justo quando a loja fechava. (Ele = João)		
$C_f(U_{4a}) = \{\text{João, loja de música}\}$	$C_p(U_{4a}) = \{\text{João}\}$	$C_b(U_{4a}) = \{\text{João}\}$
Transição($U_3 - U_{4a}$) = <i>Continue</i>		

Tabela 3.6: Entidades e transição identificada no enunciado (U_{4b})

Ela estava fechando assim que João chegou. (Ela = loja de música)		
$C_f(U_{4b}) = \{\text{loja de música, João}\}$	$C_p(U_{4b}) = \{\text{loja de música}\}$	$C_b(U_{4b}) = \{\text{loja de música}\}$
Transição($U_3 - U_{4b}$) = <i>Smooth-Shift</i>		

Desse modo, a afirmação anteriormente apresentada de que o segmento S é intuitivamente mais coerente se finalizado com o enunciado (U_{4a}) em vez de (U_{4b}) se comprova na teoria de *Centering* pelo fato da transição entre $U_3 - U_{4a}$ ser do tipo *Continue*, como visto na Tabela 3.5, enquanto a transição entre $U_3 - U_{4b}$ é do tipo *Smooth-Shift*, como visto na Tabela 3.6.

Em resumo, a Teoria de *Centering* postula que em um segmento de discurso coerente, os centros de atenção dos enunciados que o constituem são sobre uma mesma entidade (relação *Continue*) antes de introduzir uma nova entidade relacionada (relação *Retain*), a qual poderá vir a ser o novo centro de atenção de um enunciado (relações *Smooth-Shift* ou *Rough-Shift*). Logo, é uma teoria relacionada à coerência local, já que existe um determinado critério para o agrupamento de uma sequência de enunciados.

3.2 Modelo de Entidades

O Modelo de Entidades ou *Entity-grid* (Barzilay e Lapata, 2008) também tem seu foco na coerência local do texto, uma vez que analisa o relacionamento das transições entre sentenças adjacentes. Embora trate de coerência local, os autores afirmam que esse nível é incontestavelmente necessário para se conseguir a coerência global.

A premissa fundamental do modelo é de que textos com coerência local possuem certas regularidades na distribuição de substantivos e pronomes, chamados de entidades. Tal premissa é baseada em teorias como a *Centering* e em outras teorias de discurso baseada em entidades (e.g., Givón (1987); Prince (1981)).

No modelo, cada texto é representado por uma matriz (também denominada de *grid*) que captura a distribuição das entidades de discurso das sentenças, em que as linhas correspondem às sentenças e as colunas às entidades. Para cada ocorrência de uma entidade no texto, a célula correspondente da matriz contém informações sobre sua presença ou ausência em uma sequência de sentenças. Além disso, em entidades presentes em uma determinada sentença, a célula da matriz contém informações sobre o papel gramatical do sintagma nominal. Dessa maneira, cada célula da matriz é preenchida por uma letra que representa se a entidade em questão possui o papel de sujeito (S), objeto (O) ou nenhum (X) no sintagma nominal. Entidades ausentes de uma sentença são sinalizadas por lacunas (-).

Considere o exemplo de Barzilay e Lapata (2008), que apresenta um texto de seis sentenças com suas anotações sintáticas já realizadas:

1. [The Justice Department]_S is conducting an [anti-trust trial]_O against [Microsoft Corp.]_X with [evidence]_X that [the company]_S is increasingly attempting to crush [competitors]_O.
2. [Microsoft]_O is accused of trying to forcefully buy into [markets]_X where [its own products]_S are not competitive enough to unseat [established brands]_O.
3. [The case]_S revolves around [evidence]_O of [Microsoft]_S aggressively pressuring [Netscape]_O into merging [browser software]_O.
4. [Microsoft]_S claims [its tactics]_S are commonplace and good economically.
5. [The government]_S may file [a civil suit]_O ruling that [conspiracy]_S to curb [competition]_O through [collusion]_X is [a violation of the Sherman Act]_O.
6. [Microsoft]_S continues to show [increased earnings]_O despite [the trial]_X.

A partir dessas seis sentenças, a Figura 3.1 representa um fragmento da matriz de entidades correspondente.

	Department	Trial	Microsoft	Evidence	Competitors	Markets	Products	Brands	Case	Netscape	Software	Tactics	Government	Suit	Earnings	
1	S	O	S	X	O	-	-	-	-	-	-	-	-	-	-	1
2	-	-	O	-	-	X	S	O	-	-	-	-	-	-	-	2
3	-	-	S	O	-	-	-	-	S	O	O	-	-	-	-	3
4	-	-	S	-	-	-	-	-	-	-	-	S	-	-	-	4
5	-	-	-	-	-	-	-	-	-	-	-	-	S	O	-	5
6	-	X	S	-	-	-	-	-	-	-	-	-	-	-	O	6

Figura 3.1: Representação da matriz de entidades (Barzilay e Lapata, 2008)

Como o exemplo possui seis sentenças, a matriz representada na Figura 3.1 também possui seis linhas. Considere a coluna da matriz referente a entidade *Trial*, [O - - - - X]. Esse registro representa que a entidade *Trial* está presente nas sentenças 1 e 6 (como O e X, respectivamente), mas ausente nas outras sentenças. É importante destacar que a matriz leva em consideração a resolução de correferência. Mesmo que uma entidade ocorra de diferentes formas linguísticas no texto, por exemplo, *Microsoft Corp.*, *Microsoft* e *the company*, ela é mapeada como uma entrada única na matriz, como é o caso da coluna *Microsoft* [S O S S - S].

Outra particularidade do processo de preenchimento da matriz ocorre quando uma mesma entidade é referenciada mais de uma vez na mesma sentença e com diferentes funções gramaticais. O padrão é escolher para ser uma entrada na matriz de entidades a entidade que possuir a função gramatical com maior classificação, sendo considerada a seguinte ordenação: sujeito > objeto > outras funções gramaticais. Por exemplo, a entidade *Microsoft* é mencionada duas vezes na Sentença 1 com funções gramaticais diferentes, [*Microsoft Corp.*]_X e [*the company*]_S, mas é representada somente por *S* na célula correspondente da matriz, já que sujeito (*S*) é a função gramatical mais altamente classificada.

O modelo considera que matrizes de textos coerentes são suscetíveis a possuírem algumas colunas densas, isto é, em que a maior parte de seus elementos são diferentes de lacunas (-), como é o caso da coluna da entidade *Microsoft* da Figura 3.1, e muitas colunas esparsas constituídas principalmente de lacunas, como as entidades *Markets* e *Earnings* da Figura 3.1. Também se espera que as colunas densas possuam mais entidades do tipo sujeito ou objeto, pois essas características são menos presentes em textos com baixa coerência.

Inspirada na teoria de *Centering*, o modelo gira em torno dos padrões das transições de entidades locais. Uma transição de entidade local é uma sequência {S, O, X, -}ⁿ que

representa a ocorrência de entidades e seus papéis sintáticos em n sentenças adjacentes. As transições locais podem ser facilmente obtidas a partir da matriz de transições como subsequências contínuas de cada coluna e que possuem uma certa probabilidade de ocorrência na matriz. Por exemplo, a probabilidade de ocorrer a transição [S -] na matriz da Figura 3.1 é 0.08, já que a transição ocorre seis vezes em um total de 75 possíveis transições de tamanho 2, logo: $\frac{6}{75} = 0.08$.

Dessa maneira, cada texto pode ser visto como uma distribuição definida de transições locais, representada por uma notação de vetor de características. O vetor de características de uma matriz de entidades x_{i_j} de um texto ou documento d_i é representado por:

$$\Phi(x_{i_j}) = (p_1(x_{i_j}), p_2(x_{i_j}), \dots, p_m(x_{i_j}))$$

em que m é o número de todas as transições pré-definidas de entidades e $p_t(x_{i_j})$ é a probabilidade da transição t na matriz x_{i_j} .

A representação por vetor de características é muito útil pois permite a utilização desses dados em algoritmos de aprendizagem de máquina, que por sua vez podem descobrir novos padrões relevantes de distribuição de entidades a serem utilizados na avaliação de coerência.

Como se pode notar, existe um grande número de possíveis transições que podem ser observadas no vetor de características. Uma forma de limitar a quantidade de observações é limitar o número de transições, normalmente em duas ou três, ou verificar apenas as transições mais frequentes nos documentos.

Na Figura 3.2 é apresentado um exemplo com espaço de transições de tamanho dois, em que a primeira linha (d_1) representa as probabilidades verificadas no exemplo da matriz de entidades da Figura 3.1, enquanto o restante das linhas ilustram os valores de outros documentos (d_2 e d_3).

	SS	SO	SX	S-	OS	OO	OX	O-	XS	XO	XX	X-	-S	-O	-X	--
d_1	.01	.01	0	.08	.01	0	0	.09	0	0	0	.03	.05	.07	.03	.59
d_2	.02	.01	.01	.02	0	.07	0	.02	.14	.14	.06	.04	.03	.07	0.1	.36
d_3	.02	0	0	.03	.09	0	.09	.06	0	0	0	.05	.03	.07	.17	.39

Figura 3.2: Exemplo de vetor de características com todas as transições envolvendo duas entidades e as categorias sintáticas {S, O, X, -} (Barzilay e Lapata, 2008)

A forma de representação proposta pelo modelo permite extrair propriedades de textos coerentes a partir de um corpus, sem recorrer a anotação manual ou consulta à uma base de conhecimento pré-definida. No trabalho de Barzilay e Lapata (2008), os

autores demonstram o uso do modelo em três diferentes aplicações utilizando um corpus formado por textos jornalísticos: ordenação de textos, avaliação automática de coerência em resumos e avaliação de legibilidade.

Trabalhos como o de Burstein et al. (2010), combinam o Modelo de Entidades com características relacionadas a erros gramaticais, vocabulário, estilo e outras características de qualidade de escrita na avaliação de redações (*essays*) de estudantes não-nativos em inglês e vem conseguindo resultados promissores que podem ser incorporados nos modelos de avaliação de ferramentas de AES. O algoritmo desenvolvido pelos autores é baseado no algoritmo BL08 (Barzilay e Lapata, 2008), em que é gerado um vetor de características com os valores das probabilidades das transições, como apresentado na Figura 3.2, se houve ou não a resolução de correferência e se tal entidade é mencionada duas ou mais vezes. O algoritmo também extrai informações relacionadas à qualidade de escrita que normalmente fazem parte de ferramentas de AES como a presença de erros ortográficos/gramaticais e a frequência de palavras com conteúdo que não fazem parte do tópico do texto. Tais informações são utilizadas como atributos dos textos analisados e permitem a sua comparação com textos previamente anotados por humanos em relação à coerência, por meio de algoritmos de aprendizagem de máquina.

3.3 Teoria da Estrutura Retórica

A Teoria da Estrutura Retórica (*Rhetorical Structure Theory: RST*) (Mann e Thompson, 1987) é uma teoria discursiva cujo objetivo é descrever a organização do texto por meio de relações entre suas partes funcionais. Essas partes são porções de texto chamadas de *text spans*, constituídas de intervalos lineares e ininterruptos de texto, e o conjunto de relações definidas entre essas partes forma a estrutura retórica do texto.

Segundo Mann e Thompson (1987), a RST fornece meios de se obter o conteúdo proposicional implícito de um texto por meio de relações proposicionais entre suas partes. Ainda segundo os autores, essas proposições são essenciais para a coerência do texto, de tal forma que, se um texto for coerente, será sempre possível construir sua estrutura retórica. Por esse motivo, as relações retóricas também são chamadas relações de coerência.

Na RST original foram definidas 26 relações retóricas para relacionar as proposições expressas em um texto. Essas relações se estabelecem entre duas ou mais proposições de segmentos adjacentes do texto, sendo uma nuclear (N), que indica a informação principal, e outra informação adicional complementar, chamada de satélite (S). Quando ambas as informações são igualmente importantes, tem-se uma relação multinuclear, em que se tem mais de um núcleo e nenhum satélite.

O conjunto dessas relações e suas nuclearidades são apresentadas na Tabela 3.7.

Tabela 3.7: Relações e nuclearidade da RST (Mann e Thompson, 1987)

Relação	Multinuclear
Antithesis	Não
Background	Não
Circumstance	Não
Concession	Não
Condition	Não
Elaboration	Não
Enablement	Não
Evaluation	Não
Evidence	Não
Interpretation	Não
Justify	Não
Means	Não
Motivation	Não
Non-volitional Cause	Não
Non-volitional Result	Não
Otherwise	Não
Purpose	Não
Restatement	Não
Solutionhood	Não
Summary	Não
Volitional Cause	Não
Volitional Result	Não
Contrast	Sim
Joint	Sim
List	Sim
Sequence	Sim

Para o estabelecimento de uma relação entre duas proposições, Mann e Thompson (1987) definem que quatro campos devem ser observados e que o analista deve ter seu julgamento baseado no contexto e nas intenções do escritor:

1. Restrições sobre o núcleo (N);
2. Restrições sobre o satélite (S);
3. Restrições sobre a combinação do núcleo e do satélite (N+S); e
4. Efeito que a relação em questão causa no leitor.

Vamos considerar, por exemplo, a relação *Purpose*:

1. Restrições sobre (N): apresenta uma ação;
2. Restrições sobre (S): apresenta uma situação não realizada;
3. Restrições sobre (N+S): S apresenta uma situação que pode realizar N; e
4. Efeito: o leitor reconhece que a atividade em N pode ser iniciada por meio de S.

Dessa maneira, existe um conjunto de restrições para cada uma das 26 relações estabelecidas na RST que pode ser consultado na obra de Mann e Thompson (1987). O mesmo conjunto traduzido para o português pode ser consultado em Pardo (2005).

Apresentamos os exemplos descritos em Pardo (2005) para exemplificar o uso de algumas das relações. No trecho de texto “Embora você não goste, trabalhar é importante.”, a proposição expressa pela primeira oração é o satélite e a proposição expressa pela segunda oração é o núcleo da relação retórica de oposição *Concession*. A estrutura retórica construída pela RST também pode ser representada pelo formalismo de árvores, como pode ser observado na Figura 3.3, que ilustra o exemplo apresentado anteriormente.

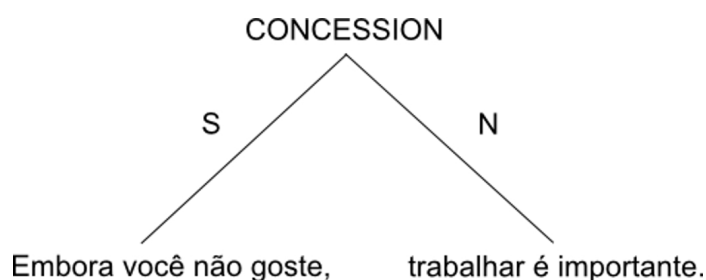


Figura 3.3: Exemplo de estrutura retórica RST (Pardo, 2005)

Ainda no exemplo apresentado na Figura 3.3, se adicionarmos a oração “O trabalho enobrece o homem.” às sentenças já existentes, temos uma nova relação, *Justify*, em que a nova oração é a satélite e as proposições que formam a relação *Concession* são o núcleo da relação *Justify*, formando uma hierarquia de relações. Esta hierarquia formada pelas relações *Concession* e *Justify* é apresentada na Figura 3.4.

3.4 Análise de Semântica Latente

A Análise de Semântica Latente (*Latent Semantic Analysis: LSA*) (Landauer et al., 1998) é um método estatístico para a extração e representação de conhecimento aplicado em corpus, originalmente desenvolvido na *Bell Laboratories* e *University of Colorado*¹ no contexto de recuperação de informações e recentemente utilizada em um grande número de aplicações.

O método simula o processo de aquisição de conhecimento de forma associativa, assim como uma pessoa associa ações/percepções/palavras com seus respectivos contextos e forma uma grande base de relações semânticas (Kintsch, 2002).

A ideia básica da LSA é formar um espaço semântico em que a semelhança entre os termos se dá pela ocorrência em contextos comuns. Por exemplo, dadas ocorrências como “O médico curou o paciente.” e “O cirurgião operou o paciente.”, as palavras “médico” e “cirurgião” podem ser consideradas similares, já que ocorrem no mesmo contexto com a palavra “paciente” (Mani, 2001). Para isso, a LSA extrai, a partir da análise de um corpus, uma matriz formada por termos e suas respectivas quantidades de ocorrências nos documentos (contextos) do corpus, como pode ser visto no exemplo da Tabela 3.8.

Tabela 3.8: Exemplo de matriz de co-ocorrência de termos

	Documento 1	Documento 2	Documento 3	Documento N
Termo 1	0	2	1	...
Termo 2	4	3	5	...
Termo 3	3	4	2	...
Termo K	...	...	...	...

Embora seja apresentada na Tabela 3.8 uma comparação apenas entre termos e documentos, o método permite comparações entre combinações de termos, formando sentenças ou parágrafos.

Após a construção da matriz, o valor de cada célula é submetido a uma função de transformação que atribui um peso a cada entrada de acordo com sua importância em relação às outras entradas. Essa função é também conhecida como normalização e é obtida por meio do modelo TF-IDF (*term frequency – inverse document frequency*). O modelo é constituído basicamente de duas etapas. Na primeira etapa (TF) atribui-se um peso às entradas de acordo com a frequência de termos em um único documento. Esse cálculo é realizado por meio de um processo simples que envolve a divisão da frequência

¹<http://lsa.colorado.edu/>

de cada termo pelo total de termos do documento. Por exemplo, considere um documento contendo 100 palavras em que determinado termo ocorre 3 vezes. O peso TF desse termo será $(3/100) = 0,03$. A segunda etapa é a indexação pela frequência inversa (IDF) dos documentos por meio do emprego da fórmula:

$$IDF = \log\left(\frac{N}{n_k}\right)$$

em que N é o número de documentos do corpus e n_k é o número de documentos em que o termo k ocorre. Por exemplo, considere 10 milhões de documentos e que determinado termo ocorre em 1.000 destes documentos do corpus, logo o peso IDF é calculado como $\log\left(\frac{10.000.000}{1.000}\right) = 4$. Com a combinação das duas etapas obtém-se o valor TF-IDF utilizado para calcular o peso de cada termo nos documentos:

$$TF-IDF = TF * IDF$$

Após normalizar as entradas da matriz, uma técnica matemática derivada da álgebra linear e denominada *Singular Value Decomposition* – SVD (Golub e Reinsch, 1970) é utilizada para reduzir a dimensionalidade da matriz e encontrar os principais padrões associativos nos dados, ignorando as influências menos importantes. A SVD fornece uma medida natural, por meio da decomposição da matriz, e não uma medida artificialmente construída (pré-estabelecida), como ocorre em um modelo simbólico, por exemplo.

Descrevemos a seguir todo processo realizado pela LSA utilizando a SVD de acordo com Deerwester et al. (1990).

Considere uma matriz de termos como a apresentada na Tabela 3.8 e com seus valores normalizados pelo modelo TF-IDF, em que:

- t = número de termos, ou linhas
- d = número de documentos, ou colunas
- X = matriz $t \times d$

A matriz X é decomposta em um produto de outras três matrizes, $X = TSD$, sendo que:

- m = número de dimensões, $m \leq \min(t, d)$
- T = matriz de vetores singulares à esquerda, $t \times m$
- S = matriz diagonal $m \times m$ de valores singulares em ordem decrescente,
- D = matriz de vetores singulares à direita, $m \times d$

A Figura 3.6 apresenta esquematicamente o processo realizado pela SVD de decomposição da matriz *termos* x *documentos* em outras três matrizes (TSD), como descrito por Deerwester et al. (1990).

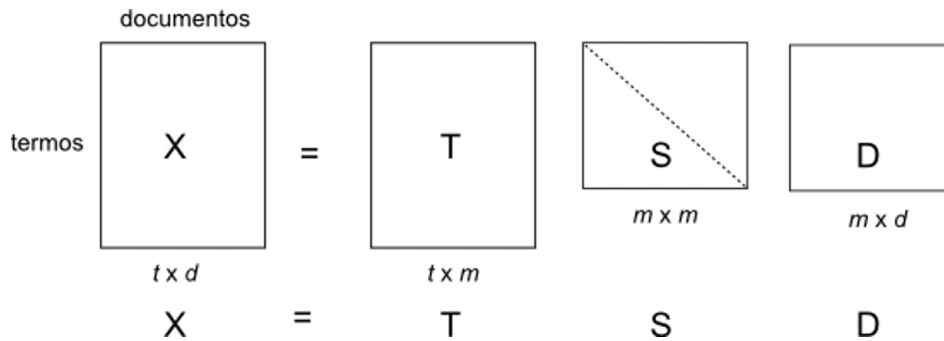


Figura 3.6: Matriz original X como um produto de três matrizes (Deerwester et al., 1990)

O próximo passo é ilustrado na Figura 3.7, em que a LSA reduz a dimensão dessas matrizes, eliminando as linhas e colunas correspondentes aos menores valores da matriz S , assim como as colunas da matriz T e as linhas da matriz D . Para a redução das dimensões, reduz-se o número m de dimensões para um valor $k < m$ e, dessa maneira, a matriz reduzida S' afeta diretamente nas dimensões das matrizes T e D .

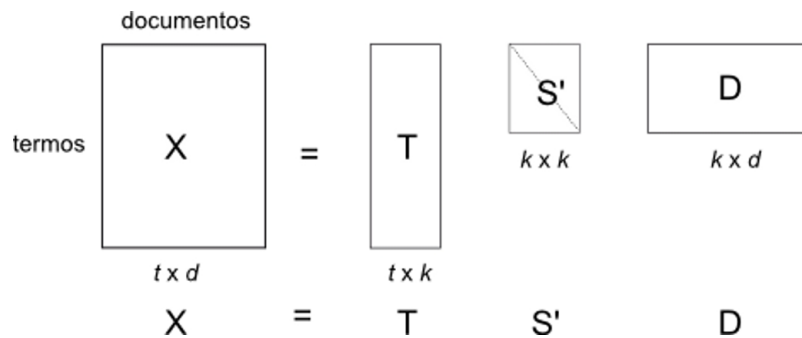


Figura 3.7: Processo de redução de dimensões da matriz criada pela SVD (Deerwester et al., 1990)

Com o produto de $TS'D$ obtém-se os elementos mais significativos da matriz e com um custo computacional reduzido para seu processamento, já que matrizes com um número grande de dimensões podem necessitar de horas ou até dias para seu processamento (Miller, 2003).

Encontrar o número correto de dimensões k para ser utilizado na redução de dimensionalidade do vetor S é uma tarefa crucial e difícil para a SVD. Se o número de dimensões for considerado pequeno, pode-se perder informações importantes e, se

considerado grande, pode-se adicionar informações sem relevância. Trabalhos empíricos envolvendo corpus volumosos apresentam valores entre 200 e 400 dimensões (Landauer et al., 1998; Wade-Stein e Kintsch, 2004).

A partir da representação em forma de vetores permitida pela LSA, pode-se medir a similaridade de conceitos relacionados entre duas palavras ou sentenças, calculando-se o produto co-seno (ou produto cartesiano) entre os vetores que representam tais palavras, sentenças ou textos. Ljungstrand e Johansson (1998) definem essa medida de similaridade por meio da fórmula que segue:

$$\text{sim}(X, Y) = \frac{\sum_{i=1}^n X_i Y_i}{\sqrt{\sum_{i=1}^n X_i^2 \sum_{i=1}^n Y_i^2}}$$

em que $X = (x_1, x_2, \dots, x_n)$ e $Y = (y_1, y_2, \dots, y_n)$ são dois vetores com n dimensões que representam os textos a serem comparados utilizando o modelo *bag of words*. O resultado apresenta valores limitados entre $[-1, 1]$, sendo -1 o menor valor possível de similaridade e 1 o mais alto grau de similaridade.

A Figura 3.8 ilustra a comparação entre três vetores que possuem respectivamente as palavras “casa”, “sacada” e “corrida”. Observa-se que as palavras “casa” e “sacada” possuem .65 de similaridade, enquanto as palavras “casa” e “corrida”, assim como “sacada” e “corrida” possuem .01 de similaridade. Esses valores se devem ao fato das palavras “casa” e “sacada” estarem presentes em contextos semelhantes, ao contrário dos outros pares de palavras.

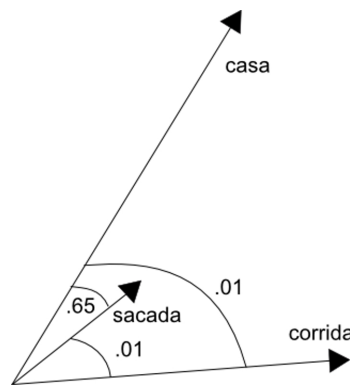


Figura 3.8: Comparação LSA entre três palavras (Kintsch, 2002)

A similaridade entre duas cadeias pode ser alta mesmo que existam poucas palavras em comum entre as cadeias. Considere o exemplo em que são analisadas três sentenças,

duas com conceitos relacionados, mas com palavras distintas, e uma terceira sentença com conceito não relacionado às outras duas sentenças, mas com uma palavra semelhante a uma delas. Na Tabela 3.9 são apresentados os valores finais² (co-seno) da comparação entre os vetores de termos.

Tabela 3.9: Valores atribuídos pela LSA na comparação entre três sentenças

	“O raio da esfera”	“O diâmetro do círculo”	“A música das esferas”
“O raio da esfera”	1	0.66	0.08
“O diâmetro do círculo”	0.66	1	0.15
“A música das esferas”	0.08	0.15	1

Na Tabela 3.9, a diagonal da matriz apresenta os valores da comparação entre as mesmas sentenças e por isso o nível máximo de similaridade é alcançado com o valor 1. Também se pode observar que a comparação entre as sentenças “O raio da esfera” e “O diâmetro do círculo” apresentam elevada similaridade (0.66), já que são sentenças de um mesmo contexto. Por outro lado, ao comparar a primeira sentença com a terceira, “A música das esferas”, a similaridade é de apenas 0.08, significando que, embora as sentenças apresentem a mesma palavra “esfera”, o seu uso é utilizado em contextos diferentes.

Como já dito, inicialmente a LSA foi desenvolvida no contexto de Recuperação de Informação com a *Latent Semantic Indexing* (LSI), mas recentemente tem sido aplicada nas mais diversas aplicações, como em ferramentas de *Automated Essay Scoring*, em que, por exemplo, se compara o texto avaliado com um texto base escrito por um profissional e a nota emitida leva em consideração a similaridade semântica entre os textos.

A LSA possui algumas limitações, como o fato de não considerar a ordem das palavras, relações sintáticas ou lógicas, nem morfológicas. Assim, frases do tipo “João ama Maria” e “Maria ama João” são interpretadas pela LSA como sendo iguais e o valor do co-seno atribuído a elas é 1, o que não é verdade. Mesmo com tais limitações, a LSA vem apresentando bons resultados nos contextos em que é aplicada.

²Valores da análise em inglês, obtidos no endereço: <http://lsa.colorado.edu/cgi-bin/LSA-matrix.html>. Os valores obtidos variam de acordo com o corpus. Acesso em 10/05/2010.

Análise Automática de Textos

Apresentamos neste capítulo ferramentas computacionais que realizam a análise automática de textos e que possuem algum critério de análise que julgamos relacionado à coerência, servindo como base para a realização deste trabalho.

4.1 CriterionTM

O *CriterionTM Online Essay Evaluation Service* (Burstein et al., 2003a; Higgins et al., 2004; Miltasakaki e Kukich, 2000a,b, 2004) desenvolvido pela empresa *Education Testing Service* – ETS¹, é um sistema educacional *online* com o objetivo de realizar a análise automática de redações (*essays*) escritas por estudantes de diferentes níveis escolares, retornando uma nota e *feedback* com sugestões específicas ao escritor e que podem melhorar o texto. Para isso, o *CriterionTM* conta com duas ferramentas complementares que podem ser utilizadas em conjunto ou separadamente. São elas:

E-raterTM Scoring Engine (Attali e Burstein, 2006): ferramenta responsável por extrair características linguísticas do texto a ser avaliado. Usa um modelo estatístico para comparar o texto escrito com uma base de referência, sendo possível atribuir uma nota em relação a aspectos como variedade sintática, conteúdo temático e coerência.

CritiqueTM (Burstein et al., 2004): ferramenta composta por um conjunto de programas responsáveis por avaliar erros de gramática, falta de pontuação, concordância

¹<http://www.ets.org/>

verbal, uso incorreto de determinadas palavras, identificar a estrutura do texto, reconhecer o uso de características linguísticas indesejáveis, entre outras características do texto.

Como parte do *Criterion*TM, Miltsakaki e Kukich (2000a,b, 2004) realizaram um estudo sobre a identificação das transições da teoria de *Centering* como um modelo indicador de incoerências locais no texto, aplicado ao *E-rater*TM. Embora os autores afirmem alguma melhora de desempenho na aplicação, também comentam que o processo depende fortemente da melhoria de outras técnicas, como as utilizadas para a identificação de sintagmas nominais e resolução de anáforas. Tais técnicas são essenciais para a correta identificação dos centros de atenção e, por consequência, das transições definidas pela teoria de *Centering*.

Em Higgins et al. (2004), os autores afirmam que os aspectos avaliados pelo *Criterion*TM relacionados à coerência levam em conta a similaridade semântica das sentenças e a estrutura do texto. Neste ponto, vale lembrar que o *Criterion*TM avalia redações escritas a partir de um tópico específico (*Prompt*) e a estrutura identificada se baseia na proposta de Burstein et al. (2003b), em que as sentenças são classificadas em uma de cinco categorias retóricas: *Background*, *Thesis*, *Support*, *Main Idea* e *Conclusion*. Após classificar as sentenças, verificam-se quatro dimensões, a saber:

- Dim_P (Análise do relacionamento com o *Prompt*): determinadas sentenças do texto devem ser semanticamente relacionadas ao tópico proposto para a escrita da redação. Em especial, sentenças classificadas como *Thesis*, *Main Idea* e *Conclusion* devem ser altamente relacionadas ao tópico. Se a similaridade for muito baixa ou não existir, pode ser um indicativo de fuga do assunto proposto, o que caracteriza um problema de coerência.
- Dim_T (Análise do relacionamento com a *Thesis*): determinadas sentenças do texto devem ser semanticamente relacionadas entre si. Em especial, sentenças classificadas como *Background*, *Main Idea* e *Conclusion* devem ser altamente relacionadas às sentenças de *Thesis*. Vale lembrar que diferentemente do *Prompt*, as sentenças classificadas como *Thesis* são escritas pelo autor no decorrer do texto e expressa a sua posição em relação ao tópico proposto. Se entre essas sentenças a similaridade for muito baixa ou não existir, pode ser um indicativo de baixa conexão entre partes importantes do texto, o que pode acarretar em problemas relacionados à coerência global.
- Dim_S (Análise do relacionamento dentro do segmento): diferentes sentenças classificadas em uma mesma categoria no mesmo segmento (adjacentes) devem ser

semanticamente relacionadas. A baixa similaridade entre sentenças adjacentes e de mesma categoria retórica pode ser um indicador de problemas relacionados à coerência local.

