

Suporte Computacional à Escrita Científica em Português*

Valéria Delisandra Feltrim

Orientadora: Prof^a. Dr^a. Maria das Graças Volpe Nunes

Co-orientadora: Prof^a. Dr^a. Sandra Maria Aluísio

Monografia apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC/USP para o Exame de Qualificação, como parte dos requisitos para a obtenção do título de Doutor na Área de Ciências de Computação e Matemática Computacional.

São Carlos – SP
2002

* Trabalho realizado com o apoio financeiro da Fapesp, # 00/04790-6.

Resumo

Ferramentas destinadas a auxiliar a escrita na língua inglesa têm se tornado cada vez mais comuns, incluindo tanto sistemas dedicados ao pós-processamento do texto como sistemas mais completos, que atuam desde a geração e organização de idéias até a escrita. Entre esses sistemas há aqueles que apóiam especificamente a escrita de textos acadêmicos, um gênero mais tratável computacionalmente dada a sua rigorosa padronização. Em contraste, para o português não há sistemas semelhantes, mas apenas ferramentas de pós-processamento, como corretores ortográficos, revisores gramaticais, dicionários para consulta, entre outras, de aplicação geral. Nesse contexto, este trabalho tem por objetivo a análise do discurso acadêmico em textos em português — em termos de estruturação esquemática, estratégias retóricas e padrões de escrita — visando uma modelagem computacional para a criação e avaliação de ferramentas de apoio à escrita. Para viabilizar o projeto, restringimos nossa análise às seções Resumo, Introdução e Conclusão, notadamente mais padronizadas, de dissertações e teses na área de Ciências da Computação.

Abstract

Writing tools for English have been created for providing help during different stages of the writing process, from the generation and organization of ideas to post-processing tasks. Academic texts have been especially focused on due to their rigorous structure. For Portuguese, otherwise, one can find only post-processing tools, such as spell-checkers, grammar checkers and on-line dictionaries. This work aims at analysing the academic discourse in Portuguese texts — in terms of schematic structure, rhetorical strategies and writing patterns — to derive a computational model to be used in the creation and evaluation of writing tools. To make this project feasible, we have focused on the Abstract, Introduction and Conclusion sections of monographs and theses on Computer Science.

Índice

1 INTRODUÇÃO	1
2 A ESTRUTURAÇÃO DE TEXTOS CIENTÍFICOS.....	7
2.1 ESTRUCTURA ESQUEMÁTICA GLOBAL.....	8
2.2 DETALHANDO O RESUMO	10
2.3 DETALHANDO A INTRODUÇÃO.....	13
2.3.1 <i>O Modelo de Swales</i>	14
2.3.2 <i>O Modelo de Weissberg & Buker</i>	15
2.3.3 <i>O Modelo de Booth, Colomb e Willians</i>	16
2.3.4 <i>O Modelo de Aluísio</i>	17
2.3.5 <i>Comentários sobre os Modelos</i>	19
2.3.6 <i>Detalhando os Elementos da Introdução</i>	20
2.3.7 <i>Considerações sobre a Introdução</i>	29
2.4 DETALHANDO OS MATERIAIS E MÉTODOS.....	30
2.5 DETALHANDO OS RESULTADOS	32
2.6 DETALHANDO A CONCLUSÃO	33
2.7 CONSIDERAÇÕES FINAIS	36
3 FERRAMENTAS DE AUXÍLIO À ESCRITA.....	37
3.1 WRITING ENVIRONMENT – WE	38
3.2 WRITER’S ASSISTANT.....	40
3.3 COMPOSER.....	42
3.4 ACADEMIC WRITER.....	44
3.5 ABSTRACT HELPER.....	47
3.6 AMADEUS	48
3.6.1 <i>Ferramenta de Referência</i>	49
3.6.2 <i>Ferramenta de Suporte</i>	52
3.6.3 <i>Ferramenta de Crítica</i>	55
3.6.4 <i>Ferramenta Tutorial</i>	57
3.7 COMENTÁRIOS SOBRE AS FERRAMENTAS	58
3.8 CONSIDERAÇÕES FINAIS	59

4 ATIVIDADES DESENVOLVIDAS.....	61
4.1 O CORPUSDT	61
4.1.1 <i>Por que Formar um Novo Corpus?</i>	61
4.1.2 <i>Os Textos Coletados</i>	62
4.2 ANÁLISE SUPERFICIAL DA ESTRUTURA ESQUEMÁTICA NO CORPUSDT	64
4.3 A ETIQUETAÇÃO DO CORPUS	67
4.4 PESQUISA REALIZADA COM USUÁRIOS ALVO.....	71
4.5 CONSIDERAÇÕES FINAIS	78
5 PROPOSTA E METODOLOGIA	79
5.1 METODOLOGIA	80
5.2 CRONOGRAMA DE TRABALHO	84
5.3 VIABILIDADE E RECURSOS DISPONÍVEIS.....	85
REFERÊNCIAS BIBLIOGRÁFICAS	87
APÊNDICE A	93
APÊNDICE B.....	99
APÊNDICE C	107