- Dim_{ERR} (Análise de erros técnicos): independente da categoria da sentença, espera-se que ela esteja livre de erros técnicos.

Em todas as dimensões, dois valores possíveis podem ser atribuídos: *low* ou *high*, como pode ser visto no exemplo de Higgins et al. (2004) apresentado na Figura 4.1, em que anotadores humanos realizaram anotações em relação às dimensões Dim_P e Dim_T nos segmentos de discurso de uma redação escrita por um estudante a partir do *Prompt* fornecido ao escritor. Em casos em que a anotação não se aplica, por exemplo, o relacionamento com a *Thesis* (Dim_T) em uma sentença classificada como *Thesis*, é atribuído o valor *N/A*. Antes da anotação em *low* ou *high*, é necessário a classificação de cada sentença em uma das cinco categorias retóricas utilizadas pela ferramenta.

Discourse Segment	Sentence	Dim_P	Dim_T
Prompt	Images of beauty—both male and female—are promoted in magazines, in movies, on billboards, and on television. Explain the extent to which you think these images can be beneficial or harmful.		
Background	A lot of people really care about how they look or how other people look.	Low	High
	A lot of people like reading magazines or watch t.v about how you can fix your looks if you don't like the way your looks are.	High	High
Thesis	People that care about how they look is because they have problems at home, their parents don't pay attention to them or even that they have a high self-esteem which that is not good.	Low	N/A
	A lot of people get to the extent of killing themselves just because they're not happy with there looks.	Low	N/A
Support	Many people go thru make-overs to experiment how they will look but, some people still don't like themself.	N/A	High
Main Point	The people that don't like themselves need some helps and they probably feel like that because they have told them oh! your ugly, you look like Blank! or maybe a guy never ask a her out.	Low	Low
Support	In case of a guy probably the same comments but he won't dare to ask a girl out because he feels that the girl is going to say no because of the way he looks.	N/A	High
	Things like this make people don't like each other.	N/A	High
Conclusion	I suggest that a those people out here that are not happy with their looks get some help.	Low	High
	Theirs alot of programs that you can get help.	Low	Low

Figura 4.1: Anotação realizada por um anotador humano em relação às dimensões Dim_P e Dim_T (Higgins et al., 2004)

Higgins et al. (2004) apontam que é relativamente raro encontrar uma sentença que não seja semanticamente relacionada com outra de mesma categoria no mesmo segmento, o que dificulta a construção de um classificador automático que exerça a tarefa da dimensão Dim_S , devido a falta de exemplos com o valor *low*.

A verificação de similaridade semântica de forma automática é então realizada nas dimensões Dim_P e Dim_T , que abordam aspectos relacionados a coerência de acordo com os

componentes *Prompt* e *Thesis*. Para tal tarefa, utiliza-se uma técnica semelhante a LSA, denominada *Random Indexing* (Kanerva et al., 2000; Sahlgren, 2002, 2005). Segundo os desenvolvedores do sistema *Criterion*TM, o uso de tal técnica se deu devido às restrições ao uso comercial da LSA (Higgins e Burstein, 2007). A partir de um conjunto de 989 redações anotadas manualmente em relação à essas dimensões e atributos provenientes da *Random Indexing*, foram gerados classificadores que exercem os seus papéis de forma automática. Já a Dim_{ERR} é baseada em regras e não depende da anotação manual do corpus.

A partir desse tipo de verificação, o *Criterion*TM retorna um conjunto de sugestões que auxiliam o escritor a melhorar o seu texto em relação aos aspectos de coerência levando em consideração a estrutura retórica identificada, como pode ser observado na Figura 4.2.

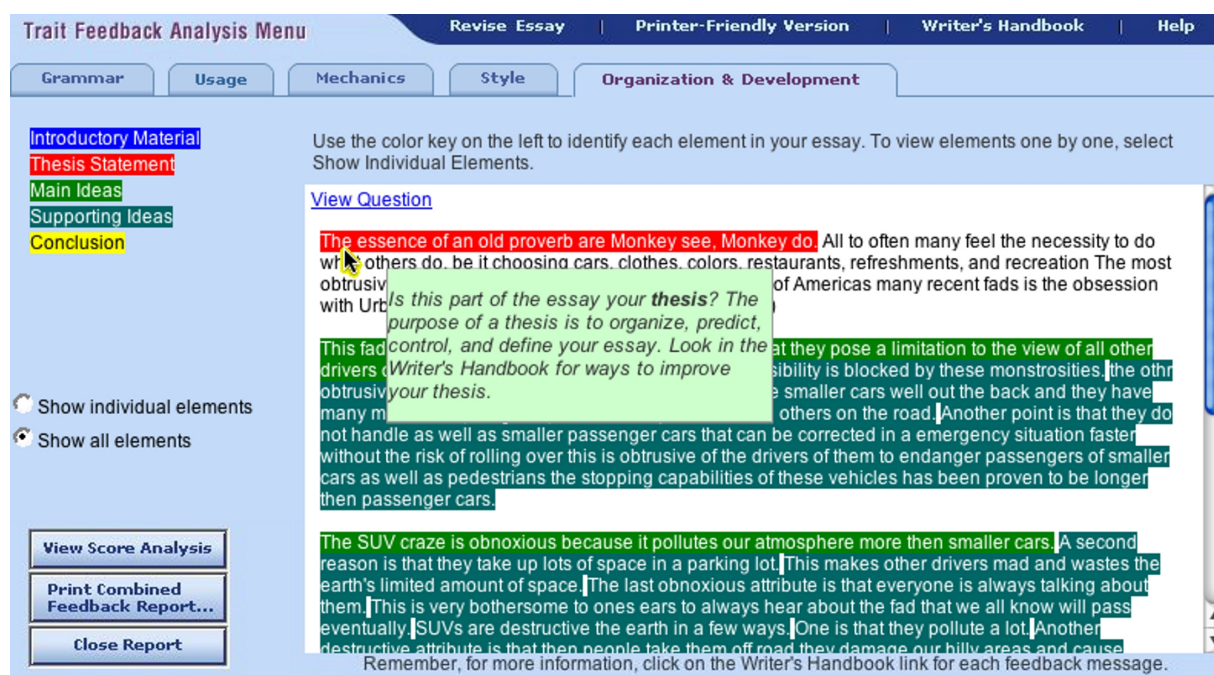


Figura 4.2: Exemplo de *feedback* apresentado pelo *Criterion*TM em relação aos aspectos de avaliação de coerência (Burstein et al., 2003a)

Na Figura 4.2 também é possível observar em sua parte superior que o sistema analisa outros aspectos do texto, tais como gramática, uso, mecanismos e estilo. Como já mencionado, o sistema também retorna uma nota (*score*) em relação a análise realizada por meio da ferramenta *E-rater*.

4.2 Knowledge Analysis Technologies™

O mecanismo *Knowledge Analysis Technologies* ou *KAT engine* foi desenvolvido pela empresa *Pearson Knowledge Analysis Technologies*², fundada em 1998 por Thomas Landauer, Peter Foltz e Darrell Laham, com o objetivo de desenvolver produtos comerciais que auxiliem estudantes no processo de escrita e de fala para não-nativos da língua inglesa.

O *KAT engine* avalia o significado do texto, examinando todas as passagens escritas, podendo ser aplicado em vários idiomas. Segundo a empresa³, o *KAT engine* é uma implementação da LSA com modificações, capaz de medir o nível de similaridade entre palavras ou sentenças e treinado utilizando um grande volume de textos significativos, sendo utilizada nos produtos desenvolvidos pela empresa. Neste trabalho, destacamos dois produtos que realizam a verificação da coerência entre o texto submetido à análise e suas referências. São eles:

Intelligent Essay Assessor™ (IEA) (Landauer et al., 2003a): é uma ferramenta que faz a análise e retorna uma nota para redações escritas por estudantes não-nativos em inglês. Embora o foco principal seja o conteúdo, o IEA também fornece resultados e comentários sobre aspectos formais do texto, como gramática, ortografia, vocabulário, uso da pontuação e estilo (Dikli, 2006). Para que o sistema funcione, o mesmo precisa ser treinado com um corpus de um domínio específico. Tal corpus deve ser representativo do domínio e receber uma classificação prévia em relação a sua qualidade (Landauer et al., 2003b). Dessa maneira, o texto avaliado é comparado com o corpus previamente classificado utilizando o *KAT engine* e também a um modelo de conhecimento, que auxilia na avaliação de estilo. Um exemplo⁴ do uso da ferramenta *Intelligent Essay Assessor™* é apresentado na Figura 6.1, em que notas de 1 à 6 são atribuídas à diferentes aspectos do texto, tais como ideias e conteúdo, organização, fluência das sentenças, variação de palavras e bom uso de convenções, além de comentários em relação a ortografia, redundância e gramática. Dessa maneira, observa-se que os aspectos relacionados à coerência do texto estão presentes na análise das ideias e conteúdo em um nível global e na fluência das sentenças em um nível local de coerência.

²<http://www.knowledge-technologies.com>

³Evidence for Reliability, Validity, and Learning Effectiveness, Disponível em: www.knowledge-technologies.com/papers/WTLReliabilityValidity-082007.pdf - Acesso em 11/05/2010

⁴www.knowledge-technologies.com/papers/IEA-FactSheet-20100401.pdf/ - Acesso em 12/05/2010

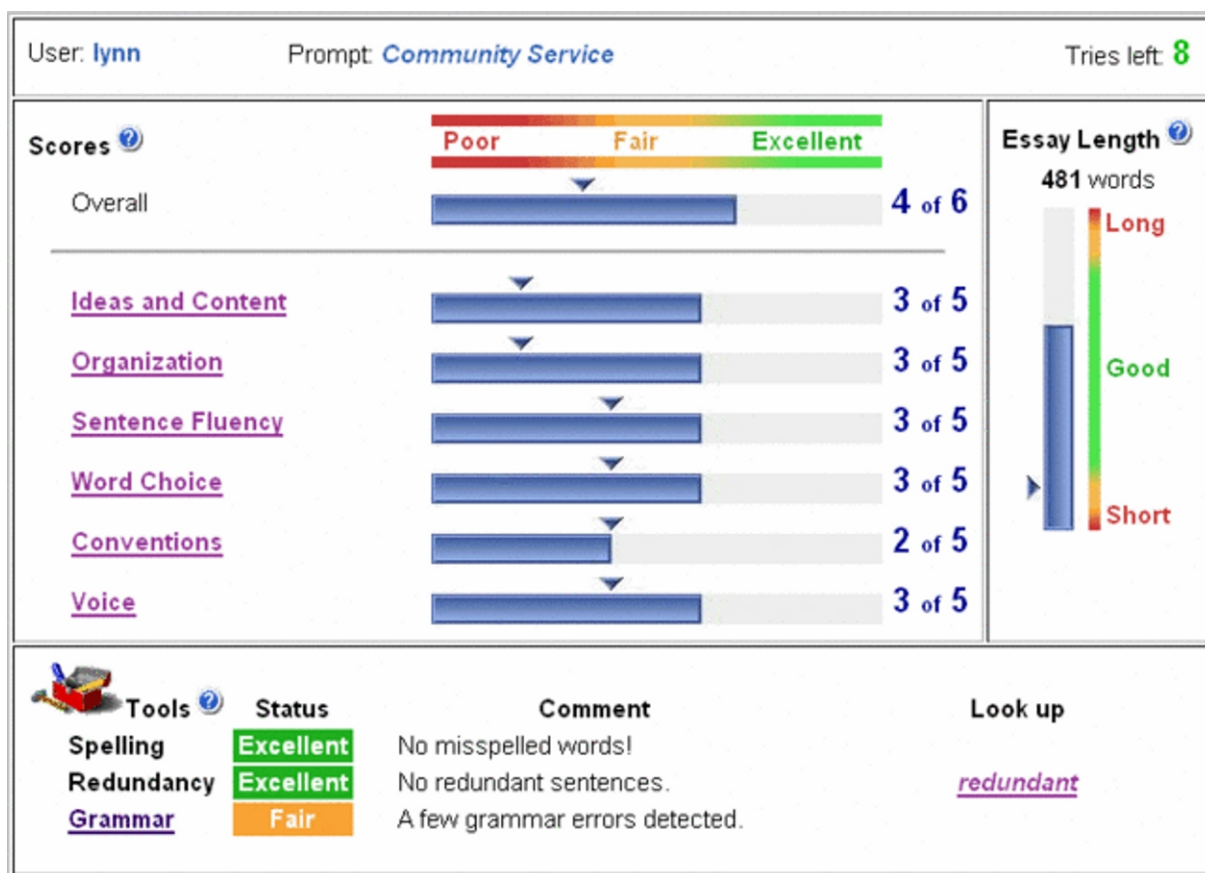


Figura 4.3: Exemplo de *feedback* apresentado pela *Intelligent Essay Assessor*TM

*Summary Street*TM (Franzke et al., 2005): é um sistema tutor inteligente que conta com uma relação de textos base sobre diversos temas e que podem ser consultados pelo usuário. Os temas abordados pela sistema incluem civilizações antigas (Maias, Incas, Astecas), fontes de energia (solar, nuclear, petróleo), sistema circulatório (coração, circulação do sangue), entre outros. Cada texto possui cerca de 2.000 palavras e é dividido em seções. O usuário submete um resumo com cerca de 200 palavras sobre determinado tópico, por exemplo, Energia Solar, que passará pela avaliação do sistema. Utilizando o *KAT engine*, o *Summary Street*TM verifica se o resumo aborda os tópicos principais de cada seção e se condensa cerca de 15% do texto base, se o estudante utilizou suas próprias palavras ou simplesmente copiou partes do texto, se existem sentenças irrelevantes que não adicionam valor ao conteúdo do resumo e se há repetição de ideias (Franzke et al., 2005). A ferramenta é direcionada tanto aos estudantes escritores quanto para seus respectivos professores, que podem consultar as atividades e avaliações realizadas por cada estudante de sua disciplina. Na Figura 4.4 é apresentado um exemplo⁵ de *feedback* da ferramenta

⁵www.knowledge-technologies.com/sstSampleStudent3.shtml - Acesso em 12/05/2010

Summary StreetTM para um estudante. São retornadas notas de 1 à 6 que classificam os aspectos avaliados em uma de três categorias *Poor*, *Fair* e *Excellent* (Ruim, Razoável e Excelente) de acordo com o nível de cobertura de cada tópico do texto submetido pelo estudante comparado ao texto base. Também são retornados comentários em relação a cópia de partes do texto base, ortografia, redundância e relevância das sentenças.



Figura 4.4: Exemplo de *feedback* apresentado pela *Summary StreetTM*

Na *Summary StreetTM*, a coerência é analisada a partir do nível de cobertura ao abordar cada tópico do texto base e da relevância das sentenças. Ou seja, o texto será coerente se o seu conteúdo for próximo ao texto base, mas sem possuir cópias e sentenças irrelevantes.

Tanto a *Intelligent Essay AssessorTM* como a *Summary StreetTM* permitem que o escritor retorne ao texto submetido para avaliação e faça as correções necessárias quantas vezes for desejado, se tornando um processo cíclico de melhoria do texto.

Além desses dois produtos, o mecanismo baseado na LSA é utilizado em outros produtos da empresa, como *WriteToLearn*^{TM6}, *CareerMap*^{TM7}, *Auto-autodidact*TM(*A2D*)⁸, *Knowledge Post*^{TM9} e *SuperManual*^{TM10}. Cada uma destas ferramentas tem o objetivo de auxiliar a escrita de textos e tratam de domínios específicos, como manuais técnicos de eletrônica.

4.3 IntelliMetricTM

A *IntelliMetric*TM (Elliot, 2003) é uma ferramenta de AES desenvolvida na *Vantage Laboratories*¹¹ e que vem sendo utilizada desde 2006 em substituição ao *E-rater*TM no exame de proficiência *Graduate Management Admission Test* (GMAT).

Por se tratar de uma ferramenta comercial pertencente a uma empresa que tem como estratégia corporativa manter em segredo os detalhes de seus tecnologias (Rudner et al., 2006), são poucas as informações sobre como a nota final é atribuída a uma redação, embora o trabalho de Elliot e Mikulas (2004) forneça algumas informações e destaque que a *IntelliMetric*TM possui um modelo de avaliação formado por cinco categorias: foco e coerência, organização, desenvolvimento e elaboração, estrutura das frases, mecanismos e convenções, em que se compara o texto submetido para avaliação à um exemplo denominado *gold standard*, escrito por um profissional de acordo com um tema ou tópico específico.

Essas categorias abordam cerca de 300 características sintáticas e semânticas e, para isso, são utilizadas técnicas de inteligência artificial, processamento de linguagem natural e estatística.

Baseado na *IntelliMetric*TM, a *Vantage Laboratories* desenvolveu um sistema *online* de auxílio à escrita denominado *MY Access!*^{TM12}, utilizado em salas de aula dos Estados Unidos e em empresas de teste, como: *The College Board*, *ACT*, *Harcourt Assessment, Inc.*, *CTB/McGraw Hill and Thomson Learning.*, *Microsoft Corp.*[®], *Apple*[®] *Computer*, *AOL*[®] e *Sun Microsystems*[®] (Rudner et al., 2006).

⁶<http://www.knowledge-technologies.com/prodWTL.shtml> - Acesso em 12/05/2010

⁷<http://www.knowledge-technologies.com/prodOther.shtml> - Acesso em 12/05/2010

⁸<http://www.knowledge-technologies.com/prodOther.shtml> - Acesso em 12/05/2010

⁹<http://www.knowledge-technologies.com/prodKP.shtml> - Acesso em 12/05/2010

¹⁰<http://www.knowledge-technologies.com/prodSM.shtml> - Acesso em 12/05/2010

¹¹<http://www.vantagelearning.com/school/products/intellimetric/> - Acesso em 12/05/2010

¹²<http://www.myaccess.com> - Acesso em 12/05/2010

4.4 Coh-Metrix

Embora não seja uma ferramenta que retorne notas em relação a um texto avaliado ou sugestões que possam melhorar a escrita, a *Coh-Metrix*¹³ (Graesser et al., 2004; McNamara et al., 2010, 2006) é uma ferramenta *online* que analisa textos escritos em inglês e retorna um conjunto de 62 índices referentes a coesão, coerência e dificuldade de compreensão do texto, a partir de vários níveis de análise linguística: léxico, sintático, discursivo e conceitual.

Os autores consideram que coesão é o grau em que as ideias do texto estão explicitamente relacionadas entre si e que coerência se refere a qualidade da representação mental construída pelo leitor durante o processo de leitura. Sendo assim, podemos associar a definição de coerência semântica apresentada neste trabalho à definição de coesão, apresentada pelos autores da *Coh-Metrix*.

Na Figura 4.5 apresentamos um exemplo com parte das métricas retornadas ao usuário pela ferramenta.

The screenshot shows the Coh-Metrix web interface. At the top, it says 'Welcome agraesser!' and 'Coh-Metrix'. Below this, there are input fields for 'Title' (The Needs of Plants), 'Source' (Research), 'User Code' (GraesserTest), 'Genre' (Science), 'LSA Space' (CollegeLevel), and 'CollegeLevel'. The main text area contains a passage about the needs of plants. To the right, there are two columns: 'Primary Measures' and 'Help Links'. The 'Primary Measures' column lists various metrics and their values, such as 'Coreference Cohesion Global 1' (0.589), 'Coreference Cohesion Global 2' (0.647), 'Coreference Cohesion Local 1' (0.75), 'Coreference Cohesion Local 2' (0.818), 'Causal Cohesion' (0.75), 'LSA Global' (0.576), 'LSA Local' (0.582), 'Reading Ease' (85.812), 'Reading Grade Level' (3.84), 'Word Frequency' (2.315), 'Number of Words' (462), 'Type-token Ratio' (0.6), and 'Connectives' (73.593). The 'Help Links' column lists various links, such as 'Causal Cohesion', 'Concept Clarity', 'Connectives', 'Coreference Cohesion', 'Densities', 'Logical Operators', 'LSA', 'Part of Speech', 'Polysemy Hypernym', 'Readability', 'Syntactic Complexity', 'Type Token Ratio', 'Word Frequency', and 'Word Information'. At the bottom, there are 'Clear' and 'Submit' buttons.

Primary Measures		Help Links
Coreference Cohesion Global 1	0.589	Causal Cohesion
Coreference Cohesion Global 2	0.647	Concept Clarity
Coreference Cohesion Local 1	0.75	Connectives
Coreference Cohesion Local 2	0.818	Coreference Cohesion
Causal Cohesion	0.75	Densities
LSA Global	0.576	Logical Operators
LSA Local	0.582	LSA
Reading Ease	85.812	Part of Speech
Reading Grade Level	3.84	Polysemy Hypernym
Word Frequency	2.315	Readability
Number of Words	462	Syntactic Complexity
Type-token Ratio	0.6	Type Token Ratio
Connectives	73.593	Word Frequency
		Word Information

Figura 4.5: Parte das métricas apresentadas pela ferramenta *Coh-Metrix* (Graesser et al., 2004)

¹³<http://cohmetrix.memphis.edu/> - Acesso em 12/05/2010

É importante destacar que a ferramenta completa possui mais de 500 índices, porém apenas 62 são disponibilizados publicamente, sendo que boa parte são métricas simples de ocorrência de palavras. Dos índices públicos, destacamos cinco que acreditamos ser mais relevantes quanto a coerência no contexto deste trabalho: presença de referência anafórica em sentenças adjacentes, distância entre referências anafóricas, similaridade semântica (utilizando LSA) entre sentenças adjacentes, similaridade semântica entre todas as combinações de sentenças e entre parágrafos.

Embora seja preciso definir qual o gênero do texto a ser avaliado, diferentemente de outras ferramentas como *CriterionTM* ou *IntelliMetricTM*, a escrita do texto submetido para avaliação da *Coh-Metrix* não se prende a um tema ou tópico em específico, porém, os índices retornados pela ferramenta devem ser interpretados pelo usuário, o que dificulta o processo de reescrita e melhoria do texto, caso seja a intenção do usuário, já que a ferramenta não aponta em quais partes do texto podem existir problemas e não apresenta sugestões para a sua alteração.

Para o português do Brasil está em desenvolvimento uma adaptação das métricas calculadas na *Coh-Metrix*. O projeto é intitulado *Coh-Metrix-Port* (Almeida e Aluísio, 2009) e atualmente apresenta 34 das 62 métricas disponíveis na versão livre da ferramenta.

Ambiente SciPo

Neste capítulo apresentamos o ambiente de auxílio à escrita SciPo - *Scientific Portuguese* (Feltrim, 2004), no qual desenvolveu-se um módulo complementar que realiza a análise automática de coerência, objetivo principal deste trabalho. Além da apresentação do ambiente, também serão abordados na Seção 5.1 os modelos teóricos utilizados pelo SciPo e na Seção 5.2 abordamos detalhes de sua implementação, como o uso do classificador de estruturas retóricas¹ AZPort.

O SciPo foi inspirado no projeto AMADEUS - *Amiable Article Development for User Support* (Aluísio et al., 2001), um ambiente voltado à escritores não-nativos da língua inglesa e que vem sendo desenvolvido desde os anos 90. Diferentemente do AMADEUS, o SciPo enfoca a escrita em português das seções Resumo e Introdução de teses e dissertações da área de Ciência da Computação. Vale lembrar que o módulo desenvolvido neste trabalho contempla somente a seção Resumo, não abordando aspectos relativos à coerência em Introduções de tais textos.

O ambiente tem por objetivo apoiar a estruturação e a realização linguística de forma flexível de textos acadêmicos, deixando o usuário livre para escolher entre dois modos de trabalho, a saber:

- i) um processo *top-down*, que parte do planejamento estrutural para a escrita propriamente dita, herdado do projeto AMADEUS; ou

¹Neste trabalho, entende-se por estrutura retórica ou esquemática, o conjunto de elementos que descrevem a organização (funcional) do texto, relacionando suas partes de forma a atribuir-lhe sentido.

- ii) um processo *bottom-up*, em que se submete um texto já escrito à análise (detecção e crítica) automática da estrutura.

Na verdade, trata-se de pontos de partida distintos para um mesmo processo cíclico de refinamento, já que a estrutura detectada e criticada em (ii) pode ser aprimorada por meio dos recursos disponíveis em (i).

O apoio fornecido durante o processo *top-down* se dá por meio de um conjunto específico de funções e recursos com base em modelos pré-estabelecidos e no uso de exemplos autênticos. As seguintes funções de suporte ao reuso de córpus e de crítica podem ser acessadas (Feltrim, 2004):

- a) navegação nas bases de exemplos;
- b) pesquisa da ocorrência de determinado componente ou estratégia retórica;
- c) apoio à composição de estruturas esquemáticas;
- d) críticas da estrutura construída; e
- e) recuperação dos exemplos com estruturas similares à estrutura construída.

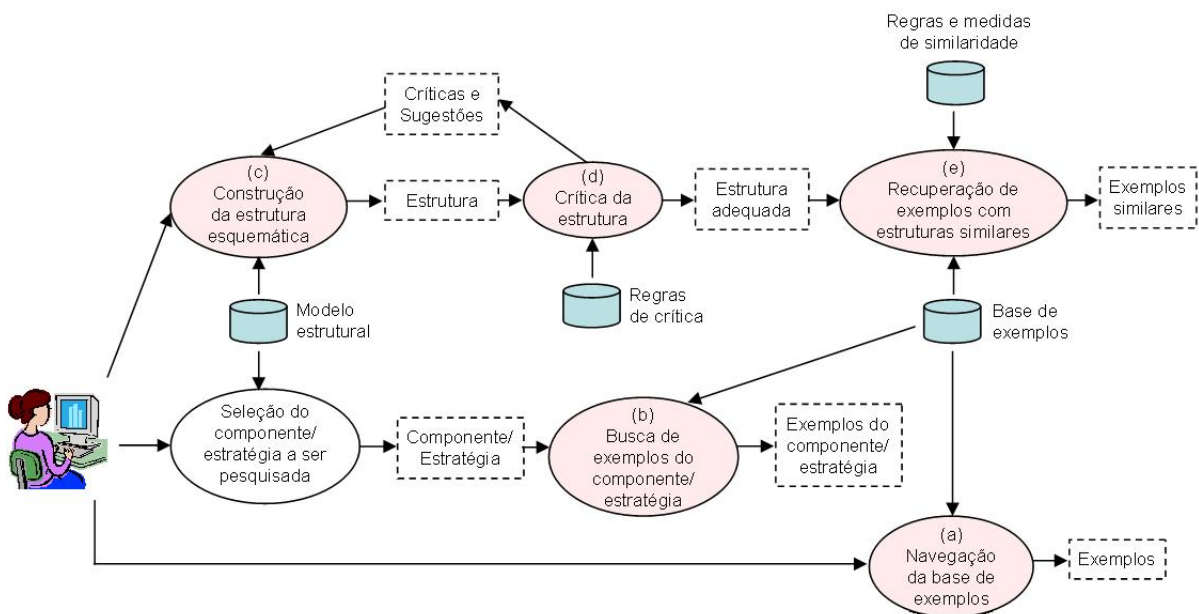


Figura 5.1: Visão geral do processo de apoio à estruturação do texto (Feltrim, 2004)

A Figura 5.1 mostra como o usuário pode interagir com essas funções para a criação da estrutura esquemática desejada, bem como as fontes de conhecimento empregadas em

cada processo. Após escolher a seção (Resumo ou Introdução) em que se deseja estruturar e escrever o texto, o usuário pode: (a) navegar na Base de exemplos, visualizando todos os exemplos anotados de acordo com suas respectivas estruturas esquemáticas, podendo também reutilizar as estruturas, se desejado; (b) selecionar um determinado componente ou estratégia retórica a partir de um modelo estrutural, para que sejam recuperadas todas as ocorrências desse componente/estratégia na Base de exemplos; ou (c) compor uma estrutura esquemática, com base no Modelo estrutural, que será submetida ao (d) processo de crítica da estrutura e, posteriormente, alimentará a (e) recuperação de exemplos similares. No processo de crítica, a estrutura construída pelo usuário é analisada com base nas Regras de crítica e, então, são retornadas ao usuário as críticas e sugestões cabíveis. O usuário poderá então modificar sua estrutura de acordo com as sugestões do sistema e submetê-la novamente à crítica. Esse ciclo se repete até que seja alcançada uma estrutura satisfatória. A recuperação de exemplos similares usa como entrada a estrutura construída pelo usuário, as estruturas esquemáticas da Base de exemplos e as Regras e medidas de similaridade, para retornar ao usuário os exemplos mais próximos à sua estrutura construída.

Na Figura 5.2 é apresentado um exemplo² de crítica da estrutura construída pelo usuário, realizada pelo processo *top-down* do ambiente SciPo.

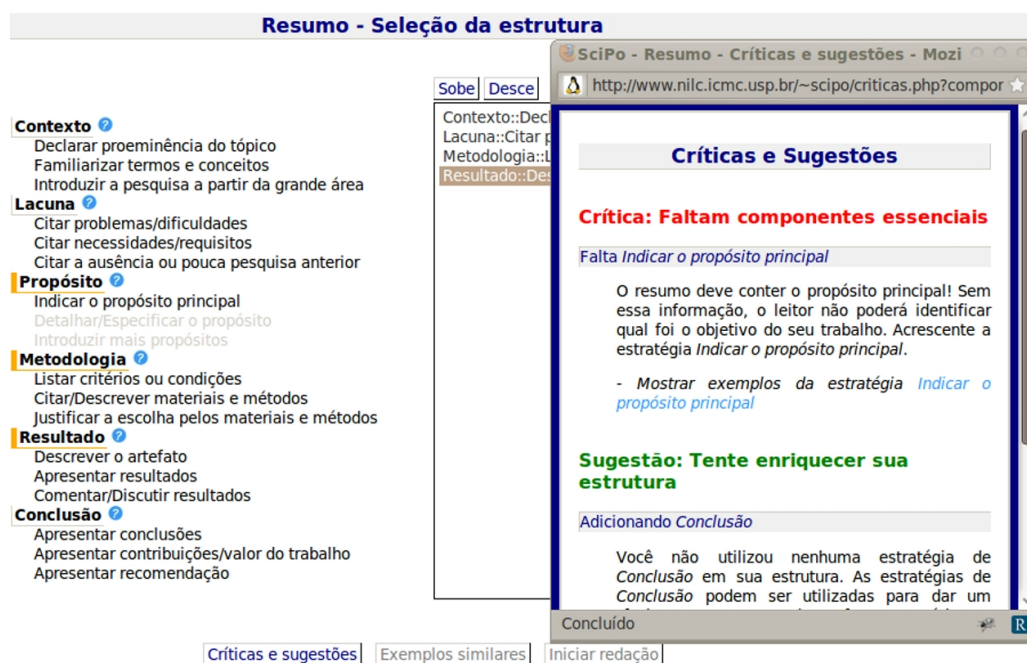


Figura 5.2: Críticas e sugestões fornecidas pelo SciPo em relação a estrutura construída pelo usuário

²www.nilc.icmc.usp.br/~scipo/ - Acesso em 10/06/2010

Conforme pode ser observado na Figura 5.2, a ferramenta detecta e relata em forma de crítica um desvio grave devido a falta do componente Propósito e também sugere a inserção do componente Conclusão para enriquecer a estrutura do texto. Também é possível observar que a ferramenta fornece a consulta de exemplos similares que podem auxiliar na escrita do texto após a sua estruturação.

Como dito anteriormente, o apoio à estruturação a partir do texto pronto (processo *bottom-up*) se dá de forma semelhante ao processo *top-down*, sendo que a diferença está no ponto de partida da interação do usuário com o ambiente. No processo *bottom-up* o sistema é capaz de detectar automaticamente a estrutura esquemática do texto redigido e submetido ao sistema, para então apresentar críticas e sugestões em relação a estrutura esquemática utilizada no texto a partir das Regras de crítica, como pode ser observado na Figura 5.3.

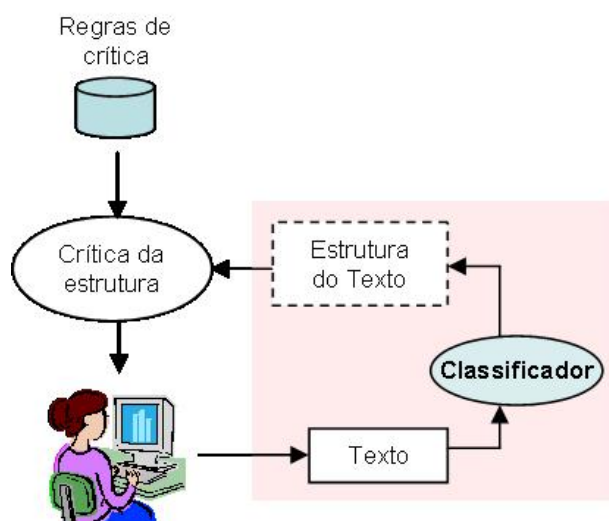


Figura 5.3: Processo *bottom-up* realizado pelo SciPo (Feltrim, 2004)

Na Figura 5.4 é apresentado um exemplo³ da detecção automática do papel retórico de cada sentença do resumo submetido para análise, inerente ao processo *bottom-up*. No exemplo, as sentenças foram classificadas respectivamente com os componentes Contexto e Metodologia. A partir desse ponto o escritor tem as opções de voltar ao seu texto, corrigir a estrutura detectada pela ferramenta ou receber as críticas e sugestões referentes a estrutura utilizada.

³www.nilc.icmc.usp.br/~scipo/ - Acesso em 10/06/2010

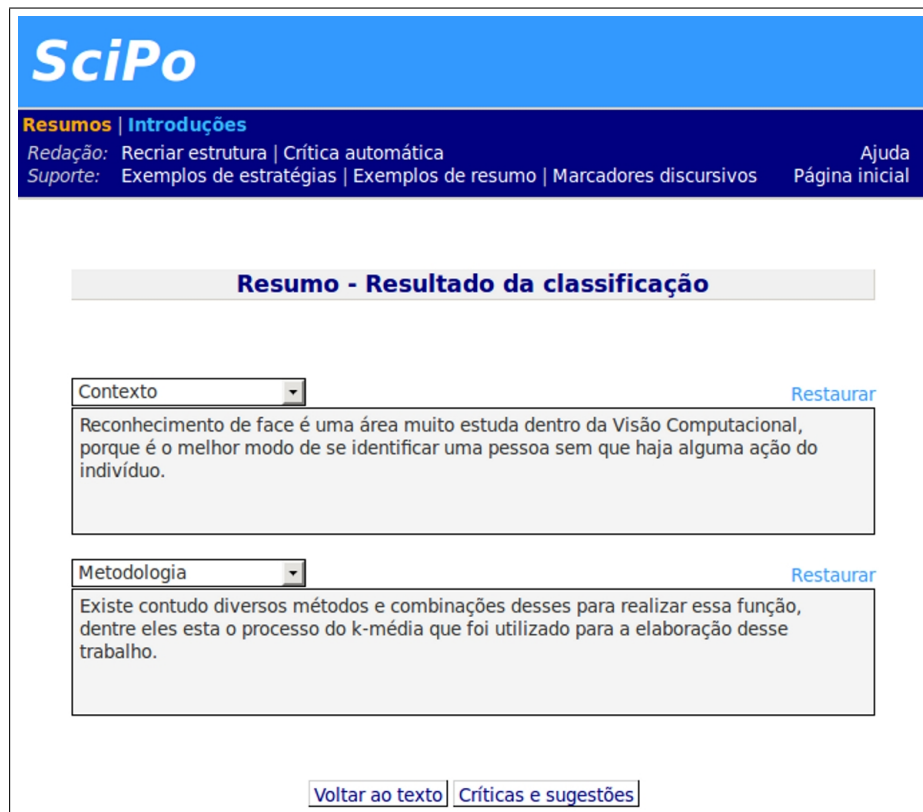


Figura 5.4: Resumo com sua estrutura detectada automaticamente pelo SciPo

5.1 Estruturação de Resumos Acadêmicos

Pode-se observar que o enfoque principal do ambiente SciPo é em relação a estruturação do texto. Nesta seção é tratado em especial, o modelo estrutural da seção Resumo, já que esta é a seção abordada neste trabalho.

Segundo Weissberg e Buker (1990), os resumos de pesquisa experimental de quase todas as áreas de estudo são escritos de maneira muito similar. Os tipos de informação e a ordem em que aparecem são muito convencionais, de modo que podem ser enunciados como modelos de resumo que objetivam guiar o escritor quanto aos tipos de informações necessárias, opcionais e a ordem convencional de apresentação.

Um dos modelos de resumo mais detalhado é o de Weissberg e Buker (1990), que nomeia e descreve os tipos de informação, também denominados componentes esquemáticos, esperados em um típico resumo acadêmico. Esse modelo é apresentado na Tabela 5.1.

Tabela 5.1: Modelo de um resumo típico, proposto por Weissberg e Buker (1990)

Componente	Descrição
Contexto	Alguma informação sobre o contexto da área de pesquisa
Propósito	A principal atividade (propósito) do estudo e seu escopo
Metodologia	Algumas informações sobre a metodologia usada no estudo
Resultados	Resultados mais importantes do estudo
Conclusão	Conclusão ou alguma recomendação

Baseado nos modelos de Aluísio e Oliveira Jr (1996); Swales (1990) e Weissberg e Buker (1990), Feltrim et al. (2003) propôs um modelo composto por seis componentes esquemáticos, representados em negrito na Tabela 5.2, em que cada componente pode ser refinado em três estratégias retóricas que os descrevem. Esse é o modelo utilizado no ambiente de auxílio à escrita SciPo. A ordem a ser seguida é a mesma disposta na Tabela 5.2, sendo que os componentes marcados com “*” são considerados obrigatórios.

Tabela 5.2: Modelo de resumo proposto por Feltrim et al. (2003)

1. Contexto (<i>Background</i>)
C1. Declarar proeminência do tópico
C2. Familiarizar termos e conceitos
C3. Introduzir a pesquisa a partir da grande área
2. Lacuna (<i>Gap</i>)
L1. Citar problemas/dificuldades
L2. Citar necessidades/requisitos
L3. Citar ausência ou falta de pesquisa anterior
3. Propósito (<i>Purpose</i>)*
P1. Indicar o propósito principal
P2. Detalhar/Especificar o propósito
P3. Introduzir mais propósitos
4. Metodologia (<i>Methodology</i>)*
M1. Listar critérios ou condições
M2. Citar/Descrever materiais e métodos
M3. Justificar a escolha pelos materiais e métodos
5. Resultado (<i>Result</i>)*
R1. Descrever os resultados
R2. Indicar os resultados
R3. Comentar/Discutir os resultados
6. Conclusão (<i>Conclusion</i>)
Co1. Apresentar conclusões
Co2. Apresentar contribuições/valor do trabalho
Co3. Apresentar recomendações

Também é apresentado na Tabela 5.2 o nome em inglês de cada componente (entre parênteses e itálico), já que em algumas ocasiões no decorrer deste trabalho, nos referimos aos componentes dessa maneira.

A proposta do modelo de Feltrim et al. (2003) foi amparada pela análise de um corpus composto por 52 resumos acadêmicos de dissertações/teses em várias áreas da Ciência da Computação⁴, em que 50% dos resumos tem entre 5 e 4 componentes, e 44,3% tem entre 3 e 2 componentes, sendo que todos apresentavam o componente Propósito, e os componentes Resultado e Metodologia constam respectivamente em 67,3% e 63,4% dos resumos analisados.

Tanto as regras de crítica como as sugestões do ambiente SciPo foram formuladas com base no modelo estrutural proposto e tentam corrigir padrões de estruturação considerados problemáticos e que podem prejudicar de alguma forma o entendimento do texto. As regras consideram desvios de conteúdo (ordem de ocorrência dos componentes/estratégias tidas como essenciais), além de diferenciar desvios graves (apresentados ao usuário como críticas) de desvios leves (apresentados ao usuário como sugestões) (Feltrim, 2004).

Embora o modelo utilizado reflita os elementos estruturais esperados, não é necessário que todos os seus componentes estejam presentes, nem que a ordem desses componentes seja rigorosamente obedecida, para que uma estrutura seja considerada boa. Dessa maneira, as regras de crítica consideram alguns componentes como obrigatórios enquanto outros são tratados como opcionais, como já observado na Tabela 5.2.

5.2 Argumentative Zoning for Portuguese - AZPort

Para permitir a análise estrutural automática de textos prontos, necessária ao processo *bottom-up*, o SciPo utiliza uma ferramenta de identificação da estrutura esquemática ou retórica de textos científicos chamada AZPort - *Argumentative Zoning for Portuguese* (Feltrim, 2004; Feltrim et al., 2006). O AZPort é um classificador estatístico construído nos moldes do AZ - *Argumentative Zoning* (Teufel e Moens, 2002), adaptado e treinado com um corpus de 52 resumos (CorpusDT) para detectar a estrutura retórica de resumos científicos em português.

O AZPort atribui uma categoria retórica para cada sentença do resumo avaliado dentre um conjunto de seis categorias, apresentadas e brevemente descritas na Tabela 5.3. Vale destacar que cada categoria retórica corresponde a um componente do modelo estrutural de resumos apresentado na Tabela 5.2.

⁴CorpusDT (Feltrim et al., 2003)

Tabela 5.3: Descrição e exemplo dos papéis retóricos identificados pelo AZPort

Etiqueta	Categoria	Descrição	Exemplo de expressão
B	Background	Conhecimento aceito pela comunidade científica	A partir do ano ...
G	Gap	Problema de pesquisa, necessidade, ...	Contudo, é necessário ...
P	Purpose	Propósito da pesquisa	Esta tese apresenta ...
M	Methodology	Metodologia utilizada	Foi utilizado o modelo ...
R	Result	Resultados obtidos	Os resultados mostram...
C	Conclusion	Conclusão, recomendação, contribuição, ...	Concluimos que...

Assim como o AZ, o AZPort atribui uma categoria retórica para cada sentença do resumo apresentado com base em um conjunto de atributos que são extraídos automaticamente de cada sentença do texto. Dessa forma, assim como outros algoritmos de aprendizagem de máquina, em vez de lidar diretamente com o objeto a ser classificado, o AZPort recebe as sentenças como vetores de atributos. O AZPort utiliza um conjunto de oito atributos, resumidos na Tabela 5.4.

Tabela 5.4: Conjunto de atributos utilizados pelo AZPort

Atributo	Descrição	Valores possíveis
1. Tamanho	Qual é o tamanho da sentença em comparação aos dois limiares (20 e 40 palavras)	Curta, média ou longa
2. Localização	Qual a posição da sentença?	Primeira, segunda, mediana, penúltima ou última
3. Citação	A sentença contém citações?	Sim ou não
4. Expressão	Que tipo de expressão padrão a sentença contém?	B, G, P, M, R, C ou noexpr
5. Tempo	Qual o tempo do primeiro verbo finito da sentença?	IMP, PRES, PAST, FUT, PRES-CPO, PAST-CPO, FUT-CPO, PRES-CT, PAST-CT, FUT-CT, PRES-CPO-CT ou <i>noverb</i>
6. Voz	Qual a voz do primeiro verbo finito da sentença?	Passiva, ativa ou <i>noverb</i>
7. Modal	O primeiro verbo finito da sentença é modal?	Sim, não ou <i>noverb</i>
8. Histórico	Qual é a categoria da sentença anterior?	-, B, G, P, M, R, C

Os atributos apresentados na Tabela 5.4 são determinados automaticamente a partir do texto de entrada e são utilizados para classificar cada sentença do texto em uma das categorias retóricas vistas na Tabela 5.3. A concordância do classificador com um anotador humano medida por meio do *Kappa* é $K=0.65$ (Feltrim, 2004).

Aprendizagem de Máquina

Neste capítulo são apresentados conceitos de Aprendizagem de Máquina (do inglês, *Machine Learning*), uma das sub-áreas da Inteligência Artificial cujo objetivo é o desenvolvimento de técnicas computacionais a serem utilizadas no processo de aprendizagem e na construção de sistemas capazes de adquirir novos conhecimentos de forma automática. Um sistema de aprendizagem é um programa de computador que toma decisões baseado em experiências acumuladas por meio da solução bem-sucedida de problemas anteriores (Mitchell, 1997). Essa premissa vem ao encontro da abordagem utilizada neste trabalho para se alcançar o objetivo principal do mesmo, a análise automática de coerência. O ponto inicial de partida se deu na análise manual de um corpus significativo, abordada no Capítulo 7, como uma forma de acumular soluções para o problema e então construir um sistema de aprendizagem capaz de tomar decisões baseadas no conhecimento adquirido nessa análise. Por isso, se vê necessário o entendimento dos conceitos de aprendizagem de máquina abordados neste capítulo.

Além dos conceitos iniciais abordados na Seção 6.1, também são apresentadas métricas utilizadas para avaliação dos algoritmos de aprendizagem de máquina na Subseção 6.1.1. Por fim, apresentamos o ambiente Weka (Hall et al., 2009; Witten e Frank, 2005) na Seção 6.2, que contém diversas implementações de algoritmos de aprendizagem de máquina e que foram utilizados na indução e avaliação dos classificadores propostos neste trabalho.

6.1 Conceitos de Aprendizagem de Máquina

O aprendizado de um novo conceito a partir de conhecimentos prévios pode ser realizado de várias maneiras. Uma delas é que será abordada neste trabalho é a indução. Indução é a forma de inferência lógica que permite que conclusões gerais sejam obtidas a partir de exemplos particulares. Na aprendizagem por indução, um conceito é adquirido por meio de inferências indutivas sobre os fatos apresentados. Desse modo, os modelos gerados por inferência indutiva podem ou não preservar a verdade (Rezende, 2003). Quando as premissas de um argumento indutivo são verdadeiras, pode-se dizer que sua conclusão será provavelmente verdadeira, mas não necessariamente verdadeira.

Os sistemas computacionais que realizam a indução são também denominados de sistemas de aprendizagem, algoritmos de aprendizagem ou simplesmente indutores e os modelos gerados por esses sistemas podem ser chamados de classificadores, quando se destinam à tarefas de classificação.

A aprendizagem indutiva pode ser dividida em supervisionada e não supervisionada, descritas a seguir (Rezende, 2003):

Aprendizagem Supervisionada: É fornecido ao sistema de aprendizagem um conjunto de exemplos $E = \{E_1, E_2, \dots, E_N\}$, também denominado de conjunto de treinamento, sendo que cada exemplo (ou instância) $E_i \in E$ possui um rótulo associado que define a classe ou categoria a qual o exemplo pertence. Formalmente, podemos dizer que cada exemplo $E_i \in E$ é uma tupla $E_i = (\vec{x}_i, y_i)$, em que $\vec{x}_i$ é um vetor de valores que representa as características do exemplo E_i , também chamado de vetor de atributos e y_i é o valor da classe. Tal representação é ilustrada na Tabela 6.1.

Tabela 6.1: Conjunto de exemplos no formato atributo-valor (Rezende, 2003)

	A_1	A_2	$\dots$	A_M	Y
E_1	x_{11}	x_{12}	$\dots$	x_{1M}	y_1
E_2	x_{21}	x_{22}	$\dots$	x_{2M}	y_2
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
E_N	x_{N1}	x_{N2}	$\dots$	x_{NM}	y_N

Na Tabela 6.1, as linhas representam as instâncias, as colunas os atributos e a última coluna representa o meta-atributo (classe) da instância, que pode assumir valores qualitativos (nominais) ou quantitativos (contínuos). Quando se trata de rótulos qualitativos,

a tarefa do sistema de aprendizagem é chamada de classificação e quando se trata de rótulos quantitativos, a tarefa é chamada de regressão. Assim como os valores das classes, os atributos podem assumir dois tipos de valores: nominal ou contínuo.

Em linhas gerais, o objetivo da aprendizagem supervisionada é induzir um mapeamento dos vetores $\vec{x}$ para os valores de classe y . Portanto, o sistema de aprendizagem deve construir um modelo $y = f(\vec{x})$, de uma função desconhecida f , que permite prever valores y para exemplos previamente não vistos. Entretanto, o número de exemplos fornecidos para a determinação do modelo não é, na maioria dos casos, suficiente para caracterizar completamente a função desconhecida f . Por isso, na verdade, o sistema de aprendizagem induz uma função h que aproxima f , ou seja, $h(\vec{x}) \approx f(\vec{x})$ e, portanto, h é chamada de hipótese ou modelo sobre a função conceito f .

Aprendizagem Não-Supervisionada: É fornecido ao sistema de aprendizagem um conjunto de exemplos E , no qual cada exemplo consiste de vetores $\vec{x}$, não incluindo a informação sobre a classe y . Sendo assim, o objetivo do sistema de aprendizagem não-supervisionada é construir um modelo que procura por regularidade nos exemplos, formando agrupamentos ou *clusters* de exemplos com características similares.

Dessa maneira, a aprendizagem indutiva pode ser representada hierarquicamente como ilustrada na Figura 6.1, dividindo-se em aprendizagem supervisionada e não-supervisionada, sendo que o foco deste trabalho é a aprendizagem indutiva supervisionada para problemas de classificação. Sendo assim, apenas a tarefa de classificação será considerada nas próximas seções.

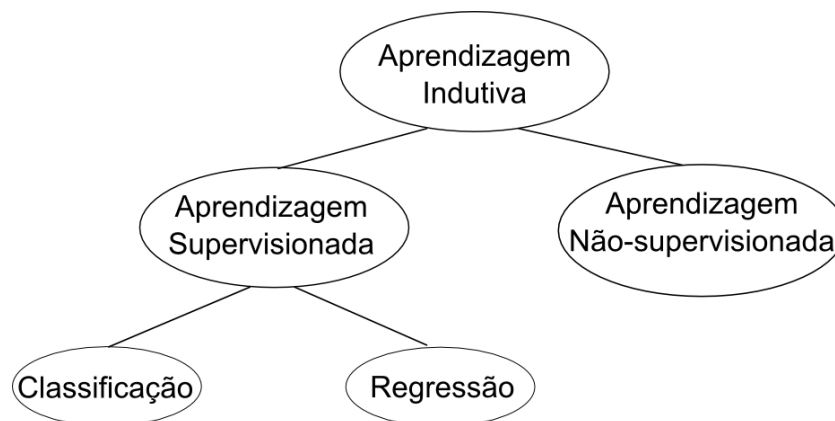


Figura 6.1: Hierarquia da aprendizagem indutiva (Rezende, 2003)

6.1.1 Métricas de Avaliação

O objetivo principal dos sistemas de aprendizagem é classificar com sucesso casos nunca vistos. Quando estão disponíveis apenas pequenos conjuntos de exemplos de treinamento ou exemplos pouco significativos, esse objetivo pode não ser satisfatoriamente alcançado. Desse modo, nesta seção são apresentadas métricas comumente utilizadas para a avaliação de classificadores e também empregadas na avaliação dos classificadores desenvolvidos neste trabalho.

Com base nos trabalhos de Rezende (2003) e Witten e Frank (2005), as principais métricas utilizadas na avaliação de classificadores são apresentadas a seguir.

Taxa de erro: Para problemas de classificação, uma medida simples e comumente utilizada para avaliar o desempenho do classificador é a taxa de erro (*error rate*). A taxa de erro é obtida comparando-se a classe verdadeira de cada exemplo com a classe atribuída pelo classificador induzido, denotada pela equação:

$$err(h) = \frac{\text{numero de erros}}{\text{numero de casos}}$$

O complemento da taxa de erro, chamado de precisão (*accuracy*) e denotado por $acc(h)$, consiste do número de acertos do classificador dividido pelo número total de exemplos classificados. Por exemplo, dado um conjunto de 10 exemplos submetidos a um classificador h_1 , em que 7 exemplos foram classificados corretamente e 3 classificados incorretamente, temos $err(h_1) = 0,3$ e $acc(h_1) = 0,7$, ou 30% de erro e 70% de acerto nas classificações de h_1 .

Matriz de confusão: Se todos os erros são de igual importância, apenas a taxa de erro seria suficiente para avaliar o desempenho do classificador, mas não é o que normalmente ocorre. Frequentemente, é necessário obter detalhes e fazer distinções entre os tipos de erros e acertos de classificação cometidos por um classificador. A matriz de confusão fornece informações sobre cada caso do conjunto de teste, mostrando a classe real da instância e a classe rotulada pelo classificador. Na Tabela 6.2, é apresentada uma matriz de confusão genérica para k classes diferentes, $C_1, C_2, \dots, C_k$, em que cada elemento m_{ij} da matriz indica o número de casos da classe C_i que foram rotulados com a classe C_j .

Tabela 6.2: Matriz de confusão genérica (Rezende, 2003)

Classe	Predita C_1	Predita C_2	...	Predita C_k	Total
Realmente C_1	m_{11}	m_{12}	...	m_{1k}	T_{R_1}
Realmente C_2	m_{21}	m_{22}	...	m_{2k}	T_{R_2}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
Realmente C_k	m_{k1}	m_{k2}	...	m_{kk}	T_{R_k}
Total	T_{C_1}	T_{C_2}	...	T_{C_k}	N

O número de acertos do classificador para cada classe se localiza na diagonal principal da matriz e os demais elementos representam os erros de classificação. Dessa maneira, a matriz de confusão de um classificador ideal que não comete erros possui o número zero em todos os elementos fora da diagonal principal. Os totais T_{R_i} e T_{C_i} representam, respectivamente, o número total de casos que realmente são da classe C_i e o número total de casos preditos pelo classificador como sendo da classe C_i .

Existem quatro classificações possíveis, como visto na Tabela 6.3 que representa a matriz de confusão das classes C_+ e C_- , em que:

- T_P = Verdadeiro Positivo (*True Positive*)
- F_N = Falso Negativo (*False Negative*)
- F_P = Falso Positivo (*False Positive*)
- T_N = Verdadeiro Negativo (*True Negative*)

Tabela 6.3: Tipos de erros de classificação em uma matriz com duas classes

Classe	Predita C_+	Predita C_-
Realmente C_+	T_P	F_N
Realmente C_-	F_P	T_N

Na Tabela 6.3, T_P representa a quantidade de instâncias classificadas como C_+ e que realmente pertencem à classe, F_N representa a quantidade de instâncias classificadas como C_- mas que pertencem à classe C_+ , F_P representa a quantidade de instâncias classificadas como C_+ mas que pertencem à classe C_- e, por fim, T_N representa a quantidade de instâncias classificadas como C_- e que realmente pertencem à classe. A partir da análise

da matriz de confusão é possível observar mais precisamente para quais pares de classe o classificador comete mais erros, além de permitir o cálculo de outras métricas.

Precision: É a razão entre as respostas corretas para uma determinada classe pelo total de respostas obtidas pelo modelo. No exemplo da matriz de confusão apresentada na Tabela 6.3, podemos expressar o valor de *Precision* da classe C_+ como:

$$Precision(C_+) = \frac{T_P}{T_P + F_P}$$

De modo geral, para uma matriz de confusão genérica, como a apresentada na Tabela 6.2, o valor de *Precision* da classe C_i é dado por:

$$Precision(C_i) = \frac{m_{ii}}{T_{C_i}}$$

Recall: É a razão das respostas corretas para uma determinada classe pelo total de respostas corretas possíveis. No exemplo da matriz de confusão apresentada na Tabela 6.3, podemos expressar o valor de *Recall* da classe C_+ como:

$$Recall(C_+) = \frac{T_P}{T_P + F_N}$$

De modo geral, para uma matriz de confusão genérica, como a apresentada na Tabela 6.2, o valor de *Recall* da classe C_i é dado por:

$$Recall(C_i) = \frac{m_{ii}}{T_{R_i}}$$

F-measure: É uma média calculada a partir dos valores das medidas *Precision* e *Recall*, sendo uma forma conveniente de expressá-las em um único valor. De modo geral, para uma matriz de confusão genérica, como a apresentada na Tabela 6.2, o valor de *F-measure* da classe C_i é dado por:

$$F-measure(C_i) = \frac{2 * Precision(C_i) * Recall(C_i)}{Precision(C_i) + Recall(C_i)}$$

Macro-F: É a média das *F-measures* de todas as categorias, sendo uma forma de conveniente de expressar a combinação das medidas *Precision* e *Recall* de todas as classes possíveis de serem classificadas. É útil para avaliar se o classificador não sacrifica o desempenho de uma ou outra classe com poucos exemplos em troca de uma melhora na

taxa de acerto. Dado um número k de classes possíveis, a medida *Macro-F* pode ser expressa pela equação:

$$Macro-F = \frac{1}{k} \sum_{i=1}^k F-measure(C_i)$$

Kappa: Mede a concordância entre k classificadores sobre N exemplos. É uma medida estatística e sua fórmula para o cálculo é:

$$Kappa = \frac{P_A - P_E}{1 - P_E}$$

em que P_A é a taxa de concordância observada para cada par de juízes (ou classificadores) e P_E estima a concordância ao acaso entre eles. A matriz de confusão pode ser utilizada para o cálculo das probabilidades P_A e P_E . Dessa modo, a partir da matriz apresentada na Tabela 6.2, temos que:

$$P_A = \frac{1}{N} \sum_{i=1}^k m_{ii}$$

$$P_E = \frac{1}{N^2} \sum_{i=1}^k T_{C_i} * T_{R_i}$$

O valor *Kappa* varia entre -1 e 1 , em que -1 indica máxima discordância, 0 indica que houve apenas concordância ao acaso e 1 indica concordância perfeita entre os juízes. A Tabela 6.4 apresenta uma possível interpretação dos níveis de concordância dados pela medida *Kappa*. Destacamos que tal interpretação pode ser diferente de acordo com o nível de complexidade da tarefa em que se está medindo a concordância.

Tabela 6.4: Níveis de concordância dos valores *Kappa* (Landis e Koch, 1977)

Kappa	Nível de concordância
< 0	Péssimo
$0,01 - 0,20$	Ruim
$0,21 - 0,40$	Razoável
$0,41 - 0,60$	Bom
$0,61 - 0,80$	Muito bom
$0,81 - 1$	Excelente

A medida *Kappa* pode ser utilizada tanto para mensurar a concordância entre juízes humanos quanto na avaliação de uma tarefa realizada automaticamente, como um classi-

ficador induzido. Dessa maneira, a classificação realizada por um modelo induzido pode ser comparada tanto às classificações realizadas por humanos quanto às classificações realizadas por outros indutores.

As medidas apresentadas nesta seção estimam o desempenho de classificadores induzidos a partir de um conjunto finito de exemplos. Para que as medidas obtenham resultados válidos e realistas, é necessário que o conjunto de exemplos utilizado para o treinamento e para a avaliação sejam diferentes e aleatoriamente distribuídos.

As duas principais técnicas de amostragem utilizadas para avaliação de algoritmos de aprendizagem de máquina supervisionada são a *Holdout* e a *Cross-Validation*.

Holdout: Consiste em dividir aleatoriamente os n exemplos em uma porcentagem fixa de p exemplos para treinamento e $(1 - p)$ para teste, normalmente considerando $p > \frac{1}{2}$. Valores típicos são $p = \frac{2}{3}$ e $(1 - p) = \frac{1}{3}$, embora não existam fundamentos teóricos sobre esses valores.

k-fold cross validation: Consiste em dividir os exemplos aleatoriamente em k partições mutuamente exclusivas (*folds*) de tamanho aproximadamente igual a $\frac{n}{k}$, sendo n o número total de exemplos. Os exemplos nos $(k - 1)$ *folds* são usados para treinamento (indução do classificador) e a hipótese induzida é testada no *fold* remanescente. Esse processo é repetido k vezes, cada vez considerando um *fold* diferente para teste. As métricas de avaliação são feitas calculando-se a média dos resultados obtidos no teste de cada um dos k *folds*. Um valor típico para k é 10 (*10-fold cross-validation*), mas outros valores podem ser utilizados. A vantagem da *cross-validation* é que todos os exemplos disponíveis são utilizados para treinamento e teste, mas sem que o mesmo exemplo seja utilizado em ambos os processos.

k-fold stratified cross validation: Similar à *cross-validation*, mas ao gerar os k *folds* mutuamente exclusivos, a distribuição de classes (proporção de exemplos em cada uma das classes) é considerada durante a amostragem. Por exemplo, se o conjunto original de exemplos possui duas classes com distribuição de 20% e 80%, então cada *fold* também terá essa proporção de classes.

6.2 Waikato Environment for Knowledge Analysis

O Weka (Hall et al., 2009; Witten e Frank, 2005) é um ambiente que possui diversas implementações de algoritmos de aprendizagem de máquina de diferentes paradigmas, além de algoritmos para preparação de dados e validação dos resultados. O ambiente foi desenvolvido na Universidade de Waikato na Nova Zelândia, sendo todo escrito em Java e de código aberto disponível¹ para uso, consulta e adequações por parte de seus usuários.

O Weka possui um formato de arquivo de entrada próprio (ARFF), basicamente dividido em duas partes: a primeira é composta por um cabeçalho que contém uma lista de todos os atributos e seus tipos e a segunda parte consiste das instâncias que serão utilizadas. A Figura 6.2 ilustra um arquivo no formato ARFF com 5 atributos (*outlook*, *temperature*, *humidity*, *windy* e *play?*) e 14 instâncias que representam a relação *weather*.

```
% ARFF file for the weather data with some numeric features
%
@relation weather

@attribute outlook { sunny, overcast, rainy }
@attribute temperature numeric
@attribute humidity numeric
@attribute windy { true, false }
@attribute play? { yes, no }

@data
%
% 14 instances
%
sunny, 85, 85, false, no
sunny, 80, 90, true, no
overcast, 83, 86, false, yes
rainy, 70, 96, false, yes
rainy, 68, 80, false, yes
rainy, 65, 70, true, no
overcast, 64, 65, true, yes
sunny, 72, 95, false, no
sunny, 69, 70, false, yes
rainy, 75, 80, false, yes
sunny, 75, 70, true, yes
overcast, 72, 90, true, yes
overcast, 81, 75, false, yes
rainy, 71, 91, true, no
```

Figura 6.2: Arquivo no formato ARFF (Witten e Frank, 2005)

¹<http://www.cs.waikato.ac.nz/ml/weka/> - Acesso em 24/11/2010

O ambiente possui uma interface gráfica que permite a fácil exploração de suas funcionalidades pelo usuário. A Figura 6.3 apresenta a ferramenta *Explorer* que é utilizada para treinamento e avaliação de classificadores. É possível observar na Figura 6.3 a apresentação de diferentes métricas de avaliação, tais como *Kappa*, *Precision*, *Recall*, *F-Measure* e matriz de confusão, além de permitir a escolha de diferentes algoritmos de classificação (no exemplo, algoritmo SMO (Keerthi et al., 2001)), forma de realização dos testes (*Cross-validation*, *Holdout*), entre outras opções. Além do ambiente visual fornecido pela interface gráfica, o Weka permite a sua utilização por linha de comando, o que facilita a integração da ferramenta com outros projetos.

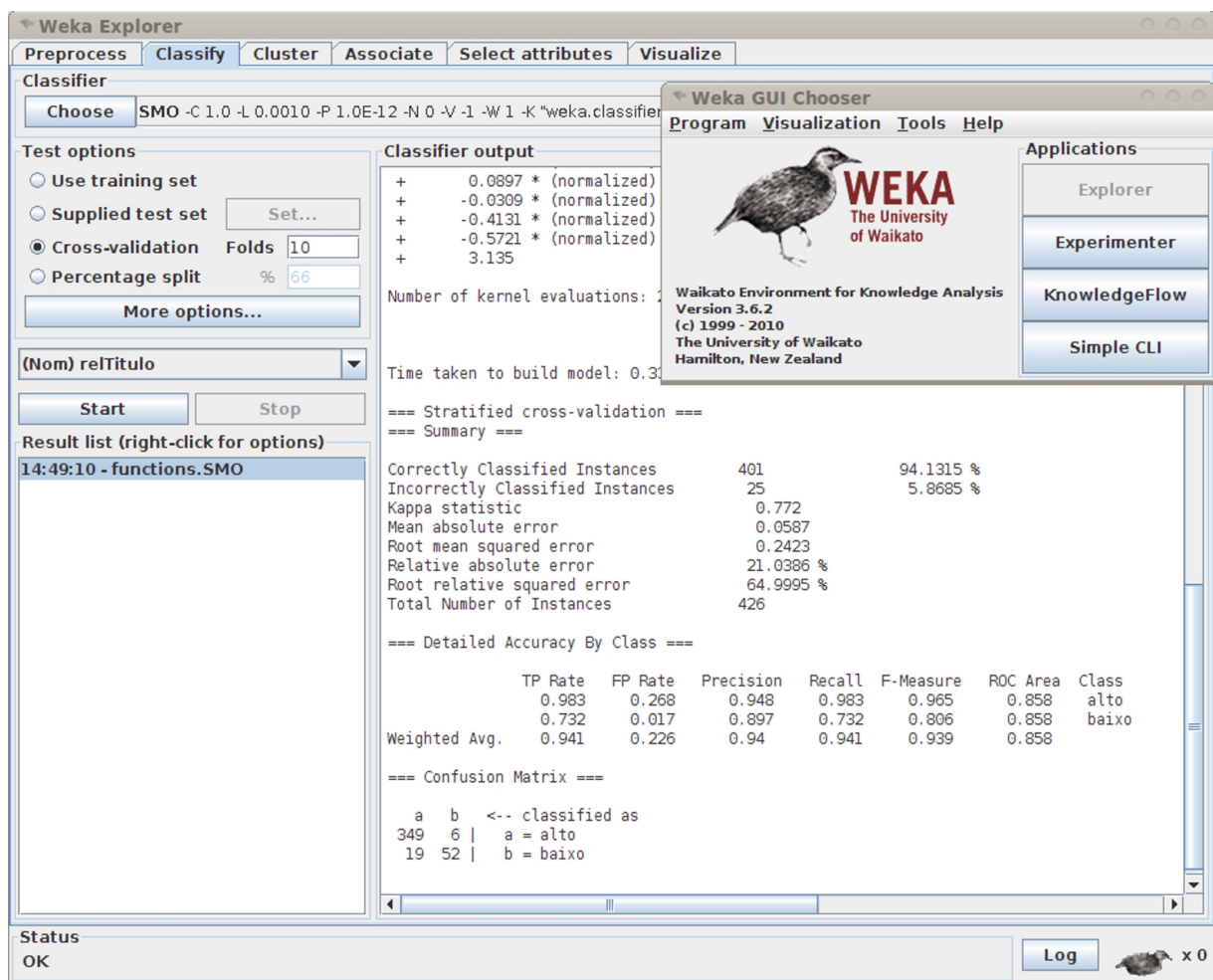


Figura 6.3: Funcionalidade *Classify* da ferramenta *Explorer* do Weka

Além da ferramenta *Explorer* utilizada para o pré-processamento, classificação, agrupamento, associação, seleção de atributos e visualização dos dados, o ambiente possui outras três ferramentas:

- ***Experimenter***: ferramenta para a realização de experimentos que comparam o desempenho de diferentes algoritmos de aprendizagem;
- ***Simple CLI***: ferramenta que permite a execução por linha de comando para sistemas operacionais que não disponibilizam sua própria interface em linha de comando; e
- ***KnowledgeFlow***: ferramenta que possui a mesma funcionalidade da ferramenta *Explorer*, mas com uma interface *drag-and-drop* (arrastar e soltar).

Coleta, Anotação e Análise de Córpus

Para a investigação dos problemas relacionados à coerência textual, fez-se necessária a formação de um córpus específico para esse propósito. Como este trabalho está inserido no contexto do ambiente SciPo, o córpus formado se restringiu a sua área de conhecimento (Ciência da Computação), mais especificamente, formado por resumos de monografias de conclusão de curso de tal área. Aspectos relacionados ao processo de coleta, anotação e análise são tratados nas Seções 7.1, 7.2 e 7.3, respectivamente. Uma segunda fase de anotação e análise é abordada na Seção 7.4.

7.1 Coleta

Todos os trabalhos foram coletados em formato digital e são provenientes de três fontes: Departamento de Informática da Universidade Estadual de Maringá (DIN-UEM), Departamento de Computação da Universidade Estadual de Londrina (DC-UEL) e Departamento de Informática do Instituto de Física e Matemática da Universidade Federal de Pelotas (Inf-UFPel). Os trabalhos provenientes do DIN-UEM foram coletados diretamente com seus autores ou respectivos orientadores, enquanto os outros trabalhos foram encontrados no Sistema de Arquivamento e Indexação de Documentos do DC-UEL¹ e no Acervo Digital da Biblioteca de Ciência e Tecnologia da UFPel².

¹<http://www2.dc.uel.br/nourau/>

²<http://www.ufpel.tche.br/prg/sisbi/bibct/acervodigital.html>

Mesmo com todos os arquivos em formato digital, foi necessário um esforço manual para a extração de seu conteúdo, já que se extraiu somente a seção resumo desses textos e alguns dos arquivos estavam em formatos que impossibilitam a extração de seu conteúdo (p.e., *PostScript*). Os trabalhos estão hoje armazenados em arquivos do tipo texto sem formatação.

Foram coletados trabalhos de várias áreas de pesquisa da Ciência da Computação, sendo 205 resumos pertencentes ao DIN-UEM, 98 ao DC-UEL e 82 ao Inf-UFPel, totalizando 385 resumos. Os trabalhos datam de 1999 a 2009. A Tabela 7.1 apresenta a distribuição dos trabalhos entre as áreas de pesquisa identificadas, bem como os totais de palavras distribuídos pelas áreas e a quantidade média de palavras de cada resumo. Também é mostrado o total de palavras de todo o corpus (64.104 palavras) e a média geral da quantidade de palavras.

Tabela 7.1: Distribuição e totais de palavras do corpus por área de conhecimento

Área de Pesquisa	No. Resumos	No. Palavras	Média
Banco de Dados	51	8.720	170,98
Inteligência Computacional	70	11.149	159,27
Engenharia de Software	92	15.482	168,28
Hipermídia	32	5.755	179,84
Sistemas Digitais	36	6.454	179,27
Sist. Distribuídos e Ling. Programação	19	4.192	220,63
Programação Gráfica e Proc. de Imagens	41	6.565	160,12
Redes de Computadores	44	5.787	131,52
Total	385	64.104	171,23

7.2 Anotação

Para auxiliar a manipulação e compreensão das informações contidas no corpus, tanto por humanos quanto por ferramentas computacionais, é necessário que o corpus esteja eletronicamente anotado com um conjunto de etiquetas apropriadas. Para essa anotação, optamos por utilizar o XML (*Extensible Markup Language*), um padrão para anotação de documentos, aprovado pelo *World Wide Consortium* - W3C³, que vem sendo mundialmente utilizado para a anotação de corpus.

O XML é um subconjunto do SGML (*Standard Generalized Markup Language*) e foi criado para a manipulação de documentos na Internet. Esse padrão incorpora

³<http://www.w3.org/>

características do SGML e amplia o poder de estruturação de documentos, facilitando o intercâmbio de dados entre aplicações e possibilitando que o formato de apresentação dos dados seja independente da anotação, uma vez que o XML não especifica como os seus elementos devem ser apresentados. Uma forma de apresentação de documentos XML consiste em transformá-los em documentos HTML (*Hypertext Markup Language*), possibilitando que sejam visualizados em um *browser* comum (Fallside, 2001).

Em um primeiro momento foi utilizado um conjunto de quatro etiquetas e quatro atributos para realizar a anotação do corpús. A Figura 7.1 apresenta um resumo com a anotação preliminar realizada automaticamente por *scripts* desenvolvidos neste trabalho.

```
<RESUMO id="3">
<Titulo>Caracterização de Tráfego de Rede</Titulo>
<P id="0">
<S id="3-0" AZ="B">A caracterização de tráfego de rede tem como objetivos
principais: garantia de qualidade de serviço (QoS), planejar e modelar o tráfego,
analisar o desempenho de redes, analisar a perda de pacotes e estudar o
comportamento do usuário.</S>
<S id="3-1" AZ="R">Neste trabalho, serão mostradas algumas técnicas de
caracterização de tráfego como: processos de renovação, processos de Markov,
séries temporais e auto-similaridade.</S>
<S id="3-2" AZ="B">Além destes processos, podem ser utilizados os sketches.</S>
<S id="3-3" AZ="B">Estes possuem a característica de fazer a análise dinâmica e
em tempo real de fluxo de dados, o que é muito importante para detectar
anomalias.</S>
</P>
</RESUMO>
```

Figura 7.1: Exemplo de resumo preliminarmente anotado

Na Figura 7.1 é possível observar as etiquetas escritas em azul e os atributos escritos em verde. A primeira etiqueta é denotada por “<RESUMO> ... </RESUMO>” e tem a função de estabelecer o início e fim do resumo, além do atributo “id” que é atribuído de acordo com o nome do arquivo em questão (neste caso, “3.xml”) sem a representação de seu formato. A etiqueta “<Titulo> ... </Titulo>” tem a função de estabelecer o início e o fim do texto referente ao título do resumo. A etiqueta “<P> ... </P>” tem a função de estabelecer o início e o fim do texto de cada parágrafo do resumo e o atributo “id” se refere a posição do parágrafo, com a contagem iniciando em zero. Por último, a etiqueta “<S> ... </S>” tem a função de estabelecer o início e fim do texto de cada sentença do resumo e possui os atributos “id” e “AZ”. O atributo “id” possui o valor do “id” da etiqueta “<RESUMO>” seguido da posição da sentença no resumo, com a contagem iniciando em zero. O atributo “AZ” se refere a classificação atribuída à sentença em um dos seis componentes que compõem a estrutura retórica de um resumo, como apresentado na Tabela 5.3 do Capítulo 5.

É importante observar que tal anotação é dita preliminar, pois novos atributos foram adicionados a essa anotação em um segundo momento, conforme descrito na Seção 8.4 do Capítulo 8.

7.3 Análise

Com base nos componentes que compõem a estrutura retórica de um resumo acadêmico proposta no modelo de Feltrim (2004), conforme apresentado na Seção 5.1, realizou-se a análise e anotação de cada uma das sentenças do corpus utilizando-se etiquetas XML apropriadas. Na Tabela 7.2 pode-se observar a frequência de cada componente no corpus, sendo que o componente *Purpose* está presente em quase todos os resumos analisados, 97,4% (375 resumos), seguido dos componentes *Background* (68,05%), *Result* (55,32%), *Gap* (40,51%), *Methodology* (37,66%) e *Conclusion* (23,11%).

Tabela 7.2: Frequência dos componentes observada no corpus

Componente	No. Resumos	Frequência
<i>Background</i>	262	68,05%
<i>Gap</i>	156	40,51%
<i>Purpose</i>	375	97,40%
<i>Methodology</i>	145	37,66%
<i>Result</i>	213	55,32%
<i>Conclusion</i>	89	23,11%

A partir do conjunto de 385 resumos, obteve-se um total de 2.293 sentenças anotadas, sendo que sua distribuição pode ser observada na Tabela 7.3.

Tabela 7.3: Distribuição dos componentes que compõem a estrutura retórica dos resumos analisados

Componente	No. Sentenças	Distribuição
<i>Background</i>	808	35,23%
<i>Gap</i>	215	9,38%
<i>Purpose</i>	426	18,58%
<i>Methodology</i>	273	11,90%
<i>Result</i>	451	19,67%
<i>Conclusion</i>	120	5,24%
Total	2.293	100%

Pode-se observar na Tabela 7.3 que as sentenças classificadas como *Background* estão presentes em maior número no corpus, constituindo um total de 808 sentenças, o que corresponde a 35,23% do total das sentenças. As sentenças classificadas como *Result*, *Purpose*, *Methodology*, *Gap* e *Conclusion* representam respectivamente 19,67%, 18,58%, 11,90%, 9,38% e 5,24% do total de sentenças analisadas. Nota-se que embora 97,4% dos resumos apresentem sentenças classificadas como *Purpose*, como observado na Tabela 7.2, esse tipo de sentença não é a de maior quantidade no corpus. Isso se deve ao fato das 426 sentenças classificadas como *Purpose* estarem distribuídas em 375 resumos, o que leva a uma média de 1,13 sentença de *Purpose* nesses textos. Já as sentenças classificadas como *Background* (em maior número, 808 sentenças), estão presentes em 262 resumos e possui média de 3,08 sentenças. Desse modo, percebe-se que o componente *Purpose* está presente na maior parte do corpus (97,4%) com aproximadamente uma única sentença, enquanto o componente *Background* está presente com uma média de três sentenças em 68,05% dos resumos. Essas e outras relações podem ser observadas na Tabela 7.4, que também apresenta a média total de 5,95 sentenças por resumo.

Tabela 7.4: Quantidade média de sentenças por resumo, distribuído por componentes da estrutura retórica

Componente	No. Resumos	No. Sentenças	Sentenças/Resumo
<i>Background</i>	262	808	3,08
<i>Gap</i>	156	215	1,38
<i>Purpose</i>	375	426	1,13
<i>Methodology</i>	145	273	1,88
<i>Result</i>	213	451	2,12
<i>Conclusion</i>	89	120	1,35
Total	385	2.293	5,95

Na verdade, o processo de anotação de cada sentença em uma das seis possíveis categorias retóricas pode ser considerado semi-automático, já que uma classificação prévia foi realizada pelo classificador estatístico AZPort e seus resultados foram corrigidos por um anotador humano. Para essa tarefa, o grau de concordância *Kappa* foi de 40% entre o AZPort e o anotador, conforme pode ser observado na matriz de confusão da Tabela 7.5, em que os resultados da classificação feita pelo AZPort estão representados nas linhas da matriz, enquanto os resultados do anotador humano estão representados nas colunas.

Tabela 7.5: Matriz de confusão: análise manual *vs.* análise automática

	Humano							
	Categoria	B	G	P	M	R	C	Total
AZPort	B	536	71	60	40	22	0	729
	G	24	65	0	5	1	1	96
	P	12	4	191	2	9	3	221
	M	10	1	8	37	19	5	80
	R	223	72	149	182	376	80	1.082
	C	3	2	18	7	24	31	85
	Total	808	215	426	273	451	120	2.293

Na Tabela 7.5 os acertos são mostrados na diagonal principal e os erros fora dessa diagonal. Nota-se que a categoria *Background* (B) obteve o melhor desempenho, dado que o AZPort classificou corretamente 536 do total de 808 sentenças da categoria anotadas pelo anotador humano. Para essa categoria, o AZPort cometeu um número maior de erros ao classificar 223 sentenças incorretamente como *Result* (R). O pior desempenho do classificador se deu na categoria *Methodology* (M), em que a concordância com o anotador humano ocorreu em 37 de um total de 273 sentenças. O AZPort é fortemente dependente do atributo Expressão, sendo esse o que mais contribui para o desempenho do classificador já que possui o maior poder de distinção entre os atributos (Feltrim et al., 2006). Conforme apresentado na Tabela 5.4 da Seção 5.2, o atributo Expressão é calculado a partir da presença de expressões padrão na sentença analisada. Na classificação realizada neste trabalho, o atributo foi calculado com base em um conjunto de expressões regulares extraídas de outro corpus, o que prejudica o desempenho do classificador, como pode ser observado na Tabela 7.5.

Ainda assim, pela matriz de confusão é possível observar que os erros do AZPort se dão na maior parte dos casos em categorias naturalmente difíceis de distinguir em trabalhos de Ciência da Computação, como é o caso das categorias Metodologia X Resultado e Conclusão X Resultado. Isso porque, muitas vezes, uma mesma sentença abrange mais de um componente retórico, o que pode causar dúvidas até mesmo na anotação humana.

Outro fator que penaliza o classificador é o desbalanceamento do corpus. Como o classificador é estatístico e o número de sentenças de *Background* é alto, o classificador tende a atribuir essa categoria mais vezes, pois sua probabilidade de ocorrência no corpus de treinamento é maior. Dessa maneira, conforme relatado anteriormente a categoria *Background* possui o melhor desempenho entre as categorias, com o valor *F-Measure* de 0.697, enquanto a categoria *Conclusion*, com baixo número de sentenças, apresenta um dos piores desempenhos com valor *F-Measure* de 0.302.

Embora o desempenho do classificador AZPort seja importante, não prejudica a construção dos classificadores referentes a este trabalho, já que são utilizados valores corrigidos pelo anotador humano. Da mesma forma, tal desempenho também não afetará no uso desses classificadores no módulo complementar ao SciPo, uma vez que o ambiente permite ao usuário a correção de tal classificação caso julgue necessária. Ainda assim, a melhoria no desempenho do AZPort é importante para aqueles usuários com pouca familiaridade com a estrutura retórica abordada pelo ambiente.

7.4 Segunda Fase de Anotação e Análise

Após a primeira fase da anotação do corpus em que uma das principais tarefas foi a atribuição de uma categoria retórica a cada uma das sentenças, conforme apresentado na Seção 7.2, foi realizada uma segunda fase da anotação. Nesta seção é apresentado o procedimento empregado nessa fase de anotação, bem como uma análise dos resultados obtidos. Nessa fase, foram analisados quatro diferentes tipos de relacionamentos entre as sentenças de diferentes componentes retóricos, denominados neste trabalho de dimensões, como segue:

Dimensão Título: verificou-se a similaridade semântica da sentença analisada com o título do resumo. Nessa dimensão são possíveis dois valores: “alto”, se identificado um relacionamento elevado da sentença com o Título e “baixo”, caso o relacionamento seja baixo. Embora seja possível a existência de um meio termo entre os valores “alto” e “baixo”, considerou-se apenas as duas faixas de valores para que a tarefa de anotação fosse facilitada e menos confusa, já que se trata de um processo subjetivo e a inclusão de mais um valor de similaridade tornaria a anotação pouco eficiente.

De um total de 2.293 sentenças, identificou-se 1.050 (46,80%) com baixo relacionamento com o Título e 1.243 (54,20%) com alto grau de relacionamento. A distribuição de acordo com as seis possíveis categorias retóricas é apresentada na Figura 7.2.

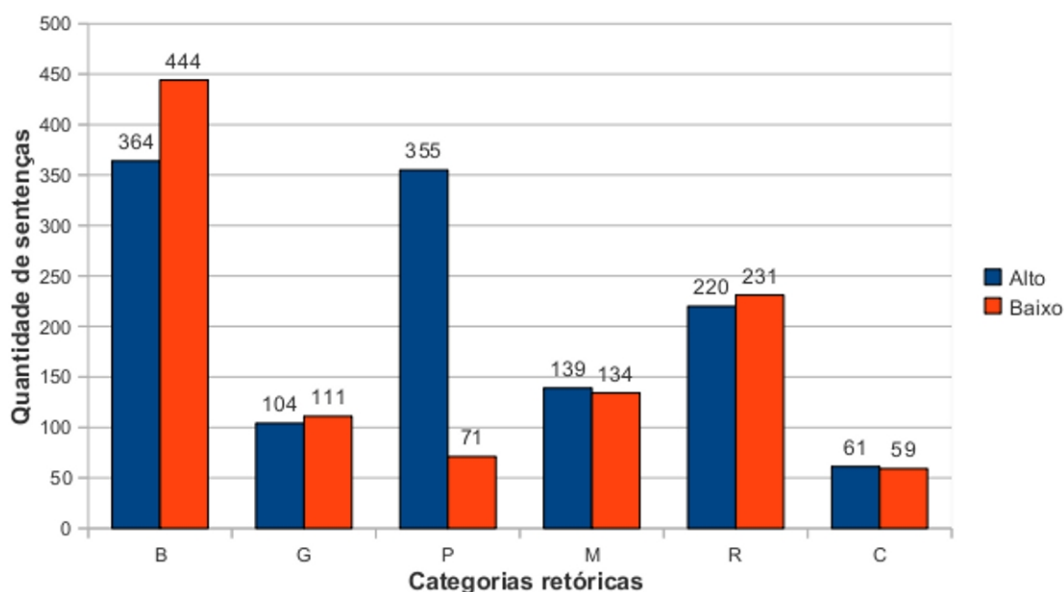


Figura 7.2: Relacionamento das sentenças com o Título de acordo com as categorias retóricas

Observa-se na Figura 7.2 um forte relacionamento entre as sentenças classificadas na categoria *Purpose* (P) e o Título do resumo, já que 355 (83,33%) dessas sentenças possuem alto relacionamento com o Título, enquanto as outras categorias possuem uma média de 48,79% de alto e 51,21% de baixo relacionamento. A categoria menos relacionada é *Background* (B), com 364 (45,05%) das sentenças com alto e 444 (54,95%) das sentenças com baixo relacionamento com o Título do resumo. Esse comportamento é justificável, uma vez que o objetivo das sentenças de *Background* é estabelecer o contexto da pesquisa ainda no início do resumo e, desse modo, em boa parte das vezes esse conteúdo é apenas introdutório e não aborda questões diretamente relacionadas ao título. Por isso, entendemos que o baixo relacionamento entre essas sentenças e o título não caracteriza um problema em relação à coerência. Por outro lado, espera-se que as sentenças de *Purpose* indiquem o propósito da pesquisa, descrevendo resumidamente o objetivo principal do trabalho, enquanto o título seja um resumo ainda mais “condensado” desse objetivo e, por isso, já era esperado um nível elevado de relacionamento semântico na maior parte dos casos analisados. Em relação às outras categorias retóricas, o relacionamento semântico com o título do resumo é bastante dividido, pois depende da natureza do trabalho e da maneira como o autor expõe o problema a ser resolvido, a metodologia utilizada, os resultados e as conclusões, sendo difícil estabelecer um padrão no nível de relacionamento dessas sentenças com o título do resumo, que podem ou não ser relacionadas ao título sem que exista um problema de coerência.

Para medir a reprodutibilidade da anotação da Dimensão Título, isto é, quão semelhante será o resultado da anotação feita por diferentes anotadores, foram realizados experimentos com dois anotadores humanos utilizando-se a medida estatística *Kappa* para medir a concordância entre os anotadores. Antes do início dos experimentos, foi realizado um treinamento com o segundo anotador com o objetivo de se estabelecer um padrão entre as anotações. Para isso, foi analisado e discutido em dupla um conjunto de treinamento com 10 resumos previamente analisados pelo primeiro anotador. Após esse treinamento, realizou-se a anotação sobre aproximadamente 10% do total de resumos do corpus para medir o nível de concordância. Foram analisados 40 resumos, totalizando 209 sentenças das seis possíveis categorias retóricas. Pode-se observar na Tabela 7.6 o valor *Kappa* para cada uma das categorias, bem como o total de sentenças analisadas.

Tabela 7.6: Valores da concordância *Kappa* entre dois anotadores humanos para a Dimensão Título

	Concordância <i>Kappa</i> (<i>K</i>)
Background (N=87)	0,750
Gap (N=46)	0,577
Purpose (N=42)	0,696
Methodology (N=14)	0,512
Result (N=16)	0,625
Conclusion (N=4)	0,500
Todas as sentenças (N=209)	0,610

Na análise do relacionamento de cada categoria com o Título, a que obteve melhor concordância entre os anotadores foi a *Background*, com o valor *Kappa* $K = 0,75$. Por outro lado, a categoria *Conclusion* apresentou a pior concordância, com $K = 0,5$. A categoria na qual espera-se um maior relacionamento com o Título, *Purpose*, obteve o valor de concordância $K = 0,696$, nível considerado “Muito bom” de acordo com as faixas de valores apresentadas na Tabela 6.4 do Capítulo 6. De modo geral, ao considerarmos todas as categorias retóricas, obteve-se o nível de concordância *Kappa*, $K = 0,61$.

Tais resultados permitem concluir que anotadores humanos treinados conseguem distinguir as classes propostas (alto/baixo) no esquema de anotação da Dimensão Título e, desse modo, a anotação resultante de tais experimentos é confiável o suficiente para ser usada como material de treinamento para o classificador automático, em especial, para a categoria retórica *Purpose*.

Concluindo, entendemos que essa dimensão pode indicar um possível problema de coerência entre importantes partes do texto, como o propósito e o título do texto, pois espera-se que o título do trabalho “condense” o seu conteúdo de maneira objetiva.

Assim, assumimos que em um texto coerente, uma de suas características é que sentenças classificadas como *Purpose* sejam semanticamente relacionadas com o seu título.

Dimensão Propósito: verificou-se a similaridade semântica da sentença analisada com a(s) sentença(s) do resumo classificada(s) como Propósito (*Purpose*). Nessa dimensão, são possíveis três valores: “alto”, se identificado um relacionamento elevado da sentença com a(s) sentença(s) classificada(s) como Propósito; “baixo”, caso tal relacionamento seja considerado pequeno pelo anotador e “N/A” (não se aplica), nos casos em que o resumo não possui sentença(s) classificada(s) como *Purpose* ou que a sentença analisada se trata do próprio propósito do resumo.

Excluindo-se um conjunto de 573 sentenças anotadas com o valor “N/A”, divididas em 426 sentenças de *Purpose* e 147 sentenças de diferentes categorias retóricas pertencentes a resumos que não possuem sentenças classificadas como *Purpose*, tem-se um total de 1.720 sentenças. Desse total, foram identificadas 704 (40,93%) sentenças com baixo relacionamento e 1.016 (59,07%) com alto relacionamento com o(s) propósito(s) de seus respectivos resumos.

É apresentada na Figura 7.3 a distribuição das sentenças anotadas nessa dimensão de acordo com as possíveis categorias retóricas, excluindo a categoria *Purpose* (P).

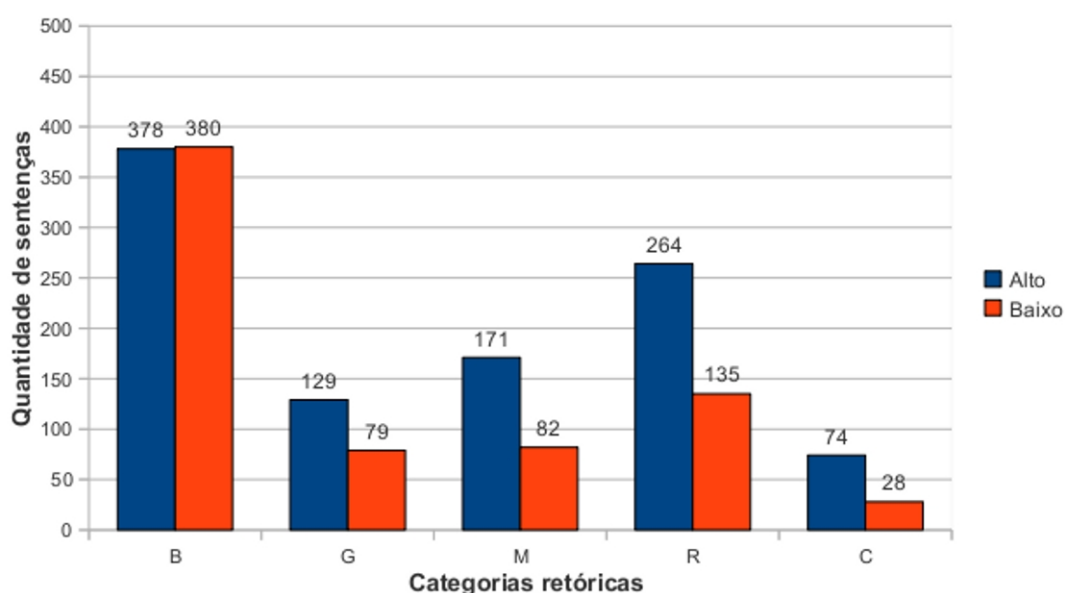


Figura 7.3: Relacionamento das sentenças com o(s) propósito(s) do resumo, de acordo com as categorias retóricas

Observa-se que as categorias mais relacionadas com o propósito são, em ordem: *Conclusion* (C) com 72,55% (74), *Methodology* (M) com 67,59% (171) e *Result* (R) com

66,17% (264) das sentenças com alto relacionamento. Destacamos que foi observado em alguns casos que as sentenças das categorias *Methodology*, *Result* e *Conclusion* retomam os objetivos apresentados no componente Propósito por meio de expressões anafóricas, o que diminui o nível de relacionamento semântico nesses casos. Dessa maneira, embora se tenha verificado um forte relacionamento entre essas sentenças e o propósito do resumo, consideramos o relacionamento semântico como baixo. Novamente, a categoria *Background* (B) se mostrou a menos relacionada, com 50,13% (380) das sentenças com baixo relacionamento com o propósito. Esse comportamento se deve ao fato das sentenças de *Background* se relacionarem com as sentenças de *Gap*, que por sua vez se relacionam fortemente com o componente Propósito em 62,01% (129) dos casos analisados.

Assim como na Dimensão Título, um segundo anotador realizou o processo de anotação da Dimensão Propósito em 40 resumos analisados pelo primeiro anotador, ou seja, 167 sentenças (já que do total de 209 sentenças selecionadas, 42 são do tipo *Purpose* e, por isso, não são analisadas nessa dimensão), com o objetivo de medir a concordância entre as anotações. Novamente, antes do início dos experimentos foi realizado um treinamento com o segundo anotador com o objetivo de se estabelecer um padrão entre as anotações. Para isso, foi analisado e discutido em dupla um conjunto de treinamento com 10 resumos previamente analisados pelo primeiro anotador. No experimento foi utilizada a medida *Kappa*, em que obteve-se $K = 0,815$, nível de concordância considerado “Excelente” de acordo com a Tabela 6.4 apresentada no Capítulo 6. Pode-se observar na Tabela 7.7 o valor *Kappa* para cada uma das cinco categorias retóricas, bem como o total de sentenças analisadas.

Tabela 7.7: Valores da concordância *Kappa* entre dois anotadores humanos para a Dimensão Propósito

	Concordância <i>Kappa</i> (K)
Background (N=87)	0,644
Gap (N=46)	0,804
Methodology (N=14)	0,811
Result (N=16)	0,818
Conclusion (N=4)	1,000
Todas as sentenças (N=167)	0,815

Os resultados apresentados na Tabela 7.7 mostram um alto nível de concordância entre os dois anotadores ao considerarmos as categorias *Conclusion*, *Result* e *Methodology*, com valores que representam um nível de concordância “Excelente”. Esses resultados

confirmam a reprodutibilidade da Dimensão Propósito, de modo que o corpus anotado pode ser utilizado como material de treinamento para essa dimensão.

Concluindo, entendemos que essa dimensão pode indicar possíveis problemas de coerência entre determinadas partes de um resumo acadêmico. Em especial, assumimos que em um resumo coerente, os componentes *Methodology*, *Result* e *Conclusion* devem ter um alto relacionamento semântico com o componente *Purpose*.

Dimensão Lacuna-Contexto: conforme observado na Dimensão Propósito, existe um forte relacionamento entre as sentenças classificadas como *Background* (Contexto) e *Gap* (Lacuna), por isso, nessa dimensão verificou-se em todos os resumos que possuem sentenças de ambas as categorias se a sentença do resumo classificada como *Gap* é semanticamente relacionada com ao menos uma sentença classificada na categoria *Background*. São possíveis dois valores: “sim”, se a sentença da categoria *Gap* é relacionada com ao menos uma sentença do categoria *Background*, e “não”, caso não exista tal relacionamento.

De um total de 215 sentenças classificadas na categoria *Gap*, 32 estão presentes em resumos que não possuem sentenças classificadas como *Background* e, por isso, não foram analisadas. Dessa maneira, do total de 183 sentenças analisadas, 137 (74,86%) foram anotadas com o valor “sim” e 46 (24,14%) foram anotadas com o valor “não”, conforme apresentado na Figura 7.4.

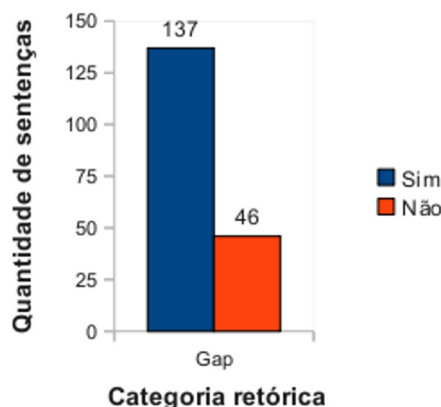


Figura 7.4: Relacionamento das sentenças de *Gap* com as sentenças de *Background*

Assim como nas outras dimensões, um segundo anotador realizou o processo de anotação da Dimensão Lacuna-Contexto em 40 resumos analisados pelo primeiro anotador com o objetivo de medir o nível de concordância. Novamente, antes do início do experimento foi realizado um treinamento com o segundo anotador com o objetivo de se estabelecer um padrão entre as anotações. Para isso, foi analisado e discutido em dupla um

conjunto de treinamento com 10 resumos previamente analisados pelo primeiro anotador. No experimento foram analisadas 46 sentenças classificadas como *Gap* e obteve-se o nível de concordância *Kappa*, $K = 0,725$, valor considerado “Muito bom”. Esse resultado confirma a reprodutibilidade da Dimensão Lacuna-Contexto, de modo que o corpus anotado pode ser utilizado como material de treinamento para essa dimensão.

Como esperado, a maior parte das sentenças do tipo *Gap* possuem um forte relacionamento com ao menos uma sentença do tipo *Background*. Dessa maneira, acreditamos que essa dimensão seja capaz de indicar possíveis problemas de coerência entre essas partes específicas do resumo.

Dimensão Quebra de Linearidade: verificou-se a existência de quebra de linearidade entre sentenças adjacentes do resumo. Nessa dimensão, são possíveis dois valores: “sim”, caso exista alguma dificuldade em se estabelecer um sentido lógico da sentença atual com a sentença anterior e com a próxima sentença; ou “não”, caso a sentença esteja de acordo com o fluxo do texto e o anotador não teve dificuldades em relacioná-la com as sentenças adjacentes. Vale lembrar que em alguns casos verificou-se apenas o relacionamento com a sentença anterior (caso a sentença analisada fosse a última do resumo) ou com a próxima sentença (caso a sentença analisada fosse a primeira do resumo).

Do total de 2.293 sentenças analisadas, houveram apenas 153 sentenças identificadas com o valor “sim” e 2.140 sentenças com o valor “não”. Como é possível observar na Figura 7.5, das 153 sentenças em que o problema foi identificado, a categoria em que o problema é mais recorrente em termos de proporção é a *Result* (R), com 41 (9,09%) das sentenças com o valor “sim”. Já a categoria *Gap* (G) é a que menos apresenta problemas, com apenas 7 (3,26%) ocorrências. Pode-se notar nessa dimensão que dificilmente existe a ocorrência de uma quebra significativa de linearidade entre sentenças adjacentes, já que os casos identificados correspondem a apenas 7,14% do total de sentenças analisadas. Além disso, em boa parte desses casos percebe-se que, em alguma parte do texto, tal sentença é justificada, mesmo não sendo uma sentença adjacente, o que dificulta ainda mais o processo de anotação e análise dessa dimensão.

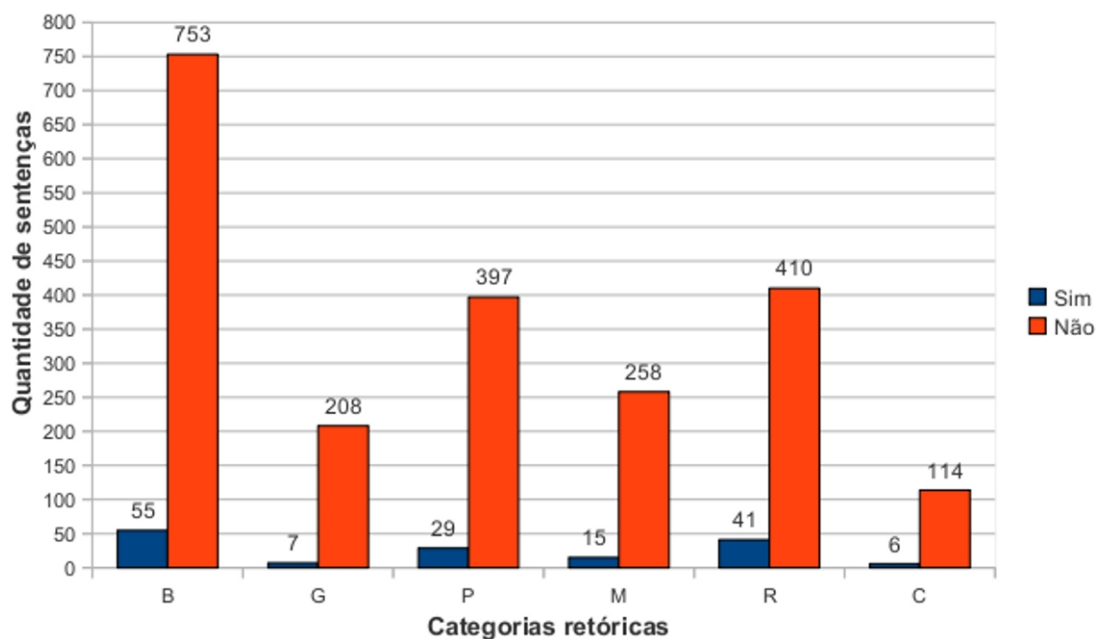


Figura 7.5: Distribuição das categorias retóricas em relação a Dimensão Quebra de Linearidade

Devido ao baixo número de casos identificados, não foi medido o nível de concordância entre diferentes anotadores em relação a essa dimensão, já que seria preciso um número de textos maior do que o utilizado nas outras dimensões (40 resumos) para que tais casos fossem identificados por ambos anotadores.

Mesmo com um baixo número de ocorrências nessa dimensão, entendemos que seu objetivo é importante na tarefa de identificação de problemas na coerência do texto, principalmente em um nível local de análise. Acreditamos que esse tipo de problema pode ser encontrado em textos com graves problemas de escrita e esse não é o caso do corpus analisado, uma vez que é constituído de resumos de trabalhos acadêmicos. Dessa maneira, acreditamos que a dimensão seja válida ao ser aplicada na análise de corpus com um nível maior de “ruído”, como textos gerados automaticamente por sumarizadores.

7.4.1 Ferramenta para Anotação

Para a anotação das dimensões propostas e correção das categorias retóricas atribuídas às sentenças pelo classificador AZPort, foi desenvolvida e utilizada uma ferramenta *online* de anotação, visando facilitar o processo.

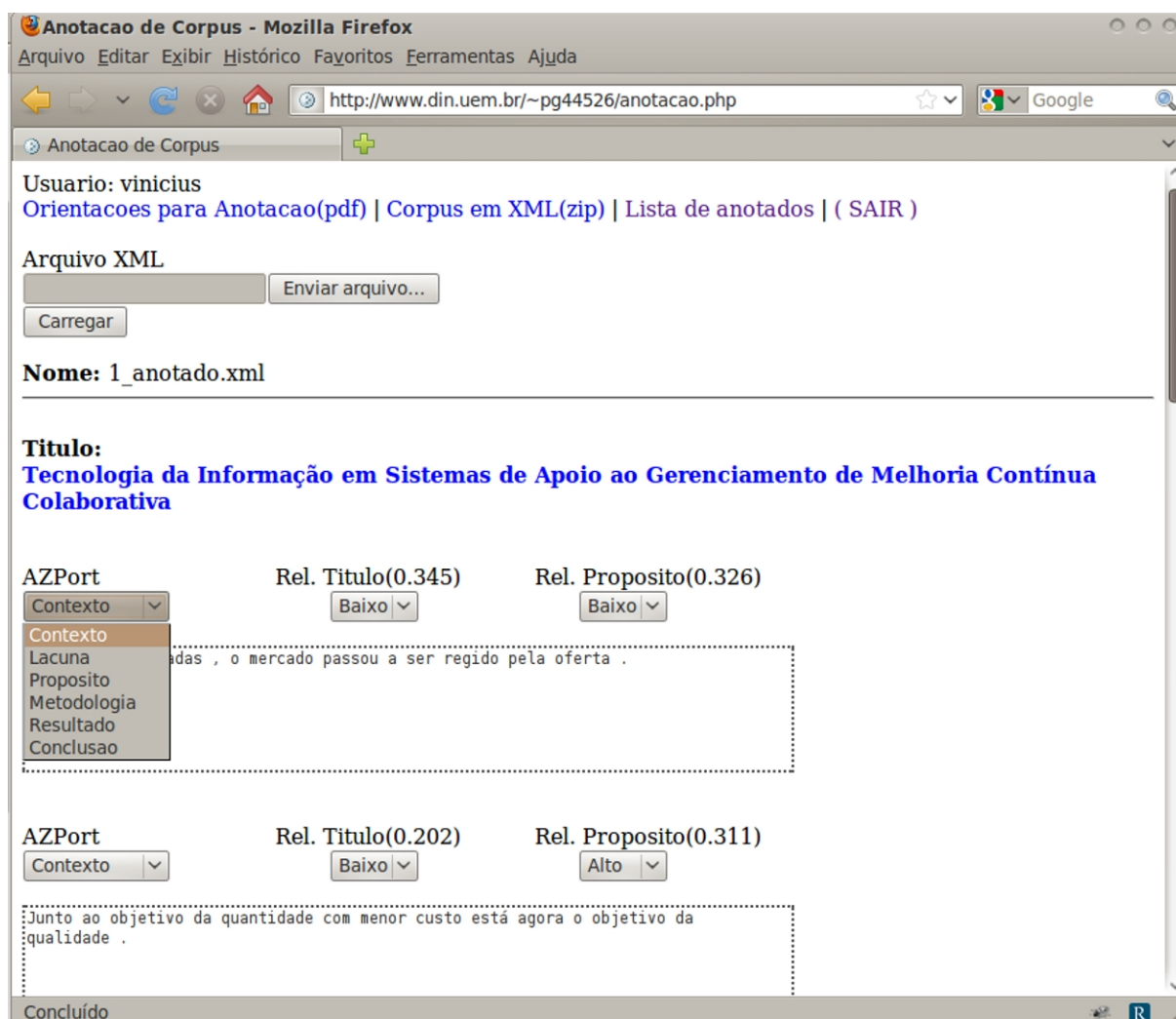


Figura 7.6: Captura de tela da ferramenta *online* para anotação do cópús

A ferramenta, conforme apresentada na Figura 7.6, permite a anotação dos resumos sem o contato direto por parte do anotador com a notação XML utilizada, permitindo assim que o anotador concentre-se nos conceitos envolvidos na anotação, sem se preocupar com questões relacionadas à implementação, como é o caso das etiquetas XML.

Análise Automática de Coerência

O objetivo principal deste trabalho foi o desenvolvimento de um módulo complementar a ser incorporado ao ambiente de auxílio à escrita SciPo. Esse módulo deve ser capaz de identificar aspectos relacionados à coerência semântica em resumos submetidos para análise no SciPo, que já é responsável por fornecer *feedback* ao usuário-escritor sobre a estrutura retórica utilizada.

O *feedback* retornado pelo módulo complementar proposto neste trabalho tem o objetivo de indicar a existência de potenciais problemas relacionados à coerência semântica, mais especificamente, problemas relacionados as quatro dimensões propostas e discutidas no Capítulo 7. A partir da identificação de um potencial problema, a ferramenta permite que o usuário-escritor possa acatar as sugestões apresentadas e corrigir os problemas para, então, novamente submeter o resumo para análise, em um processo cíclico de melhoria. Ou ainda, o usuário-escritor pode simplesmente ignorar as sugestões apresentadas pela ferramenta. É importante ressaltar que os aspectos semânticos são, muitas vezes, ambíguos por natureza, de modo que existe a possibilidade de não se estabelecer uma concordância com o usuário-escritor quanto à identificação de problemas, por isso, a escolha de correção fica a critério do usuário.

Para a automatização das quatro dimensões propostas neste trabalho, decidiu-se pela aplicação de algoritmos de aprendizagem de máquina supervisionada utilizando atributos extraídos da superfície do texto e a partir do método LSA. Para a avaliação intrínseca dos classificadores construídos foram utilizados as métricas apresentadas no Capítulo 6.

Este capítulo apresenta a metodologia de trabalho empregada para se atingir o objetivo de pesquisa proposto, bem como os resultados alcançados. Na Seção 8.1 é apresentada a etapa de pré-processamento do *corpus*, necessária para as etapas de construção do espaço semântico e comparação entre as sentenças, abordadas nas Seções 8.2 e 8.3. Na Seção 8.4 abordamos os atributos extraídos dos resumos analisados e, por fim, apresentamos as etapas de construção e avaliação dos classificadores na Seção 8.6.

8.1 Pré-processamento

Após a coleta, construção, anotação e análise realizada no *corpus*, iniciou-se a fase de desenvolvimento do trabalho a partir dos textos coletados e anotados. A primeira tarefa se refere ao pré-processamento do *corpus*. Tal tarefa foi dividida em três partes, como segue:

Case folding

Com o objetivo de uniformizar os dados, todos os textos do *corpus* passaram por um processo de *case folding*, que transforma todas as letras das palavras em letras minúsculas. Apesar de simples, é um processo importante, já que são realizadas comparações entre as palavras que constituem as sentenças e essas diferenças alteram o resultado esperado. Além disso, também foram removidas as acentuações das palavras.

Remoção de *stopwords*

As *stopwords* consistem em palavras sem relevância aparente para o entendimento do texto, que não são semanticamente significativas ou que se repetem inúmeras vezes no decorrer do texto, como artigos ou preposições e que, por isso, podem ser retiradas do texto. A remoção levou em consideração uma lista elaborada pelo autor com aproximadamente 300 palavras, como apresentado na Tabela A.1 do Apêndice A. Além disso, também foram removidos quaisquer caracteres que não sejam uma letra, como pontuações e números. A remoção de *stopwords* é importante para a redução do custo computacional na construção do espaço semântico utilizado pela LSA e na comparação entre as sentenças.

Radicalização (ou *stemming*)

Após a remoção das *stopwords*, o restante das palavras passaram por um processo de radicalização, também chamado de *stemming*. Esse processo consiste na redução de uma palavra a sua raiz (*stem*) e, dessa maneira, é possível obter uma representação única de variáveis de uma mesma palavra. Por exemplo, as palavras “músico”, “músicos”, “música”

e “musicalização” são representadas por um mesmo *stem*, “music”. O pré-processamento realizado pelo *stemming* é comumente utilizado no método LSA antes da construção do espaço semântico e comparação entre termos.

Existem diferentes algoritmos de *stemming* para diferentes idiomas (Viera e Virgil, 2007). Neste trabalho foi utilizado o algoritmo *Porter Stemmer Algorithm*¹, para o idioma português e implementado na linguagem de programação *Python*², disponível no pacote NLTK³ (*Natural Language Toolkit*).

8.2 Construção do Espaço Semântico

Após o pré-processamento dos dados, foi gerada uma matriz de representação do corpus em que as linhas correspondem às palavras selecionadas (que ocorrem em ao menos dois documentos) e as colunas correspondem aos documentos (resumos). Inicialmente, a cada entrada dessa matriz foi atribuída o valor da frequência de cada palavra em cada documento, seguindo a abordagem *TF-IDF*, em que os valores das entradas recebem um peso de acordo com sua ocorrência (conforme apresentado na Seção 3.4). Obteve-se nessa etapa uma matriz com 4.570 linhas por 385 colunas, ou seja, um total de 4.570 palavras distintas distribuídas em 385 resumos.

O passo seguinte foi a redução da dimensionalidade da matriz para obtenção dos principais padrões associativos nos dados. Para isso, foi aplicada a técnica de decomposição de matrizes em vetores singulares (SVD), em que a matriz é decomposta em três matrizes e o seu produto é a matriz originalmente construída. Com essa decomposição, é possível reduzir a dimensão das três matrizes e, conseqüentemente, da matriz original, mantendo somente os maiores valores e, dessa maneira, descartando informações de menor relevância presentes nos dados. Conforme discutido em Landauer et al. (1998), o número de dimensões a serem reduzidas de forma a otimizar o resultado não é absoluto, variando de 200 à 400 dimensões. Neste trabalho optou-se por um valor intermediário próximo de 300 dimensões após alguns testes.

Com as matrizes reduzidas foi então possível obter um espaço semântico condensado que revela as melhores relações entre as sentenças.

¹<http://tartarus.org/martin/PorterStemmer/python.txt> - Acesso em 01/05/2010

²<http://www.python.org/> - Acesso em 01/05/2010

³<http://www.nltk.org/> - Acesso em 01/05/2010

8.3 Comparação Semântica entre Sentenças

Com o espaço semântico significativo construído a partir do *cópus*, foi possível realizar comparações entre as sentenças que o constituem. Tal comparação levou em conta a similaridade semântica entre cada par de sentenças a partir do modelo estatístico definido na LSA, em que o grau de similaridade se dá pela ocorrência das sentenças em contextos similares.

Para essa tarefa, cada sentença a ser comparada possui uma representação vetorial aceita pelo modelo. Para cada um dos 385 resumos, foram realizadas todas as comparações possíveis entre os pares de sentenças de um mesmo texto, inclusive com as sentenças que constituem o Título. Os valores retornados pela comparação são limitados entre $[-1, 1]$, sendo -1 o menor valor possível de similaridade e 1 o maior valor de similaridade entre as sentenças. Tais valores são utilizados em conjunto com outros atributos para a construção dos modelos de classificação.

Tanto para a construção do espaço semântico como para comparação semântica entre sentenças, foi utilizado o *framework* de desenvolvimento *Gensim* (Řehůřek e Sojka, 2010), um *framework* para a linguagem de programação *Python* que auxilia na conversão de textos para a representação em vetores VSM (*Vector Space Model*), além de contar com a implementação de algoritmos como *Term Frequency - Inverse Document Frequency (TF-IDF)*, *Singular Value Decomposition (SVD)*, *Latent Semantic Analysis (LSA)*, *Latent Dirichlet Allocation (LDA)* e *Random Projections (RP)*.

8.4 Extração de Atributos

Após as etapas de anotação, pré-processamento e comparação semântica entre sentenças, foi possível extrair o conjunto de atributos de cada sentença analisada. Esses atributos foram utilizados na indução de classificadores que tem o objetivo de identificar problemas de coerência nas quatro dimensões propostas: Dimensão Título, Dimensão Propósito, Dimensão Lacuna-Contexto e Dimensão Quebra de Linearidade.

Na Figura 8.1 é apresentado um exemplo de resumo anotado com o conjunto de atributos utilizados pelos classificadores.

```

<RESUMO id="2" tamanhoTitulo="10">
<Titulo>A TECNOLOGIA DA INFORMAÇÃO EM SISTEMAS DE MEDIÇÃO DE DESEMPENHO</Titulo>
<P id="0">
<S AZ="B" AZanterior="N/A" AZproximo="B" elemento_anaforico="nao" id="2-0" lsaAnterior="?"
lsaProposito="0.399" lsaProxima="0.522" lsaTitulo="0.495" offset="0" palavra_transicao="nao"
quebra_linearidade="nao" relProposito="baixo" relTitulo="baixo" tamanhoSentenca="31" >A
crescente competição global que muitos setores industriais vem enfrentando nas últimas
décadas, associado às rápidas mudanças tecnológicas tem forçado as organizações a implementar
novas tecnologias para a prática de gerenciamento.</S>

<S AZ="B" AZanterior="B" AZproximo="P" elemento_anaforico="nao" id="2-1" lsaAnterior="0.522"
lsaProposito="0.675" lsaProxima="0.675" lsaTitulo="0.793" offset="1" palavra_transicao="nao"
quebra_linearidade="nao" relProposito="alto" relTitulo="alto" tamanhoSentenca="38" >A
utilização de sistemas de medição de desempenho (SMD) apoiado na tecnologia da informação e
nos sistemas de informação visa dar suporte aos gestores na escolha do melhor rumo para sua
organização visando melhorar continuamente o seu desempenho.</S>

<S AZ="P" AZanterior="B" AZproximo="N/A" elemento_anaforico="nao" id="2-2" lsaAnterior="0.675"
lsaProposito="1" lsaProxima="?" lsaTitulo="0.677" offset="2" palavra_transicao="nao"
quebra_linearidade="nao" relProposito="n/a" relTitulo="alto" tamanhoSentenca="21" >Esse
trabalho faz uma revisão bibliográfica de ferramentas e tecnologias da computação que oferecem
apoio aos SISTEMAS DE MEDIÇÃO DE DESEMPENHO.</S>
</P>
</RESUMO>

```

Figura 8.1: Exemplo de resumo com anotação final

São extraídas um total de 16 atributos, resumidos na Tabela 8.1. Com exceção dos atributos marcados com o símbolo “*” (*relTitulo*, *relProposito* e *relContexto*), que representam os meta-atributos (ou classe) de cada um dos classificadores e tiveram todos os seus valores atribuídos por um anotador humano, todos os atributos foram extraídos automaticamente de cada sentença pertencente ao corpus.

Os valores dos atributos *AZ*, *AZanterior* e *AZproximo* são extraídos automaticamente pelo classificador estatístico AZPort. Neste trabalho, para evitar que erros da classificação do AZPort influenciassem nos resultados, a classificação foi corrigida por um anotador humano. Para o atributo *AZanterior*, o valor “N/A” é atribuído à primeira sentença de cada resumo, já que a mesma não possui uma sentença anterior. Para o atributo *AZproximo*, o valor “N/A” é atribuído à última sentença de cada resumo, já que a mesma não possui uma próxima sentença.

O atributo *elemento_anaforico* é calculado a partir de uma lista de pronomes elaborada pelo autor e que pode ser consultada na Tabela A.2 no Apêndice A deste trabalho. O objetivo do atributo é a identificação simplificada de algum elemento que possa caracterizar uma anáfora pronominal. Sabemos que o processo de resolução anafórica é muito mais complexo e envolve uma série de questões que fogem do escopo deste trabalho, por isso, reforçamos que esse atributo se trata de uma indicação simplificada.

Tabela 8.1: Conjunto de atributos extraídos do corpus

	Nome	Descrição	Valores
1.	<i>AZ</i>	Categoria retórica da sentença atual	B, G, P, M, R, C
2.	<i>AZanterior</i>	Categoria retórica da sentença anterior	B, G, P, M, R, C, N/A
3.	<i>AZproximo</i>	Categoria retórica da próxima sentença	B, G, P, M, R, C, N/A
4.	<i>elemento_anaforico</i>	Presença de palavra que caracteriza um elemento anafórico	Sim ou Não
5.	<i>lsaContexto</i>	Maior valor de similaridade semântica (LSA) da sentença de <i>Gap</i> com alguma sentença de <i>Background</i>	Valores reais limitados em $[-1, 1]$
6.	<i>lsaAnterior</i>	Valor da similaridade semântica (LSA) da sentença atual com a sentença anterior	Valores reais limitados em $[-1, 1]$
7.	<i>lsaProposito</i>	Valor da similaridade semântica (LSA) da sentença atual com a sentença classificada como <i>Purpose</i>	Valores reais limitados em $[-1, 1]$
8.	<i>lsaProxima</i>	Valor da similaridade semântica (LSA) da sentença atual com a próxima sentença	Valores reais limitados em $[-1, 1]$
9.	<i>lsaTitulo</i>	Valor da similaridade semântica (LSA) da sentença atual com o Título do resumo	Valores reais limitados em $[-1, 1]$
10.	<i>offset</i>	Posição da sentença atual no resumo	Valores inteiros a partir de 0
11.	<i>palavra_transicao</i>	Presença de palavra que caracteriza uma transição	Sim ou Não
12.	<i>tamanhoSentenca</i>	Quantidade de palavras da sentença atual	Valores inteiros a partir de 1
13.	<i>tamanhoTitulo</i>	Quantidade de palavras do Título do resumo	Valores inteiros a partir de 1
14.	<i>relTitulo*</i>	Relacionamento da sentença atual com o Título do resumo	Alto ou Baixo
15.	<i>relProposito*</i>	Relacionamento da sentença atual com a sentença classificada como <i>Purpose</i>	Alto ou Baixo
16.	<i>relContexto*</i>	Presença de uma sentença de <i>Background</i> que se relaciona com a sentença classificada como <i>Gap</i>	Sim ou Não

Os valores dos atributos *lsaContexto*, *lsaAnterior*, *lsaProposito*, *lsaProxima* e *lsaTitulo* são calculados a partir do espaço semântico construído e a comparação entre sentenças, como descrito nas Seções 8.2 e 8.3. Em especial, o atributo *lsaContexto* é calculado e atribuído somente nas sentenças classificadas como *Gap* de resumos que contenham ao menos uma sentença do tipo *Background*.

O atributo *offset* representa a posição da sentença analisada em relação ao início do texto, em que o valor atribuído à primeira sentença do resumo é 0.

O atributo *palavra_transicao* é calculado a partir de uma lista de palavras elaborada pelo autor e que podem indicar algum tipo de transição na sentença, como “no entanto”

ou “embora”. A lista completa de palavras de transição utilizada pode ser consultada na Tabela A.3 do Apêndice A.

Os atributos *tamanhoSentenca* e *tamanhoTitulo* expressam respectivamente a quantidade de palavras da sentença analisada e do Título do resumo analisado.

8.5 Seleção de Atributos

O uso de atributos irrelevantes pode diminuir drasticamente o potencial de acertos de um algoritmo de aprendizagem de máquina e, por isso, é comum uma fase de seleção de atributos. Além disso, a manipulação de um conjunto menor de atributos é menos custosa computacionalmente e simplifica o modelo de classificação (Jain et al., 2000). Essa seleção pode ser realizada manualmente quando se tem um conhecimento profundo do problema modelado ou automaticamente, em que se utiliza métodos de seleção que julgam a relevância dos atributos a partir de medidas estatísticas.

A seleção de atributos de maneira automática pode ser realizada no ambiente Weka, apresentado anteriormente na Seção 6.2. O ambiente fornece duas abordagens distintas para a seleção: métodos *Wrapper*, que utilizam o próprio algoritmo de classificação e um método de busca para encontrar o melhor conjunto de atributos (de acordo com a taxa de acerto) e métodos *Filter*, que independem do algoritmo de aprendizagem escolhido para a indução do classificador e utilizam análises baseadas na medida de entropia para determinar um conjunto de atributos.

De acordo com as conclusões de Hall e Holmes (2003), em um estudo no qual foi realizado um comparativo de técnicas de seleção de atributos, o método *Wrapper* é o melhor mecanismo de seleção em um conjunto de dados pequeno, dado que para grandes conjuntos o método pode ser lento. Em geral, a utilização de métodos *Wrapper* resultam em melhores taxas de acerto do que os métodos *Filter*, uma vez que os subconjuntos de atributos selecionados são avaliados utilizando o próprio algoritmo de classificação que será utilizado posteriormente. Porém, devido a execução do algoritmo previamente e inúmeras vezes, seu custo computacional é superior.

Dado o número de instâncias anotadas em cada dimensão, optou-se pela utilização do método *Wrapper* em conjunto com o método de busca *Best-First* para a escolha de um conjunto de atributos para a indução dos classificadores referentes às dimensões propostas. O método *Wrapper* exige a escolha de um algoritmo específico para a indução do classificador. Neste trabalho, optou-se pelo algoritmo SMO (Keerthi et al., 2001). Tal algoritmo é uma implementação da técnica SVM – *Support Vector Machine* (Vapnik, 2000), um método de aprendizagem baseado em teorias estatísticas em que cria-se um

hiperplano ótimo para a separação das classes. A técnica tem sido utilizada em aplicações de reconhecimento de padrões, tais como categorização de textos (Aizawa, 2001) e categorização de *spam* (Drucker et al., 1999). Também foi utilizada por Higgins et al. (2004) na tarefa de análise de coerência. Desse modo, acreditamos que o algoritmo SMO seja adequado para a tarefa. Os resultados referentes à seleção de atributos para cada dimensão proposta são mostrados na Seção 8.6.

8.6 Construção dos Modelos de Classificação

A partir da anotação apresentada no Capítulo 7 e do conjunto de 16 atributos extraídos das sentenças do corpus (conforme a Tabela 8.1), foram gerados e avaliados modelos de classificação para as dimensões propostas. Foi realizada a etapa de seleção de atributos conforme relatado na Seção 8.5 para cada um dos modelos. Desse modo, o algoritmo SMO foi utilizado para a seleção de atributos e para a indução dos classificadores. Utilizou-se o ambiente Weka para as etapas de seleção de atributos, treinamento e teste dos indutores e na avaliação intrínseca dos métodos construídos. Para o treinamento dos classificadores foi utilizado o método de amostragem *10-fold stratified cross-validation* e alterado o parâmetro *filterType* do algoritmo SMO com o valor “*Standardize training data*”, responsável por normalizar os atributos numéricos de maneira que sua média seja zero e o intervalo de variância seja unitária.

Nas Seções 8.6.1 à 8.6.5 são abordadas as etapas de seleção de atributos, treinamento, teste e avaliação intrínseca dos modelos de classificação referentes às tarefas da Dimensão Título, Dimensão Propósito e Dimensão Lacuna-Contexto.

8.6.1 Dimensão Título

Nesta seção é apresentado o classificador para a Dimensão Título. Resumidamente, a tarefa de tal classificador é prever o grau de relacionamento das sentenças do resumo classificadas como *Purpose* com o seu respectivo Título, podendo assumir os valores “alto” ou “baixo”. O classificador tem o objetivo de identificar um potencial problema de coerência entre esses componentes.

Previamente à indução do classificador final da Dimensão Título, foi realizada a etapa de seleção de atributos para se alcançar o melhor desempenho do classificador.

Seleção de Atributos

A partir do conjunto de 11 atributos extraídos das sentenças para essa dimensão: *AZanterior*, *AZproximo*, *elemento_anaforico*, *lsaAnterior*, *lsaProxima*, *lsaTitulo*, *offset*, *palavra_transicao*, *tamanhoSentenca*, *tamanhoTitulo* e *relTitulo* (classe), foi realizada a seleção automática de atributos. Como resultado, o seguinte conjunto de atributos foi selecionado: *AZproximo*, *elemento_anaforico*, *lsaTitulo*, *offset*, *palavra_transicao*, *tamanhoSentenca* e *relTitulo* (classe). Com o objetivo de comparar o desempenho do classificador utilizando o conjunto de atributos selecionados automaticamente, foram consideradas duas medidas de *baseline*. A primeira *baseline* considera a medida de concordância *Kappa* entre a anotação humana e um classificador treinado com o conjunto de 11 atributos para essa dimensão, conforme mencionado anteriormente. Como *Baseline 2*, considerou-se o valor *Kappa* entre a anotação humana e um classificador com o atributo mais preditivo para a dimensão de acordo com o anotador humano. Dessa maneira, para a *Baseline 2* foram selecionados os atributos *lsaTitulo* e *relTitulo* (classe). Na Tabela 8.2 são apresentados os valores da concordância *Kappa* de cada um dos classificadores com os diferentes conjuntos de atributos.

Tabela 8.2: Concordância *Kappa* dos classificadores da Dimensão Título com diferentes conjuntos de atributos

	Atributos	Concordância <i>Kappa</i> (<i>K</i>)
Classificador	<i>AZproximo</i> , <i>elemento_anaforico</i> , <i>lsaTitulo</i> , <i>offset</i> , <i>palavra_transicao</i> , <i>tamanhoSentenca</i> , <i>relTitulo</i> (classe)	0,871
<i>Baseline 1</i>	<i>AZanterior</i> , <i>AZproximo</i> , <i>elemento_anaforico</i> , <i>lsaAnterior</i> , <i>lsaProxima</i> , <i>lsaTitulo</i> , <i>offset</i> , <i>palavra_transicao</i> , <i>tamanhoSentenca</i> , <i>tamanhoTitulo</i> , <i>relTitulo</i> (classe)	0,844
<i>Baseline 2</i>	<i>lsaTitulo</i> , <i>relTitulo</i> (classe)	0,836

Nota-se na Tabela 8.2 que a concordância *Kappa* entre a anotação humana e o classificador com o conjunto de atributos selecionados automaticamente supera os valores considerados nas duas *baselines*. Assim, os atributos *AZproximo*, *elemento_anaforico*, *lsaTitulo*, *offset*, *palavra_transicao*, *tamanhoSentenca*, e *relTitulo* (classe) foram utilizados para a indução do classificador final da Dimensão Título.

Avaliação Intrínseca

Dado os 7 atributos selecionados, um conjunto de 426 sentenças classificadas como *Purpose* foi utilizado para treinamento e teste do classificador final da Dimensão Título. Desse total, 411 (96,48%) foram corretamente classificadas e 15 (3,52%) foram incorretamente classificadas. Os detalhes do teste do classificador são apresentados na matriz de confusão da Tabela 8.3, em que os acertos são mostrados na diagonal principal e os erros fora dessa diagonal.

Tabela 8.3: Matriz de confusão do classificador final da Dimensão Título

Anotador	Classe	Classificador		Total
		Alto	Baixo	
	Alto	349	6	355
	Baixo	9	62	71
Humano	Total	358	68	426

Pode-se observar na Tabela 8.3 que do total de 355 sentenças que realmente pertencem à classe “alto” (classificadas pelo anotador humano), 349 sentenças foram corretamente classificadas como “alto” e 6 sentenças foram incorretamente classificadas como “baixo” pelo classificador. Já do total de 71 sentenças que realmente pertencem à classe “baixo” (classificadas pelo anotador humano), 62 sentenças foram corretamente classificadas como “baixo” e 9 sentenças foram incorretamente classificadas como “alto” pelo classificador. No total, foram classificadas 358 sentenças como “alto”, em que 9 sentenças pertencem à classe “baixo”, ou seja, incorretamente classificadas, e 68 sentenças classificadas como “baixo”, em que 6 pertencem à classe “alto” e, portanto, incorretamente classificadas.

Na Tabela 8.4 é apresentado o desempenho do classificador em relação às medidas de avaliação *Precision*, *Recall*, *F-Measure* e *Macro-F*. A Tabela 8.4 também apresenta os resultados de uma *baseline* simples que sempre atribui a classe prevalente como saída para um classificador hipotético. Como relatado anteriormente na Seção 7.4, a classe prevalente na anotação da Dimensão Título, em especial para as sentenças classificadas como *Purpose*, é a classe “alto” com 83,33% das sentenças.

Tabela 8.4: Desempenho do classificador final da Dimensão Título

	Alto			Baixo			Macro-F
	Precision	Recall	F-measure	Precision	Recall	F-measure	
Classificador	0,975	0,983	0,979	0,912	0,873	0,892	0,936
Baseline	0,833	1,000	0,908	0,000	0,000	0,000	0,454

Observa-se na Tabela 8.4 que o classificador supera os resultados da *baseline*. Nota-se também que o valor da medida *F-measure* da classe “alto” é superior ao da classe “baixo”. Embora exista um desbalanceamento das classes que possa favorecer a classe “alto”, acredita-se que o comportamento do classificador se deve ao fato de ser mais fácil afirmar a existência de um relacionamento alto entre uma sentença da categoria *Purpose* e o Título do resumo do que um relacionamento baixo. Logo, existe uma ambiguidade maior para identificação de um relacionamento baixo. Esse fato foi constatado durante o processo de anotação manual e acredita-se que a dificuldade, ainda que sensível, tenha refletido no classificador.

Desse modo, os resultados positivos obtidos na avaliação do classificador possibilitam a sua utilização para desempenhar a tarefa proposta pela Dimensão Título.

8.6.2 Dimensão Propósito (componente Methodology)

Nesta seção é apresentado o classificador para a Dimensão Propósito referente ao componente *Methodology*. Resumidamente, a tarefa de tal classificador é prever o grau de relacionamento das sentenças do resumo classificadas como *Methodology* com a sentença classificada como *Purpose*, podendo assumir os valores “alto” ou “baixo”. O objetivo do classificador é identificar um potencial problema de coerência desse componente com o propósito do resumo.

Previamente à indução do classificador final da Dimensão Propósito com o componente *Methodology*, foi realizada a etapa de seleção de atributos para se alcançar o melhor desempenho do classificador.

Seleção de Atributos

A partir do conjunto de 12 atributos extraídos das sentenças para essa dimensão: *AZanterior*, *AZproximo*, *elemento_anaforico*, *lsaAnterior*, *lsaProposito*, *lsaProxima*, *lsaTitulo*, *offset*, *palavra_transicao*, *tamanhoSentenca*, *tamanhoTitulo* e *relProposito* (classe), foi realizada a seleção automática de atributos. Como resultado, o seguinte conjunto

de atributos foi selecionado: *AZanterior*, *lsaAnterior*, *lsaProposito*, *palavra_transicao* e *relProposito* (classe). Com o objetivo de comparar o desempenho do classificador utilizando o conjunto de atributos selecionados automaticamente, foram consideradas duas medidas de *baseline*. A primeira *baseline* considera a medida de concordância *Kappa* entre a anotação humana e o classificador treinado com o conjunto de 12 atributos para essa dimensão, conforme mencionado anteriormente. Como *Baseline 2*, considerou-se o valor *Kappa* entre a anotação humana e um classificador com o atributo mais preditivo para a dimensão de acordo com o anotador humano. Dessa maneira, para a *Baseline 2* foram selecionados os atributos *lsaProposito* e *relProposito* (classe). Na Tabela 8.5 são apresentados os valores da concordância *Kappa* de cada um dos classificadores com esses conjuntos de atributos.

Tabela 8.5: Concordância *Kappa* dos classificadores da Dimensão Propósito (envolvendo o componente *Methodology*) com diferentes conjuntos de atributos

	Atributos	Concordância <i>Kappa</i> (<i>K</i>)
Classificador	<i>AZanterior</i> , <i>lsaAnterior</i> , <i>lsaProposito</i> , <i>palavra_transicao</i> , <i>relProposito</i> (classe)	0,683
<i>Baseline 1</i>	<i>AZanterior</i> , <i>AZproximo</i> , <i>elemento_anaforico</i> , <i>lsaAnterior</i> , <i>lsaProposito</i> , <i>lsaProxima</i> , <i>lsa-Titulo offset</i> , <i>palavra_transicao</i> , <i>tamanhoSentenca</i> , <i>relProposito</i> (classe)	0,647
<i>Baseline 2</i>	<i>lsaProposito</i> , <i>relProposito</i> (classe)	0,643

Nota-se na Tabela 8.5 que a concordância *Kappa* entre a anotação humana e o classificador com o conjunto de atributos selecionados automaticamente supera os valores considerados nas duas *baselines*. Assim, os atributos *AZanterior*, *lsaAnterior*, *lsaProposito*, *palavra_transicao* e *relProposito* (classe) foram utilizados para a indução do classificador final da Dimensão Propósito envolvendo o componente *Methodology*.

Avaliação Intrínseca

Dado os 5 atributos selecionados, um conjunto de 253 sentenças classificadas como *Methodology* foi utilizado para treinamento e teste do classificador final da Dimensão Propósito envolvendo o componente *Methodology* (uma vez que do total de 273 sentenças de *Methodology* do cópua, 20 estão resumos que não possuem o componente *Purpose*). Desse total, 218 (86,17%) foram corretamente classificadas e 35 (13,83%) foram incorretamente classificadas. Os detalhes do teste do classificador são apresentados na matriz de

confusão da Tabela 8.6, em que os acertos são mostrados na diagonal principal e os erros fora dessa diagonal.

Tabela 8.6: Matriz de confusão do classificador final da Dimensão Propósito (componente *Methodology*)

Anotador	Classificador			
	Classe	Alto	Baixo	Total
	Alto	154	17	171
	Baixo	18	64	82
Humano	Total	172	81	253

Pode-se observar na Tabela 8.6 que do total de 171 sentenças que realmente pertencem à classe “alto” (classificadas pelo anotador humano), 154 sentenças foram corretamente classificadas como “alto” e 17 sentenças foram incorretamente classificadas como “baixo” pelo classificador. Já do total de 82 sentenças que realmente pertencem à classe “baixo” (classificadas pelo anotador humano), 64 sentenças foram corretamente classificadas como “baixo” e 18 sentenças foram incorretamente classificadas como “alto” pelo classificador. No total, foram classificadas 172 sentenças como “alto”, em que 18 sentenças pertencem à classe “baixo”, ou seja, incorretamente classificadas, e 81 sentenças classificadas como “baixo”, em que 17 pertencem à classe “alto” e, portanto, incorretamente classificadas.

Na Tabela 8.7 é apresentado o desempenho do classificador em relação às medidas de avaliação *Precision*, *Recall*, *F-Measure* e *Macro-F*. A Tabela 8.7 também apresenta os resultados de uma *baseline* simples que sempre atribui a classe prevalente da anotação como saída para um classificador hipotético. Como relatado anteriormente na Seção 7.4, a classe prevalente na anotação da Dimensão Propósito, em especial para as sentenças classificadas como *Methodology*, é a classe “alto” com 67,59% das sentenças.

Tabela 8.7: Desempenho do classificador final da Dimensão Propósito (componente *Methodology*)

	Alto			Baixo			Macro-F
	Precision	Recall	F-measure	Precision	Recall	F-measure	
Classificador	0,895	0,901	0,898	0,79	0,78	0,785	0,842
Baseline	0,675	1,000	0,795	0,000	0,000	0,000	0,398

Observa-se na Tabela 8.7 que o classificador supera os resultados da *baseline*. Nota-se que o valor da medida *F-measure* da classe “alto” é superior ao da classe “baixo”. Embora

exista um desbalanceamento das classes que possa favorecer a classe “alto”, assim como na Dimensão Título, acredita-se que o comportamento do classificador se deve ao fato de ser mais fácil afirmar a existência de um relacionamento alto entre uma sentença da categoria *Methodology* e o propósito do resumo do que um relacionamento baixo. Logo, existe uma ambiguidade maior para identificação de um relacionamento baixo. Esse fato foi constatado durante o processo de anotação manual e acredita-se que a dificuldade tenha refletido no classificador. Além disso, em boa parte dos casos o conteúdo das sentenças de *Methodology* introduz o nome de técnicas e métodos utilizados no desenvolvimento do trabalho. Esses termos possuem nomes próprios e isso faz com que a similaridade semântica com as sentenças de *Purpose*, calculada a partir do atributo *lsaProposito*, seja baixa e, dessa maneira, contradiz o anotador humano que consegue analisar além da superfície do texto. Assim, o estabelecimento da classe “baixo” é dificultado. Entretanto, o classificador obteve um bom desempenho de modo geral com o valor 0,842 para a medida *Macro-F*, que leva em consideração ambas as classes.

Desse modo, os resultados positivos obtidos na avaliação do classificador possibilitam a sua utilização para desempenhar a tarefa proposta pela Dimensão Propósito envolvendo o componente *Methodology*.

8.6.3 Dimensão Propósito (componente Result)

Nesta seção é apresentado o classificador para a Dimensão Propósito referente ao componente *Result*. Resumidamente, a tarefa de tal classificador é predizer o grau de relacionamento das sentenças do resumo classificadas como *Result* com a sentença classificada como *Purpose*, podendo assumir os valores “alto” ou “baixo”. O objetivo do classificador é identificar um potencial problema de coerência desse componente com o propósito do resumo.

Previamente à indução do classificador final da Dimensão Propósito com o componente *Result*, foi realizada a etapa de seleção de atributos para se alcançar o melhor desempenho do classificador.

Seleção de Atributos

A partir do conjunto de 12 atributos extraídos das sentenças para essa dimensão: *AZanterior*, *AZproximo*, *elemento_anaforico*, *lsaAnterior*, *lsaProposito*, *lsaProxima*, *lsaTitulo*, *offset*, *palavra_transicao*, *tamanhoSentenca*, *tamanhoTitulo* e *relProposito* (classe), foi realizada a seleção automática de atributos. Como resultado, o seguinte conjunto de atributos foi selecionado: *lsaProposito*, *lsaProxima*, *lsaTitulo* e *relProposito* (classe). Com

o objetivo de comparar o desempenho do classificador utilizando o conjunto de atributos selecionados automaticamente, foram consideradas duas medidas de *baseline*. A primeira considera a medida de concordância *Kappa* entre a anotação humana e o classificador treinado com o conjunto de 12 atributos dessa dimensão, conforme mencionado anteriormente. Como *Baseline 2*, considerou-se o valor *Kappa* entre a anotação humana e um classificador com o atributo mais preditivo para a dimensão de acordo com o anotador humano. Dessa maneira, para a *Baseline 2* foram selecionados os atributos *lsaProposito* e *relProposito* (classe). Na Tabela 8.8 são apresentados os valores da concordância *Kappa* de cada um dos classificadores com esses conjuntos de atributos.

Tabela 8.8: Concordância *Kappa* dos classificadores da Dimensão Propósito (envolvendo o componente *Result*) com diferentes conjuntos de atributos

	Atributos	Concordância <i>Kappa</i> (<i>K</i>)
Classificador	<i>lsaProposito</i> , <i>lsaProxima</i> , <i>lsaTitulo</i> , <i>relProposito</i> (classe)	0,763
<i>Baseline 1</i>	<i>AZanterior</i> , <i>AZproximo</i> , <i>elemento_anaforico</i> , <i>lsaAnterior</i> , <i>lsaProposito</i> , <i>lsaProxima</i> , <i>lsaTitulo</i> , <i>offset</i> , <i>palavra_transicao</i> , <i>tamanhoSentenca</i> , <i>relProposito</i> (classe)	0,685
<i>Baseline 2</i>	<i>lsaProposito</i> , <i>relProposito</i> (classe)	0,728

Nota-se na Tabela 8.8 que a concordância *Kappa* entre a anotação humana e o classificador com o conjunto de atributos selecionados automaticamente supera os valores considerados nas duas *baselines*. Assim, os atributos *lsaProposito*, *lsaProxima*, *lsaTitulo* e *relProposito* (classe) foram utilizados para a indução do classificador final da Dimensão Propósito envolvendo o componente *Result*.

Avaliação Intrínseca

Dado os 4 atributos selecionados, um conjunto de 399 sentenças classificadas como *Result* foi utilizado para treinamento e teste do classificador final da Dimensão Propósito envolvendo o componente *Result* (uma vez que do total de 451 sentenças de *Result* do corpus, 52 estão em resumos que não possuem o componente *Purpose*). Desse total, 357 (89,47%) foram corretamente classificadas e 42 (10,52%) foram incorretamente classificadas. Os detalhes do teste do classificador são apresentados na matriz de confusão da Tabela 8.9, em que os acertos são mostrados na diagonal principal e os erros fora dessa diagonal.

Tabela 8.9: Matriz de confusão do classificador final da Dimensão Propósito (componente *Result*)

Anotador	Classificador			
	Classe	Alto	Baixo	Total
	Alto	245	19	264
	Baixo	23	112	135
Humano	Total	268	131	399

Pode-se observar na Tabela 8.9 que do total de 264 sentenças que realmente pertencem à classe “alto” (classificadas pelo anotador humano), 245 sentenças foram corretamente classificadas como “alto” e 19 sentenças foram incorretamente classificadas como “baixo” pelo classificador. Já do total de 135 sentenças que realmente pertencem à classe “baixo” (classificadas pelo anotador humano), 112 sentenças foram corretamente classificadas como “baixo” e 23 sentenças foram incorretamente classificadas como “alto” pelo classificador. No total, foram classificadas 268 sentenças como “alto”, em que 23 sentenças pertencem à classe “baixo”, ou seja, incorretamente classificadas, e 131 sentenças classificadas como “baixo”, em que 19 pertencem à classe “alto” e, portanto, incorretamente classificadas.

Na Tabela 8.10 é apresentado o desempenho do classificador em relação às medidas de avaliação *Precision*, *Recall*, *F-Measure* e *Macro-F*. A Tabela 8.10 também apresenta os resultados de uma *baseline* simples que sempre atribui a classe prevalente da anotação como saída para um classificador hipotético. Como relatado anteriormente na Seção 7.4, a classe prevalente na anotação da Dimensão Propósito, em especial para as sentenças classificadas como *Result*, é a classe “alto” com 66,17% das sentenças.

Tabela 8.10: Desempenho do classificador final da Dimensão Propósito (componente *Result*)

	Alto			Baixo			Macro-F
	Precision	Recall	F-measure	Precision	Recall	F-measure	
Classificador	0,914	0,928	0,921	0,855	0,83	0,842	0,882
Baseline	0,661	1,000	0,840	0,000	0,000	0,000	0,420

Observa-se na Tabela 8.10 que o classificador supera os resultados da *baseline*. Assim como nas outras dimensões apresentadas, nota-se que o valor da medida *F-measure* da classe “alto” é superior ao da classe “baixo”. Acredita-se que o comportamento do classificador se deve ao fato de ser mais fácil afirmar a existência de um relacionamento alto

entre uma sentença da categoria *Result* e o propósito do resumo do que um relacionamento baixo. Logo, existe uma ambiguidade maior para identificação de um relacionamento baixo. Esse fato foi constatado durante o processo de anotação manual e acredita-se que a dificuldade tenha refletido no classificador. Além disso, em boa parte dos casos o conteúdo das sentenças de *Result* introduz o nome de métricas utilizadas na avaliação dos resultados. Esses termos possuem nomes próprios e isso faz com que a similaridade semântica com as sentenças de *Purpose*, calculada a partir do atributo *lsaProposito*, seja baixa e, dessa maneira, contradiz o anotador humano que consegue analisar além da superfície do texto. Assim, o estabelecimento da classe “baixo” é dificultado. Entretanto, o classificador obteve um bom desempenho de modo geral com o valor 0,882 para a medida *Macro-F*, que leva em consideração ambas as classes.

Desse modo, os resultados positivos obtidos na avaliação do classificador possibilitam a sua utilização para desempenhar a tarefa proposta pela Dimensão Propósito envolvendo o componente *Result*.

8.6.4 Dimensão Propósito (componente Conclusion)

Nesta seção é apresentado o classificador para a Dimensão Propósito referente ao componente *Conclusion*. Resumidamente, a tarefa de tal classificador é predizer o grau de relacionamento das sentenças do resumo classificadas como *Conclusion* com a sentença classificada como *Purpose*, podendo assumir os valores “alto” ou “baixo”. O objetivo do classificador é identificar um potencial problema de coerência desse componente com o propósito do resumo.

Previamente à indução do classificador final da Dimensão Propósito com o componente *Conclusion*, foi realizada a etapa de seleção de atributos para se alcançar o melhor desempenho do classificador.

Seleção de Atributos

A partir do conjunto de 12 atributos extraídos das sentenças para essa dimensão: *AZanterior*, *AZproximo*, *elemento_anaforico*, *lsaAnterior*, *lsaProposito*, *lsaProxima*, *lsaTitulo*, *offset*, *palavra_transicao*, *tamanhoSentenca*, *tamanhoTitulo* e *relProposito* (classe), foi realizada a seleção automática de atributos. Como resultado, os atributos *AZproximo*, *lsaProposito* e *relProposito* (classe) foram selecionados. Com o objetivo de comparar o desempenho do classificador utilizando o conjunto de atributos selecionados automaticamente, foram consideradas duas medidas de *baseline*. A primeira considera a medida de concordância *Kappa* entre a anotação humana e um classificador treinado com o conjunto

de 12 atributos selecionados para essa dimensão, conforme relatado anteriormente. Como *Baseline 2*, considerou-se o valor *Kappa* entre a anotação humana e um classificador com o atributo mais preditivo para a dimensão de acordo com o anotador humano. Dessa maneira, para a *Baseline 2* foram selecionados os atributos *lsaProposito* e *relProposito* (classe). Na Tabela 8.11 são apresentados os valores da concordância *Kappa* de cada um dos classificadores com esses conjuntos de atributos.

Tabela 8.11: Concordância *Kappa* dos classificadores da Dimensão Propósito (envolvendo o componente *Conclusion*) com diferentes conjuntos de atributos

	Atributos	Concordância <i>Kappa</i> (<i>K</i>)
Classificador	<i>AZproximo</i> , <i>lsaProposito</i> , <i>relProposito</i> (classe)	0,720
<i>Baseline 1</i>	<i>AZanterior</i> , <i>AZproximo</i> , <i>elemento_anaforico</i> , <i>lsaAnterior</i> , <i>lsaProposito</i> , <i>lsaProxima</i> , <i>lsaTitulo</i> , <i>offset</i> , <i>palavra_transicao</i> , <i>tamanhoSentenca</i> , <i>relProposito</i> (classe)	0,546
<i>Baseline 2</i>	<i>lsaProposito</i> , <i>relProposito</i> (classe)	0,748

Nota-se na Tabela 8.11 que a concordância *Kappa* entre a anotação humana e o classificador utilizando somente os atributos *lsaProposito* e *relProposito* (classe) (*Baseline 2*) supera a *Baseline 1* e o classificador com o conjunto de atributos selecionados automaticamente. Assim, somente os atributos *lsaProposito* e *relProposito* (classe) foram utilizados para a indução do classificador final da Dimensão Propósito envolvendo o componente *Conclusion*.

Avaliação Intrínseca

Dado os atributos *lsaProposito* e *relProposito* (classe) selecionados, um conjunto de 102 sentenças classificadas como *Conclusion* foi utilizado para treinamento e teste do classificador final da Dimensão Propósito envolvendo o componente *Conclusion* (uma vez que do total de 120 sentenças de *Conclusion* do corpus, 18 estão resumos que não possuem o componente *Purpose*). Desse total, 92 (90,19%) foram corretamente classificadas e 10 (9,80%) foram incorretamente classificadas. Os detalhes do teste do classificador são apresentados na matriz de confusão da Tabela 8.12, em que os acertos são mostrados na diagonal principal e os erros fora dessa diagonal.

Tabela 8.12: Matriz de confusão do classificador final da Dimensão Propósito (componente *Conclusion*)

Anotador	Classificador			
	Classe	Alto	Baixo	Total
	Alto	70	4	74
	Baixo	6	22	28
Humano	Total	76	26	102

Pode-se observar na Tabela 8.12 que do total de 74 sentenças que realmente pertencem à classe “alto” (classificadas pelo anotador humano), 70 sentenças foram corretamente classificadas como “alto” e 4 sentenças foram incorretamente classificadas como “baixo” pelo classificador. Já do total de 28 sentenças que realmente pertencem à classe “baixo” (classificadas pelo anotador humano), 22 sentenças foram corretamente classificadas como “baixo” e 6 sentenças foram incorretamente classificadas como “alto” pelo classificador. No total, foram classificadas 76 sentenças como “alto”, em que 6 sentenças pertencem à classe “baixo”, ou seja, incorretamente classificadas, e 26 sentenças classificadas como “baixo”, em que 4 pertencem à classe “alto” e, portanto, incorretamente classificadas.

Na Tabela 8.13 é apresentado o desempenho do classificador em relação às medidas de avaliação *Precision*, *Recall*, *F-Measure* e *Macro-F*. A Tabela 8.13 também apresenta os resultados de uma *baseline* simples que sempre atribui a classe prevalente da anotação como saída para um classificador hipotético. Como relatado anteriormente na Seção 7.4, a classe prevalente na anotação da Dimensão Propósito, em especial para as sentenças classificadas como *Conclusion*, é a classe “alto” com 72,55% das sentenças.

Tabela 8.13: Desempenho do classificador final da Dimensão Propósito (componente *Conclusion*)

	Alto			Baixo			Macro-F
	Precision	Recall	F-measure	Precision	Recall	F-measure	
Classificador	0,921	0,946	0,933	0,846	0,786	0,815	0,874
Baseline	0,725	1,000	0,840	0,000	0,000	0,000	0,420

Observa-se na Tabela 8.13 que o classificador supera os resultados da *baseline*. Assim como nas dimensões anteriores, nota-se que o valor da medida *F-measure* da classe “alto” é superior ao da classe “baixo”. Embora exista um desbalanceamento das classes que possa favorecer a classe “alto”, acredita-se que o comportamento do classificador se deve ao fato de ser mais fácil afirmar a existência de um relacionamento alto entre uma sentença

da categoria *Conclusion* e o propósito do resumo do que um relacionamento baixo. Logo, existe uma ambiguidade maior para identificação de um relacionamento baixo. Esse fato foi constatado durante o processo de anotação manual e acredita-se que a dificuldade tenha refletido no classificador. Entretanto, o classificador obteve um bom desempenho de modo geral com o valor 0,874 para a medida *Macro-F*, que leva em consideração ambas as classes.

Desse modo, os resultados positivos obtidos na avaliação do classificador possibilitam a sua utilização para desempenhar a tarefa proposta pela Dimensão Propósito envolvendo o componente *Conclusion*.

8.6.5 Dimensão Lacuna-Contexto

Nesta seção é apresentado o classificador para a Dimensão Lacuna-Contexto. Resumidamente, a tarefa de tal classificador é verificar no resumo analisado, se existe a presença de ao menos uma sentença da categoria *Background* que seja semanticamente relacionada à sentença de *Gap*, podendo assumir os valores “sim” ou “não”, com o objetivo de identificar um potencial problema de coerência entre esses componentes.

Previamente à indução do classificador final da Dimensão Lacuna-Contexto, foi realizada a etapa de seleção de atributos para se alcançar o melhor desempenho do classificador.

Seleção de Atributos

A partir do conjunto de 11 atributos extraídos das sentenças para essa dimensão: *AZanterior*, *AZproximo*, *elemento_anaforico*, *lsaContexto*, *lsaAnterior*, *lsaProposito*, *lsaProxima*, *offset*, *palavra_transicao*, *tamanhoSentenca* e *relProposito* (classe), foi realizada a seleção automática de atributos. Como resultado, os atributos *lsaContexto*, *palavra_transicao* e *relContexto* (classe) foram selecionados. Com o objetivo de comparar o desempenho do classificador utilizando o conjunto de atributos selecionados automaticamente, foram consideradas duas medidas de *baseline*. A primeira considera a medida de concordância *Kappa* entre a anotação humana e o classificador treinado com o conjunto de 12 atributos dessa dimensão, conforme mencionado anteriormente. Como *Baseline 2*, considerou-se o valor *Kappa* entre a anotação humana e um classificador com o atributo mais preditivo para a dimensão de acordo com o anotador humano. Dessa maneira, para a *Baseline 2* foram selecionados os atributos *lsaContexto* e *relContexto* (classe). Na Tabela 8.14 são apresentados os valores da concordância *Kappa* de cada um dos classificadores com esses conjuntos de atributos.

Tabela 8.14: Concordância *Kappa* dos classificadores da Dimensão Lacuna-Contexto com diferentes conjuntos de atributos

	Atributos	Concordância <i>Kappa</i> (<i>K</i>)
Classificador	<i>lsaContexto</i> , <i>palavra_transicao</i> , <i>relContexto</i> (classe)	0,679
<i>Baseline 1</i>	<i>AZanterior</i> , <i>AZproximo</i> , <i>elemento_anaforico</i> , <i>lsaContexto</i> , <i>lsaAnterior</i> , <i>lsaProposito</i> , <i>lsaProxima</i> , <i>offset</i> , <i>palavra_transicao</i> , <i>tamanhoSentenca</i> , <i>relContexto</i> (classe)	0,605
<i>Baseline 2</i>	<i>lsaContexto</i> , <i>relContexto</i> (classe)	0,653

Nota-se na Tabela 8.14 que a concordância *Kappa* entre a anotação humana e o classificador com o conjunto de atributos selecionados automaticamente supera os valores considerados nas duas *baselines*. Assim, os atributos *lsaContexto*, *palavra_transicao* e *relContexto* (classe) foram utilizados para a indução do classificador final da Dimensão Lacuna-Contexto.

Avaliação Intrínseca

Dado os 3 atributos selecionados, um conjunto de 183 sentenças classificadas como *Gap* foi utilizado para treinamento e teste do classificador final da Dimensão Lacuna-Contexto (uma vez que do total de 215 sentenças de *Gap* do cópulo, 32 estão resumos que não possuem o componente *Background*). Desse total, 162 (88,52%) foram corretamente classificadas e 21 (11,48%) foram incorretamente classificadas. Os detalhes do teste do classificador são apresentados na matriz de confusão da Tabela 8.15, em que os acertos são mostrados na diagonal principal e os erros fora dessa diagonal.

Tabela 8.15: Matriz de confusão do classificador final da Dimensão Lacuna-Contexto

		Classificador		Total
Anotador Humano	Classe	Sim	Não	
	Sim	130	7	137
	Não	14	32	46
	Total	144	39	183

Pode-se observar na Tabela 8.15 que do total de 137 sentenças que realmente pertencem à classe “sim” (classificadas pelo anotador humano), 130 sentenças foram corretamente classificadas como “sim” e 7 sentenças foram incorretamente classificadas como “não” pelo classificador. Já do total de 46 sentenças que realmente pertencem à classe “não”

(classificadas pelo anotador humano), 32 sentenças foram corretamente classificadas como “não” e 14 sentenças foram incorretamente classificadas como “sim” pelo classificador. No total, foram classificadas 144 sentenças como “sim”, em que 14 sentenças pertencem à classe “não”, ou seja, incorretamente classificadas, e 39 sentenças classificadas como “não”, em que 7 pertencem à classe “sim” e, portanto, incorretamente classificadas.

Na Tabela 8.16 é apresentado o desempenho do classificador em relação às medidas de avaliação *Precision*, *Recall*, *F-Measure* e *Macro-F*. A Tabela 8.16 também apresenta os resultados de uma *baseline* simples que sempre atribui a classe prevalente da anotação como saída para um classificador hipotético. Como relatado anteriormente na Seção 7.4, a classe prevalente na anotação da Dimensão Lacuna-Contexto é a classe “sim” com 74,86% das sentenças.

Tabela 8.16: Desempenho do classificador final da Dimensão Lacuna-Contexto

	Sim			Não			Macro-F
	Precision	Recall	F-measure	Precision	Recall	F-measure	
Classificador	0,903	0,949	0,925	0,821	0,696	0,753	0,839
Baseline	0,748	1,000	0,855	0,000	0,000	0,000	0,428

Observa-se na Tabela 8.16 que o classificador supera os resultados da *baseline*. Assim como nas outras dimensões, nota-se que o valor da medida *F-measure* da classe “sim” é superior ao da classe “não”. Como observado no Capítulo 7, cada resumo possui em média uma sentença classificada como *Gap* e três sentenças classificadas como *Background*. Desse modo, a probabilidade de existir um alto relacionamento da sentença de *Gap* com uma das três sentenças de *Background* é maior do que a probabilidade de um baixo relacionamento. Entretanto, o classificador obteve um bom desempenho de modo geral com o valor 0,839 para a medida *Macro-F*, que leva em consideração ambas as classes.

Desse modo, os resultados positivos obtidos na avaliação do classificador possibilitam a sua utilização para desempenhar a tarefa proposta pela Dimensão Lacuna-Contexto.

8.7 Sugestões Apresentadas no Protótipo

Como discutido em capítulos anteriores, o auxílio fornecido pelo SciPo é dado por meio de críticas e sugestões que são apresentadas ao usuário em relação ao texto submetido a análise. A ferramenta considera que desvios graves são apresentados ao usuário como críticas, enquanto os desvios leves são apresentados como sugestões. Até então, somente aspectos relacionados a estrutura retórica identificada no texto eram abordados pela

ferramenta. Nesta seção, apresentamos as novas sugestões baseadas nos classificadores desenvolvidos no decorrer deste trabalho e que visam auxiliar o usuário em relação aos aspectos semânticos do texto, como a coerência. Estes aspectos são analisados em um protótipo do ambiente SciPo que permite ao usuário acatar as sugestões apresentadas e corrigir os problemas identificados para, então, novamente submeter o texto para análise, em um processo cíclico de melhoria. Ou ainda, ignorar as sugestões apresentadas, assim como ocorre no ambiente SciPo. Escolhemos por apresentar somente sugestões, uma vez que tratam-se de aspectos subjetivos no qual não podemos classificá-los como desvios graves e que podem ser ignorados pelo usuário, diferentemente das críticas que devem ser obrigatoriamente corrigidas.

Dessa maneira, de acordo com os classificadores desenvolvidos e analisados, apresentamos o conjunto de sugestões utilizadas no protótipo.

8.7.1 Dimensão Título

Sugestão: O Título e as sentenças de Propósito devem ser relacionadas.

— Espera-se que as sentenças de Propósito indiquem o principal objetivo da pesquisa. Da mesma maneira, o Título do resumo deve “condensar” esse objetivo de modo que seja compreendido pelo leitor em poucas palavras (entre uma ou duas linhas de texto). Verifique se o Título de seu resumo aborda de forma sucinta o objetivo principal citado nas sentenças de Propósito ou se as sentenças de Propósito remetem ao conteúdo abordado no Título, além de informações complementares. O não cumprimento destas condições poderá implicar em um problema de coerência no seu resumo. Considere a reescrita do Título e das sentenças de Propósito de seu resumo.

8.7.2 Dimensão Propósito

Sugestão: Metodologia e Propósito podem ser mais relacionados.

— As sentenças de Metodologia indicam os materiais e métodos utilizados ou que serviram de base para a pesquisa. É comum que os métodos sejam descritos ou, pelo menos, indicados. Em resumos com um nível maior de coerência, espera-se que a descrição ou indicação dos métodos utilizados sejam relacionadas com o objetivo principal do trabalho para que seu uso seja justificado. Desse modo, verifique em seu texto se o conteúdo das sentenças de Metodologia e Propósito está relacionado. A reescrita das sentenças de Metodologia pode ser necessária em seu resumo.

Sugestão: Resultado e Propósito podem ser mais relacionados.

— As sentenças de Resultado descrevem qualquer artefato desenvolvido pelo autor ou indicam os resultados de experimentos e avaliações. Em resumos com um nível maior de coerência, espera-se que a descrição dos resultados seja relacionada com o objetivo principal do trabalho, de modo que o leitor compreenda, por meio dos resultados, qual a relevância e a contribuição do trabalho de acordo com o objetivo exposto nas sentenças de Propósito. Desse modo, verifique em seu texto se o conteúdo das sentenças de Resultado e Propósito está relacionado. A reescrita das sentenças de Resultado pode ser necessária em seu resumo.

Sugestão: Conclusão e Propósito podem ser mais relacionados.

— As sentenças de Conclusão têm o papel de “encerrar” o texto, além de indicar recomendações, contribuições e expressar o valor do trabalho realizado. Para que esse papel seja exercido de forma coerente, espera-se que as sentenças de Conclusão remetam ao conteúdo abordado nas sentenças de Propósito, de modo que o leitor lembre o objetivo principal do trabalho. Desse modo, verifique em seu resumo se o conteúdo das sentenças de Conclusão e Propósito está relacionado. A reescrita das sentenças de Conclusão pode ser necessária em seu resumo.

8.7.3 Dimensão Lacuna-Contexto

Sugestão: Lacuna e Contexto devem ser relacionados.

— As sentenças de Lacuna indicam alguma questão de pesquisa que vale a pena ser investigada. Desse modo, espera-se que em um resumo coerente exista ao menos uma sentença responsável por contextualizar tal questão antes de sua apresentação propriamente dita. Verifique em seu resumo se o conteúdo das sentenças de Lacuna e Contexto está relacionado. A reescrita dessas sentenças pode ser necessária.

8.8 Avaliação do Protótipo

Para avaliar o protótipo em seu contexto de uso em conjunto com o ambiente SciPo, foi realizado um experimento com usuários reais com o objetivo de verificar o quanto as novas sugestões que tratam aspectos relativos à coerência são de fato efetivas no auxílio à escrita de resumos acadêmicos em português.

O experimento foi realizado com oito estudantes voluntários do curso de mestrado em Ciência da Computação da Universidade Estadual de Maringá. Dos oito estudantes participantes do experimento, quatro escreveram recentemente o Trabalho de Conclusão de Curso (TCC) exigido no curso de graduação da mesma instituição, enquanto os outros quatro escreveram suas dissertações de mestrado recentemente.

Foi solicitado individualmente aos participantes que utilizassem o protótipo para a reescrita do resumo de suas dissertações ou trabalhos de conclusão de curso, submetendo seus resumos já escritos à análise automática de coerência. Durante a execução do experimento, o autor do protótipo estava presente, mas só foram feitas intervenções quando requeridas pelo participante. Antes do experimento ser iniciado, foi apresentado os objetivos do protótipo e uma breve explicação a respeito dos componentes retóricos que compõem a estrutura esquemática de um resumo acadêmico. Após a reescrita do resumo, todos os participantes preencheram um questionário relatando suas impressões sobre o protótipo. O questionário utilizado nesse experimento é apresentado ao final do trabalho, no Apêndice B.

O protótipo utiliza um sistema de detecção automática da estrutura esquemática do resumo, proveniente do classificador AZPort. Antes de submeter o resumo à análise de coerência, foi permitido e solicitado aos estudantes que corrigissem o resultado da detecção automática, sendo que, de um total de oito estudantes, três não tinham nenhum conhecimento prévio sobre a estruturação de um resumo acadêmico em termos de componentes esperados. Durante a correção do resultado da detecção automática, observou-se a dificuldade de dois estudantes em distinguir as sentenças de Contexto e Metodologia e um terceiro estudante teve dificuldade em distinguir as sentenças de Resultado e Conclusão. Para esses estudantes, o autor do protótipo auxiliou a correção e notou ambiguidade entre os componentes. Os outros estudantes não relataram dificuldades na correção. A quantidade total de classificações realizadas pelo AZPort nesse experimento foi de 63 sentenças, em que 26 (41,3%) sentenças foram corrigidas pelos estudantes.

Para seis estudantes, o protótipo apresentou uma única sugestão em relação à coerência para cada um deles. Para três desses estudantes, a sugestão referiu-se ao baixo relacionamento entre o par de componentes Título-Propósito. Para outros dois estudantes, a sugestão referiu-se ao baixo relacionamento entre os componentes Propósito-Metodologia; e o último estudante recebeu uma sugestão referente ao baixo relacionamento entre o par Propósito-Resultado. Para outros dois participantes do experimento, o protótipo apresentou duas sugestões em relação à coerência para cada um. Para o primeiro estudante, as sugestões envolviam o relacionamento entre os componentes Título-Propósito e Propósito-Metodologia, enquanto para o segundo, as sugestões foram referentes aos com-

ponentes Propósito-Metodologia e Propósito-Resultado. Notou-se no experimento que o protótipo não apresentou nenhuma sugestão em relação aos componentes Lacuna-Contexto e Propósito-Conclusão. Isso se deve não só ao fato dos resumos analisados não apresentarem problemas em relação a esses componentes, mas também pela ausência de sentenças dos tipos Lacuna e Conclusão em boa parte dos resumos, ainda que solicitados previamente aos estudantes e explicitado a importância de cada componente em um resumo acadêmico.

Todos os estudantes acataram as sugestões emitidas pelo protótipo, sendo que seis acataram integralmente e dois acataram parcialmente, de maneira que mesmo após a reescrita e adequação do resumo, o protótipo ainda apresentava alguma sugestão.

Em relação a relevância das sugestões apresentadas, cinco estudantes consideraram como relevantes, dois estudantes consideraram como muito relevantes e um estudante considerou a sugestão apresentada como pouco relevante.

Ao comparar o resumo reescrito após a adequação de acordo com as sugestões e a versão inicial do mesmo resumo, quatro estudantes consideraram que o nível de coerência do resumo final aumentou razoavelmente, enquanto três estudantes consideraram um bom aumento no nível de coerência e um estudante considerou que as alterações no resumo não alteraram o seu nível de coerência.

A reescrita e adequação do resumo de acordo com as sugestões apresentadas pelo protótipo gerou algumas dúvidas para três estudantes e muitas dúvidas para um estudante. Por outro lado, quatro estudantes relataram que não tiveram dúvidas para reescrever o resumo ao acatar as sugestões. Acredita-se que as dificuldades relatadas são inerentes ao processo de escrita, uma vez que a maioria dos estudantes (6) julgaram as informações das sugestões como muito informativas, enquanto dois estudantes julgaram como razoavelmente informativas e, mesmo aqueles com dificuldades em reescrever partes do resumo, não consideraram as informações das sugestões como pouco informativas.

Parte da análise realizada e discutida anteriormente é apresentada graficamente na Figura 8.2. Nesta figura são apresentados gráficos referentes as opiniões dos participantes do experimento em relação a relevância das sugestões emitidas pelo protótipo, o nível de coerência após a adequação das sugestões e as dificuldades encontradas quanto a tarefa de adequação das sugestões.

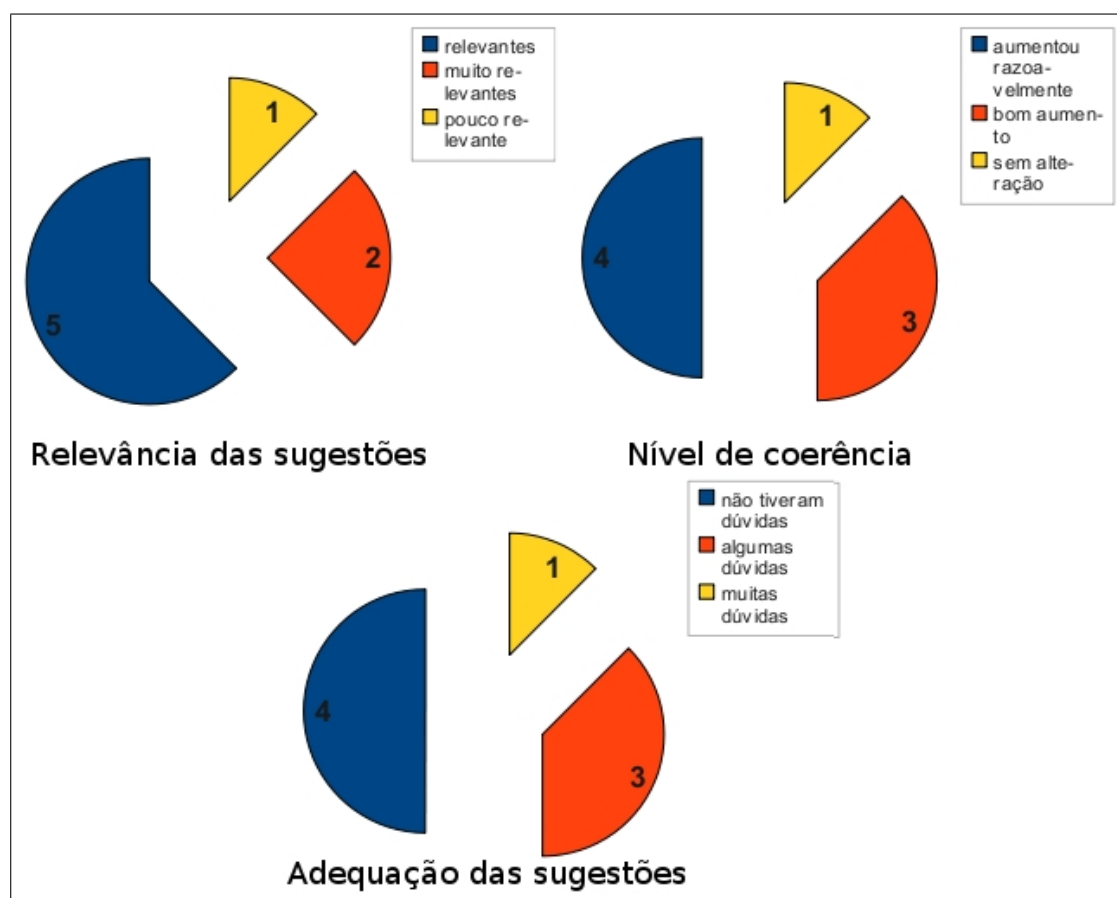


Figura 8.2: Gráficos que correspondem às respostas referentes às opiniões sobre a relevância das sugestões, nível de coerência após a reescrita e dificuldade de adequação das sugestões

Com o objetivo de apresentar o resultado final de um resumo que passou pelo processo de reescrita de acordo com as sugestões apresentadas pelo protótipo em relação à coerência e compará-lo à sua versão inicial, foi solicitado e autorizado por um dos participantes do experimento o uso de seu resumo para demonstração. Na Tabela 8.17 é apresentada a versão inicial do resumo selecionado, composto por sete sentenças, no qual considerou-se a seguinte classificação: 1) Propósito; 2) Propósito; 3) Metodologia; 4) Metodologia; 5) Metodologia; 6) Resultado e 7) Conclusão.

Tabela 8.17: Versão inicial de um resumo acadêmico

Título
Classificação de Papéis Retóricos Baseada em Expressões Automaticamente Extraídas de Textos Científicos
Resumo
1) Este trabalho apresenta a JBoW, uma ferramenta para a geração automática de listas de expressões padrão, ou seja, expressões que caracterizam o papel retórico de uma sentença dentro de um texto. 2) Em especial, essa ferramenta visa auxiliar na geração do conhecimento necessário para a extração do atributo Expressão do classificador AZPort. 3) A JBoW utiliza ferramentas de pré-processamento de texto e implementa o cálculo de medidas estatísticas que estimam a relevância das expressões geradas. 4) Para a extração das listas de expressões são empregados dois modelos: um Bag of Clusters original e um Bag of Clusters modificado. 5) Para testar a eficiência das listas geradas, foram realizados testes com um classificador bayesiano semelhante ao AZPort. 6) Os resultados obtidos indicam que a automatização da geração das listas utilizadas para a extração do atributo Expressão é viável. 7) Espera-se que este trabalho possa auxiliar na adaptação do AZPort para diferentes domínios, minimizando o esforço manual para a geração do conhecimento utilizado pelo classificador.

Foram emitidas pelo protótipo duas sugestões em relação ao resumo inicial:

- Sugestão 1: O Título e as sentenças de Propósito devem ser relacionadas;
- Sugestão 2: Resultado e Propósito podem ser mais relacionados.

Dadas as sugestões apresentadas, o estudante considerou a reescrita tanto do Título quanto da sentença 1) Propósito, para que as mesmas possuam um relacionamento maior. Posteriormente, o estudante considerou a reescrita da sentença 6) Resultado, com o objetivo de relacioná-la com o componente Propósito. Após essas mudanças na versão inicial do resumo, o estudante realizou uma nova submissão para análise na qual o protótipo emitiu uma nova sugestão:

- Sugestão: Metodologia e Propósito podem ser mais relacionados.

Dada a sugestão, o estudante considerou a reescrita das sentenças 4) Metodologia e 5) Metodologia, com o objetivo de relacioná-las com o componente Propósito. Após tais alterações e uma nova submissão, o protótipo não emitiu mais sugestões. A versão final do resumo é apresentada na Tabela 8.18.

Tabela 8.18: Versão final de um resumo acadêmico após as sugestões de coerência

Título
JBoW: Uma Ferramenta para Classificação de Papéis Retóricos Baseada em Expressões Automaticamente Extraídas de Textos Científicos
Resumo
1) Este trabalho apresenta a JBoW, uma ferramenta para a geração automática de listas de expressões padrão, utilizadas na classificação de sentenças de textos científicos em papéis retóricos. 2) Em especial, essa ferramenta visa auxiliar na geração do conhecimento necessário para a extração do atributo Expressão do classificador AZPort. 3) A JBoW utiliza ferramentas de pré-processamento de texto e implementa o cálculo de medidas estatísticas que estimam a relevância das expressões geradas. 4) Para a geração automática das listas de expressões a serem utilizadas pelo classificador AZPort, são empregados dois modelos: um Bag of Clusters original e um Bag of Clusters modificado. 5) Para testar a eficiência das listas de expressões padrão geradas, foram realizados testes com um classificador bayesiano semelhante ao do classificador AZPort. 6) Os resultados obtidos indicam que é viável a automatização da geração das listas utilizadas para a classificação de sentenças de textos científicos em papéis retóricos. 7) Espera-se que este trabalho possa auxiliar na adaptação do AZPort para diferentes domínios, minimizando o esforço manual para a geração do conhecimento utilizado pelo classificador.

Nota-se nas Tabelas 8.17 e 8.18 que a reescrita de determinadas sentenças pode ser sutil para que os componentes sejam mais relacionados e, conseqüentemente, o resumo mais coerente. Entretanto, em determinados casos a reescrita total das sentenças pode ser necessária, bem como a remoção ou inserção de novas sentenças.

De modo geral, o experimento mostrou que o protótipo é capaz de apresentar sugestões relevantes e que guiam o estudante a escrever um resumo acadêmico com um nível maior de coerência. Como o auxílio é fornecido por meio de sugestões que podem ser rejeitadas, o protótipo não garante que o texto final seja realmente mais coerente, mas indica ao estudante interessado em melhorar o seu resumo determinadas partes importantes do texto que podem ser reescritas.

Por fim, apresentamos na Figura 8.3 a captura de tela do protótipo desenvolvido. É possível observar a sugestão apresentada ao usuário que submeteu um resumo com problemas de coerência em relação ao componente Propósito e o Título do trabalho.

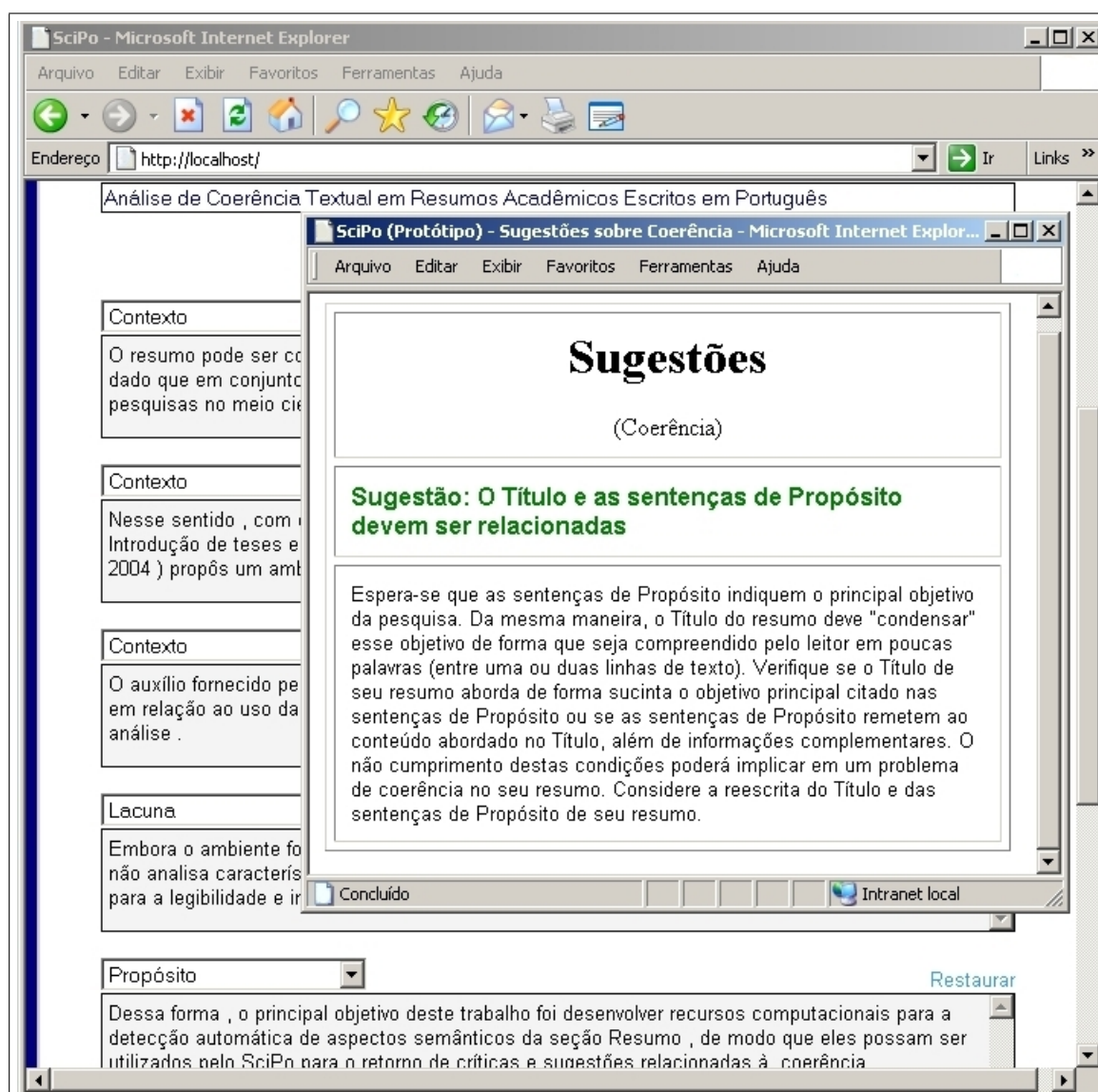


Figura 8.3: Captura de tela do protótipo desenvolvido

Conclusões e Trabalhos Futuros

Este trabalho teve como objetivo estudar, analisar, avaliar e implementar métodos para a detecção automática de problemas em relação a coerência semântica em resumos acadêmicos escritos em português a partir da proposta de quatro diferentes dimensões. Em específico, nos propusemos a desenvolver o protótipo de um módulo complementar ao ambiente de auxílio à escrita SciPo, para que o mesmo seja capaz de apresentar sugestões em relação a aspectos semânticos da seção resumo.

Com o objetivo de desenvolver o protótipo proposto, primeiramente estudamos os conceitos teóricos ligados à coerência textual, além de teorias e métodos empregados na análise da coerência, tanto de base linguística quanto estatística, e que têm sido utilizados em diferentes sistemas computacionais na análise de forma automática. Também estudamos as ferramentas de avaliação automática de textos que possuem algum critério de análise relacionado à coerência, embora boa parte dos sistemas seja comercial e, por isso, de difícil acesso. Além disso, também apresentamos o ambiente de auxílio à escrita SciPo com o qual este trabalho está diretamente relacionado. Por meio da abordagem de estrutura retórica utilizada no SciPo e do estudo de métodos e outras ferramentas que analisam a coerência, definimos as características desejáveis em nosso protótipo.

O protótipo desenvolvido é resultado da indução e avaliação de diversos classificadores com algoritmos de aprendizagem de máquina aplicados às dimensões propostas. Inicialmente foram propostas quatro dimensões: Dimensão Título, Dimensão Propósito, Dimensão Lacuna-Contexto e Dimensão Quebra de Linearidade. Porém, não foi induzido um classificador para a Dimensão Quebra de Linearidade devido ao baixo número de

exemplos identificados ainda no processo de anotação. Na indução dos classificadores foi utilizado o algoritmo SMO – *Sequential Minimal Optimization*. Tal algoritmo é uma implementação da técnica SVM – *Support Vector Machine* que tem sido utilizada com sucesso em aplicações de reconhecimento de padrões tais como categorização de textos (Aizawa, 2001) e categorização de *spam* (Drucker et al., 1999). Também foi utilizada por Higgins et al. (2004) na tarefa de análise de coerência.

Para o classificador da Dimensão Título, em especial na detecção do relacionamento semântico entre o título do resumo e o componente retórico *Purpose*, obteve-se o valor *Kappa* de 0.871 e os valores 0.979 para “alto” e 0.892 para “baixo” relacionamento, em termos da medida *F-measure*. Esses valores demonstram um nível de concordância elevada entre o classificador e o anotador humano em relação a tarefa desempenhada na Dimensão Título, que permitiu o uso do classificador no protótipo proposto para a identificação de problemas de coerência entre o título e o propósito do resumo.

Para desempenhar as tarefas propostas na Dimensão Propósito foram construídos três classificadores. O primeiro classificador é responsável pela detecção do relacionamento semântico das sentenças classificadas como *Methodology* e o propósito do resumo. Tal classificador obteve o valor *Kappa* de 0,683 e os valores 0,898 para “alto” e 0,785 para “baixo” relacionamento, em termos da medida *F-measure*. O segundo classificador é responsável pela detecção do relacionamento semântico das sentenças classificadas como *Result* e o propósito do resumo. Tal classificador obteve o valor *Kappa* de 0,763 e os valores 0,921 para “alto” e 0,842 para “baixo” relacionamento, em termos da medida *F-measure*. Por fim, o terceiro classificador é responsável pela detecção do relacionamento semântico das sentenças classificadas como *Conclusion* e o propósito do resumo e obteve o valor *Kappa* de 0,748 e os valores 0,933 para “alto” e 0,815 para “baixo” relacionamento, em termos da medida *F-measure*. Esses valores demonstram um bom nível de concordância entre os classificadores e o anotador humano, que permitiu o uso dos classificadores na geração de sugestões sobre coerência entre as sentenças classificadas como *Methodology* e *Purpose*, *Result* e *Purpose*, e também *Conclusion* e *Purpose*.

Para o classificador Dimensão Lacuna-Contexto, ao detectar o relacionamento entre os componentes retóricos *Gap* e *Background*, obteve-se o valor de concordância *Kappa* $K = 0.679$, com o anotador humano. Além dos valores 0.925 e 0.753 para a medida *F-measure* em relação as classes “sim” e “não”, respectivamente. Esses valores demonstram um bom nível de concordância entre o classificador e a anotação humana e que permitiu o uso do classificador proposto na geração de sugestões sobre a coerência entre as sentenças classificadas na categoria *Gap* e *Background*.

Por outro lado, a Dimensão Quebra de Linearidade apresentou um alto nível de ambiguidade ainda no processo de anotação, no qual o anotador humano teve dificuldades em selecionar casos em que fosse identificada a quebra de sentido lógico entre sentenças adjacentes. Os casos identificados correspondem a apenas 7,14% do total de sentenças analisadas. Essa dificuldade e o baixo número de exemplos refletiriam diretamente no desempenho do classificador. Percebeu-se que no corpus analisado, apenas em raras ocasiões se encontra uma sentença que não esteja relacionada com a sentença anterior ou posterior e, ainda assim, em alguns casos, tal sentença é justificada no decorrer do texto, sem que seja por uma sentença adjacente. Por outro lado, acreditamos que embora essa dimensão não possa ser utilizada como um indicativo nos problemas relacionados à coerência em resumos acadêmicos, pode ser válida se aplicada na análise de corpus com um nível maior de problemas, como textos gerados automaticamente por sumarizadores.

Devido as dificuldades encontradas para a construção do classificador da Dimensão Quebra de Linearidade, pretende-se, como extensão deste trabalho, usar o Modelo de Entidades proposto por Barzilay e Lapata (2008), para a análise do corpus e extração de novos atributos que possam colaborar para a construção de um novo classificador com o papel de identificar a existência de quebra de sentido lógico em sentenças adjacentes.

Durante as etapas de treinamento e teste dos classificadores foi realizada a etapa de seleção de atributos com o objetivo de melhorar o desempenho dos classificadores. Porém, acredita-se que o estudo aprofundado sobre o algoritmo SMO a fim de otimizar o ajuste de seus parâmetros obtenha resultados mais promissores. O estudo de algoritmos de diferentes abordagens também deve ser considerado.

Também deve-se considerar o desbalanceamento das classes no desempenho dos classificadores. Desse modo, se vê necessário em trabalhos futuros a aplicação e análise detalhada do impacto de técnicas para o balanceamento artificial das classes, tais como *Undersampling* para eliminação de exemplos da classe majoritária (Kubat e Matwin, 1997) ou *Oversampling* para replicação de exemplos da classe minoritária (Chawla et al., 2002). Ou ainda, a anotação de novos exemplos que contribuam para o balanceamento das classes.

Outra questão a ser abordada em trabalhos futuros é a adequação das dimensões propostas para outras seções de trabalhos acadêmicos, tais como Introdução ou Conclusão. Além de apresentarem uma estrutura retórica diferente do resumo, essas seções apresentam textos mais longos e um número maior de variações, tanto em relação à estrutura quanto ao conteúdo.

Por fim, pode-se concluir que tanto as avaliações intrínsecas dos classificadores quanto a avaliação do protótipo em um contexto de uso com usuários reais, demonstraram o potencial das dimensões propostas. O experimento mostrou que o protótipo é capaz de apresentar sugestões relevantes e que guiam o estudante a escrever um resumo acadêmico com um nível maior de coerência. Além disso, como o auxílio é fornecido por meio de sugestões que podem ser rejeitadas, o estudante pode ignorar qualquer erro dos classificadores.

Acreditamos que este trabalho possa beneficiar não só o ambiente de auxílio à escrita SciPo, mas também outras ferramentas de Processamento de Linguagem Natural, como sumarizadores automáticos.

Referências

AIZAWA, A. Linguistic techniques to improve the performance of automatic text categorization. In: *Proceedings of NLPRS-01, 6th Natural Language Processing Pacific Rim Symposium*, Tokyo, Japan, 2001, p. 307–314.

ALMEIDA, D.; ALUÍSIO, S. Text readability analysis with natural language processing tools: The adaptation of coh-metrix metrics for portuguese. In: *Proceedings of Seventh Brazilian Symposium in Information and Human Language Technology (STIL)*, 2009, p. 53–62.

ALUÍSIO, S.; OLIVEIRA JR, O. A detailed schematic structure of research paper introductions: An application in support-writing tools. *Procesamiento del Lenguaje Natural*, v. 19, p. 141–147, 1996.

ALUÍSIO, S.; SCHUSTER, E.; FELTRIM, V. D.; PESSOA, A.; OLIVEIRA, O. Evaluating scientific abstracts with a genre-specific rubric. In: *Proceeding of the 2005 Conference on Artificial Intelligence in Education*, Amsterdam, The Netherlands: IOS Press, 2005, p. 738–740.

ALUÍSIO, S. M.; BARCELOS, I.; SAMPAIO, J.; OLIVEIRA JR., O. N. How to learn the many unwritten "rules of the game" of the academic discourse: A hybrid approach based on critiques and cases. In: *Proceedings of the IEEE International Conference on Advanced Learning Technologies*, 2001, p. 257–260.

ANTHONY, L.; LASHKIA, G. Mover: A machine learning tool to assist in the reading and writing of technical papers. *Professional Communication, IEEE Transactions on*, v. 46, n. 3, p. 185–193, 2003.

ATTALI, Y.; BURSTEIN, J. Automated essay scoring with e-rater v. 2. *Journal of Technology, Learning, and Assessment*, v. 4, n. 3, p. 1–31, 2006.

BARZILAY, R.; LAPATA, M. Modeling local coherence: An entity-based approach. *Computational Linguistics*, v. 34, n. 1, p. 1–34, 2008.

DE BEAUGRANDE, R.; DRESSLER, W. *Introduction to text linguistics*. New York, NY: Longman Publisher Group, 288 p., 1981.

BROADY, E.; SHURVILLE, S. Developing academic writer: Designing a writing environment for novice academic writers. *Second Language Writing in a Computer Environment, CILT, London*, p. 131–151, 2000.

BURSTEIN, J.; CHODOROW, M.; LEACOCK, C. Criterionsm: Online essay evaluation: An application for automated evaluation of student essays. In: *Proceedings of the fifteenth annual conference on innovative applications of artificial intelligence*, Acapulco, Mexico, 2003a, p. 3–10.

BURSTEIN, J.; CHODOROW, M.; LEACOCK, C. Automated essay evaluation: The criterion online writing service. *AI Magazine*, v. 25, n. 3, p. 27–36, 2004.

BURSTEIN, J.; MARCU, D.; KNIGHT, K.; TECHNOL, E.; PRINCETON, N. Finding the write stuff: Automatic identification of discourse structure in student essays. *IEEE Intelligent Systems*, v. 18, n. 1, p. 32–39, 2003b.

BURSTEIN, J.; TETREAULT, J.; ANDREYEV, S. Using entity-based features to model coherence in student essays. In: *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, Stroudsburg, PA, USA: Association for Computational Linguistics, 2010, p. 681–684.

CHAROLLES, M. Introdução aos problemas da coerência dos textos: abordagem teórica e estudo das práticas pedagógicas. In: *GALVES, C.; ORLANDI, EP; OTONI, P. (Orgs.) O texto: leitura e escrita*, Campinas: Pontes, 1978, p. 39–90.

CHAWLA, N.; BOWYER, K.; HALL, L.; KEGELMEYER, W. Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, v. 16, p. 321–357, 2002.

DEERWESTER, S.; DUMAIS, S.; FURNAS, G.; LANDAUER, T.; HARSHMAN, R. Indexing by latent semantic analysis. *Journal of the American society for information science*, v. 41, n. 6, p. 391–407, 1990.

VAN DIJK, T. A. Studies in the pragmatics of discourse. *American Anthropologist*, v. 85, n. 1, p. 190–192, 1983.

VAN DIJK, T. A.; KINTSCH, W. *Strategies of discourse comprehension*. New York: Academic Press, 418 p., 1983.

DIKLI, S. An overview of automated scoring of essays. *The Journal of Technology, Learning, and Assessment*, v. 5(1), p. 3–35, 2006.

DRUCKER, H.; WU, D.; VAPNIK, V. N. Support vector machines for spam categorization. *IEEE TRANSACTIONS ON NEURAL NETWORKS*, v. 10, n. 5, p. 1048–1054, 1999.

ELLIOT, S. Intellimetric: From here to validity. In: SHERMIS, M.; BURSTEIN, J., eds. *Automatic Essay Scoring: A Cross-Disciplinary Perspective*, Hillsdale, NJ: Lawrence Erlbaum Associates, 2003, p. 71–86.

ELLIOT, S.; MIKULAS, C. The impact of my access! use on student writing performance: A technology overview and four studies from across the nation. In: *Annual Meeting of the National Council on Measurement in Education*, San Diego, CA, 2004.

FALLSIDE, D. C. *Xml schema part 0: Primer*. W3C, <http://www.w3.org/TR/2001/REC-xmlschema-0-20010502/>, 2001.

FELTRIM, V. D. *Uma abordagem baseada em corpus e em sistemas de crítica para a construção de ambientes web de auxílio à escrita acadêmica em português*. Tese de Doutorado, Instituto de Computação e Matemática Computacional - Universidade de São Paulo, São Carlos, SP, 2004.

FELTRIM, V. D.; ALUÍSIO, S. M.; NUNES, M. D. G. V. Analysis of the rhetorical structure of computer science abstracts in portuguese. In: ARCHER, D.; RAYSON, P.; WILSON, A.; MCENERY, T., eds. *Proceedings of Corpus Linguistics 2003*, 2003, p. 212–218 (*UCREL Technical Papers*, v.16, part 1, special issue).

FELTRIM, V. D.; TEUFEL, S.; NUNES, M.; ALUÍSIO, S. Argumentative zoning applied to critiquing novices scientific abstracts. In: SHANAHAN, J.; QU, Y.; WIEBE, J., eds. *Computing Attitude and Affect in Text: Theory and Applications*, Dordrecht, Netherlands: Springer, 2006, p. 233–246 (*The Information Retrieval Series*, v.20).

FRANZKE, M.; KINTSCH, E.; CACCAMISE, D.; JOHNSON, N.; DOOLEY, S. Summary street: Computer support for comprehension and writing. *Journal of Educational Computing Research*, v. 33, n. 1, p. 53–80, 2005.

- GIVÓN, T. Beyond foreground and background. In: TOMLIN, R. S., ed. *Coherence and grounding in discourse: outcome of a symposium, Eugene, Oregon, June 1984*, Amsterdam: John Benjamins Publishing Company, 1987, p. 88–175.
- GOLUB, G.; REINSCH, C. Singular value decomposition and least squares solutions. *Numerische Mathematik*, v. 14, n. 5, p. 403–420, 1970.
- GRAESSER, A.; MCNAMARA, D.; LOUWERSE, M.; CAI, Z. Coh-metrix: Analysis of text on cohesion and language. *Behavior Research Methods, Instruments, & Computers*, v. 36, n. 2, p. 193–202, 2004.
- GROSZ, B. J.; WEINSTEIN, S.; JOSHI, A. K. Centering: a framework for modeling the local coherence of discourse. *Computational Linguistics*, v. 21, n. 2, p. 203–225, 1995.
- HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. The weka data mining software: an update. *ACM SIGKDD Explorations Newsletter*, v. 11, n. 1, p. 10–18, 2009.
- HALL, M. A.; HOLMES, G. Benchmarking attribute selection techniques for discrete class data mining. *IEEE Transactions on Knowledge and Data Engineering*, v. 15, p. 1437–1447, 2003.
- HALLIDAY, M.; HASAN, R. *Cohesion in english*. Longman London, 392 p., 1976.
- HASLER, L. An investigation into the use of centering transitions for summarisation. In: *7th Annual CLUK Research Colloquium*, Birmingham, UK, 2004, p. 100–107.
- HASLER, L. Centering theory for evaluation of coherence in computer-aided summaries. In: *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco: European Language Resources Association (ELRA), 2008, p. 2698–2705.
- HIGGINS, D.; BURSTEIN, J. Sentence similarity measures for essay coherence. In: *Proceedings of the 7th International Workshop on Computational Semantics (IWCS-7)*, Tilburg, The Netherlands, 2007, p. 1–12.
- HIGGINS, D.; BURSTEIN, J.; MARCU, D.; GENTILE, C. Evaluating multiple aspects of coherence in student essays. In: *Human Language Technologies: The 2004 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, Association for Computational Linguistics, 2004.

- JAIN, A.; DUIN, R.; MAO, J. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 22, n. 1, p. 4–37, 2000.
- JORDÃO, R.; BELLEZI, C. *Linguagens*. São Paulo: Escala Educacional, 566 p., 2007.
- KANERVA, P.; KRISTOFERSSON, J.; HOLST, A. Random indexing of text samples for latent semantic analysis. In: *Proceedings of the 22nd Annual Conference of the Cognitive Science Society (CogSci'00)*, Philadelphia, Pennsylvania, USA: Erlbaum, 2000, p. 1036–1037.
- KARAMANIS, N. *Entity coherence for descriptive text structuring*. Tese de Doutorado, Institute for Communicating and Collaborative Systems, School of Informatics - University of Edinburgh, Edinburgh, Scotland, 2003.
- KARAMANIS, N.; POESIO, M.; MELLISH, C.; OBERLANDER, J. Evaluating centering-based metrics of coherence for text structuring using a reliably annotated corpus. In: *ACL '04: Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, Morristown, NJ, USA: Association for Computational Linguistics, 2004, p. 391–398.
- KEERTHI, S.; SHEVADE, S.; BHATTACHARYYA, C.; MURTHY, K. Improvements to platt's smo algorithm for svm classifier design. *Neural Computation*, v. 13, n. 3, p. 637–649, 2001.
- KINTSCH, W. The potential of latent semantic analysis for machine grading of clinical case summaries. *Journal of biomedical informatics*, v. 35, n. 1, p. 3–7, 2002.
- KOCH, I. V.; TRAVAGLIA, L. C. *Texto e coerência*, v. 4. São Paulo: Cortez Editora, 1989.
- KOCH, I. V.; TRAVAGLIA, L. C. *A coerência textual*. São Paulo: Editora Contexto, 2003.
- KUBAT, M.; MATWIN, S. Addressing the curse of imbalanced training sets: One-sided selection. In: *Machine learning: Proceedings of the Fourteenth International Conference (ICML'97)*, Nashville, Tennessee: Morgan Kaufmann Pub, 1997, p. 179–186.
- KUKICH, K. Beyond automated essay scoring. *IEEE Intelligent Systems*, v. 15, n. 5, p. 22–27, 2000.

- LANDAUER, T.; FOLTZ, P.; LAHAM, D. An introduction to latent semantic analysis. *Discourse processes*, v. 25, p. 259–284, 1998.
- LANDAUER, T.; LAHAM, D.; FOLTZ, P. Automated scoring and annotation of essays with the intelligent essay assessor. *Automated essay scoring: A cross-disciplinary perspective*, p. 87–112, 2003a.
- LANDAUER, T. K.; LAHAM, D.; FOLTZ, P. W. *Automated essay scoring and annotation of essays with the intelligent essay assessor*. Mahwah, NJ: Lawrence Erlbaum Associates, 87–112 p., 2003b.
- LANDIS, J.; KOCH, G. The measurement of observer agreement for categorical data. *Biometrics*, v. 33, n. 1, p. 159, 1977.
- LANSMAN, M. Using the "writing environment" to study writer. *Computers and Composition*, v. 10, n. 2, p. 71–92, 1993.
- LJUNGSTRAND, P.; JOHANSSON, H. *Intranet indexing using semantic document clustering*. Dissertação de Mestrado, Department of Informatics - Göteborg University, Göteborg, Sweden, 1998.
- MANI, I. *Automatic summarization*. Natural Language Processing. John Benjamins Publishing Company, 285 p., 2001.
- MANN, W. C.; THOMPSON, S. A. *Rhetorical structure theory: A theory of text organization*. Relatório Técnico ISI/RS-87-190, Information Sciences Institute, 1987.
- McNAMARA, D.; LOUWERSE, M.; MCCARTHY, P.; GRAESSER, A. Coh-metrix: Capturing linguistic features of cohesion. *Discourse Processes*, v. 47, n. 4, p. 292–330, 2010.
- McNAMARA, D.; OZURU, Y.; GRAESSER, A.; LOUWERSE, M. Validating coh-metrix. In: *Proceedings of the 28th annual conference of the cognitive science society*, 2006, p. 573–578.
- MILLER, T. Essay assessment with latent semantic analysis. *Journal of Educational Computing Research*, v. 29, n. 4, p. 495–512, 2003.
- MILTSAKAKI, E.; KUKICH, K. Automated evaluation of coherence in student essays. In: *Proceedings of the Workshop on Language Resources and Tools in Educational Applications (LREC 2000)*, Athens, Greece, 2000a, p. 7–14.

- MILTSAKAKI, E.; KUKICH, K. The role of centering theory's rough-shift in the teaching and evaluation of writing skills. In: *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics*, 2000b, p. 408–415.
- MILTSAKAKI, E.; KUKICH, K. Evaluation of text coherence for electronic essay scoring systems. *Natural Language Engineering*, v. 10, n. 1, p. 25–55, 2004.
- MITCHELL, T. *Machine learning*. McGraw-Hill Science/Engineering/Math, 432 p., 1997.
- NARITA, M. Constructing a tagged ej parallel corpus for assisting japanese software engineers in writing english abstracts. In: *LREC 2000. Proceedings of the Second International Conference on Language Resources and Evaluation*, 2000a, p. 1187–1191.
- NARITA, M. Corpus-based english language assistant to japanese software engineers. *Proceedings of MT-2000 Machine Translation and Multilingual Applications in the New Millennium*, p. 24–1, 2000b.
- ORĂSAN, C. PALinkA: a highly customizable tool for discourse annotation. In: *Proceedings of the 4th SIGdial Workshop on Discourse and Dialog*, Sapporo, Japan, 2003, p. 39–43.
- PARDO, T. *Métodos para análise discursiva automática*. Tese de Doutorado, Instituto de Computação e Matemática Computacional - Universidade de São Paulo, São Carlos - SP, 2005.
- PEMBERTON, L.; SHURVILLE, S.; HARTLEY, T. Motivating the design of a computer assisted environment for writers in a second language. *Computer Aided Learning and Instruction in Science and Engineering*, p. 141–148, 1996.
- PRINCE, E. Toward a taxonomy of given-new information. *Radical pragmatics*, p. 223–255, 1981.
- REZENDE, S. *Sistemas inteligentes: fundamentos e aplicações*. São Paulo: Editora Manole, 550 p., 2003.
- RUDNER, L. M.; GARCIA, V.; WELCH, C. An evaluation of intellimetric essay scoring system. *The Journal of Technology, Learning, and Assessment*, v. 4(4), p. 1–21, 2006.
- SAHLGREN, M. Vector-based semantic analysis: Representing word meanings based on random labels. In: *Proceedings of the ESSLLI 2001 Workshop on Semantic Knowledge Acquisition and Categorisation*, Helsinki, Finland, 2002, p. 1–21.

- SAHLGREN, M. An introduction to random indexing. In: *Methods and Applications of Semantic Indexing Workshop at the 7th International Conference on Terminology and Knowledge Engineering (TKE 2005)*, Copenhagen, Denmark, 2005, p. 1–9.
- SHARPLES, M.; GOODLET, J.; CLUTTERBUCK, A. A comparison of algorithms for hypertext notes network linearization. *International Journal of Human-Computer Studies*, v. 40, n. 4, p. 727–752, 1994.
- SHARPLES, M.; PEMBERTON, L. Representing writing: External representations and the writing process. *Holt and Williams*, p. 319–336, 1992.
- SHURVILLE, S.; PEMBERTON, L.; HARTLEY, A. A development methodology for composer: a computer support tool for academic writing in a second language. *The Production of Scientific Texts in the Age of the Computer*, Lang, Frankfurt, p. 171–82, 1997.
- SOUZA, V. M. A.; FELTRIM, V. D. An analysis of textual coherence in academic abstracts written in portuguese (to be published). In: *Proceedings of the Sixth Corpus Linguistics Conference: CL 2011*, Birmingham, UK, 2011.
- SWALES, J. M. *Genre analysis: English in academic and research settings*. Cambridge applied linguistics series. Cambridge, UK: Cambridge University Press, 272 p., 1990.
- TABOADA, M.; MANN, W. C. Rhetorical structure theory: looking back and moving ahead. *Discourse Studies*, v. 8, n. 3, p. 423–459, 2006.
- TEUFEL, S.; MOENS, M. Summarising scientific articles — experiments with relevance and rhetorical status. *Computational Linguistics*, v. 28, n. 4, p. 409–446, 2002.
- VAPNIK, V. *The nature of statistical learning theory*. New York, NY, USA: Springer Verlag, 314 p., 2000.
- VIERA, A.; VIRGIL, J. Uma revisão dos algoritmos de radicalização em língua portuguesa. *Information Research*, v. 12, n. 3, p. 12–3, 2007.
- ŘEHŮŘEK, R.; SOJKA, P. Software framework for topic modelling with large corpora. In: *Proceedings of LREC 2010 workshop New Challenges for NLP Frameworks*, 2010, p. 45–50.
- WADE-STEIN, D.; KINTSCH, E. Summary street: Interactive computer support for writing. *Cognition and Instruction*, v. 22, n. 3, p. 333–362, 2004.

WEISSBERG, R.; BUKER, S. *Writing up research: Experimental research report writing for students of english.* Prentice Hall Regents, 208 p., 1990.

WIDDOWSON, H. *Teaching language as communication.* Oxford Applied Linguistics. Oxford University Press, 168 p., 1984.

WITTEN, I.; FRANK, E. *Data mining: Practical machine learning tools and technique.* Morgan Kaufmann Publisher, 560 p., 2005.

Listas de Palavras Utilizadas nos Atributos dos Classificadores

Tabela A.1: Lista de *stopwords*

a	desta	feita	mesmos	poder	tampouco
agora	destas	feitas	meu	poderia	te
ainda	deste	feito	meus	poderiam	tem
alem	destes	feitos	minha	podia	temos
alguem	deve	fez	minhas	podiam	tendo
algum	devem	foi	muita	pois	tenha
alguma	devendo	for	muitas	por	ter
algumas	dever	foram	muito	porem	terei
alguns	devera	fosse	muitos	porque	teu
ampla	deverao	fossem	na	posso	teus
amplas	deveria	grande	nas	pouca	teve
amplo	deveriam	grandes	nao	poucas	ti
amplos	devia	ha	nem	pouco	tido
ante	deviam	hei	nenhum	poucos	tinha
antes	disse	havemos	nessa	primeiro	tinham
ao	disso	hao	nessas	primeiros	tive
aos	disto	houve	nesta	propria	tivera
apos	dito	houvemos	nestas	proprias	tiveram
aquela	diz	houveram	ninguem	proprio	tivessemos
aquelas	dizem	houvera	no	proprios	tivessem

aquele	do	houveramos	nos	quais	tiver
aqueles	dos	haja	nos	qual	tivermos
aquilo	e	hajamos	nossa	quando	tiverem
as	ela	hajam	nossas	quanto	toda
ate	elas	houvesse	nosso	quantos	todas
atraves	ele	houvessemos	nossos	que	todavia
cada	eles	houvessem	num	quem	todo
coisa	em	houver	numa	sao	todos
coisas	enquanto	houvermos	nunca	se	tu
com	entre	houverem	o	seja	tua
como	era	houverei	os	sejam	tuas
contra	essa	houvera	ou	sem	tudo
contudo	essas	houveremos	outra	sempre	ultima
da	esse	houverao	outras	sendo	ultimas
dar	esses	houveria	outro	ser	ultimo
daquele	esta	houveriamos	outros	sera	ultimos
daqueles	estamos	houveriam	para	serao	um
das	estao	isso	pela	seu	uma
de	estas	isto	pelas	seus	umas
dela	estava	ja	pelo	si	uns
delas	estavam	la	pelos	sido	vendo
dele	estavamos	lhe	pequena	so	ver
deles	este	lhes	pequenas	sob	vez
dentre	estes	lo	pequeno	sobre	vindo
depois	estou	mas	pequenos	sua	vir
dessa	eu	me	per	suas	vos
dessas	faz	mesma	perante	tal	

Tabela A.2: Lista de possíveis elementos anafóricos

dele	deles	dela	delas	desta	destas
daquele	daqueles	daquela	daquelas	deste	destes
dessa	dessas	desse	desses	ele	eles
ela	elas	este	estes	esta	estas
nele	neles	nela	nelas	naquele	naqueles
isso	isto	tal	tais		

Tabela A.3: Lista de palavras que podem representar uma transição

a fim de a princípio acima de tudo agora aliás anteriormente antes que assim como assim também cada vez que com a finalidade de com o intuito de como como também da mesma forma de fato de súbito decerto desde que dessarte do mesmo modo em primeiro lugar em seguida em virtude de enquanto freqüentemente igualmente indubitavelmente inopinadamente junto a logo logo que nem bem nesse íterim não raro ocasionalmente outrossim para que por analogia	a fim de que a priori ademais ainda cima além disso antes de mais nada ao mesmo tempo assim que atualmente caso com efeito com o propósito de como resultado conforme de acordo com de maneira idêntica de tal forma que demais dessa forma desse modo em conclusão em princípio em suma enfim eventualmente haja vista imediatamente inegavelmente inquestionavelmente junto de logo após mais adiante nesse hiato no momento em que não só ou por outra para exemplificar perto de por causa de	a posteriori a saber afinal ainda mais analogamente antes de tudo assim assim sendo bem como certamente com o fim de com toda a certeza como se constantemente às vezes de conformidade com de repente de tal sorte que depois que dessa maneira destarte em outras palavras em resumo em síntese enfim finalmente hoje imprevistamente inesperadamente isto é já que logo depois mas também nesse meio tempo não apenas não é certo ou seja para ilustrar pois por certo
--	---	---

por conseguinte por fim por vezes portanto pouco antes primordialmente próximo a quem sabe raramente sem dúvida sempre que sob o mesmo ponto de vista surpreendentemente tanto quanto uma vez que	por consequência por isso porquanto possivelmente pouco depois principalmente próximo de quer dizer se é que semelhantemente similarmente sobretudo tal qual tanto que visto que	por exemplo por outro lado porque posteriormente primeiramente provavelmente quando quicá segundo sempre simultaneamente subitamente talvez todas as vezes que é provável
--	---	--

Questionário Utilizado na Avaliação do Protótipo

1. Você tinha conhecimento prévio sobre a estruturação de um resumo acadêmico (em termos de componentes esperados)?

- ☐ não
- ☐ sim, mas mediano
- ☐ sim

2. Você corrigiu a classificação retórica dada pelo ambiente para as sentenças do resumo? Em quantas sentenças do total classificadas?

- ☐ não
- ☐ sim: (/) total

3. Quantas mensagens em relação à coerência foram emitidas pelo protótipo na primeira análise?

- | | |
|----------------------------------|---------------------------------|
| ☐ nenhuma | ☐ três |
| ☐ uma | ☐ quatro |
| ☐ duas | ☐ cinco |

4. As sugestões apresentadas em relação à coerência envolvem quais tipos de sentenças?

- | | |
|--|--|
| ☐ Título-Propósito | ☐ Propósito-Resultado |
| ☐ Lacuna-Contexto | ☐ Propósito-Conclusão |
| ☐ Propósito-Metodologia | |

5. Você acatou as sugestões emitidas pelo protótipo em relação a coerência?

- ☐ integralmente
- ☐ parcialmente
- ☐ não acatei

6. Na sua opinião, qual foi a relevância das sugestões emitidas pelo protótipo em relação a coerência?

- ☐ muito relevante
- ☐ relevante
- ☐ pouco relevante

7. Caso tenha acatado as sugestões, como você considera o nível de coerência do seu resumo final em relação a primeira versão?

- ☐ aumentou o nível de coerência
- ☐ aumentou razoavelmente o nível de coerência
- ☐ diminuiu o nível de coerência
- ☐ não afetou o nível de coerência

8. Quão difícil foi adequar o resumo de acordo com as sugestões apresentadas pelo protótipo?

- ☐ não fiz as correções
- ☐ não tive problemas para adequar o meu texto
- ☐ tive algumas dúvidas para adequar o meu texto
- ☐ tive muitas dúvidas para adequar o meu texto

9. Como você julga a informação das sugestões apresentadas pelo protótipo?

- ☐ muito informativa
- ☐ razoavelmente informativa
- ☐ pouco informativa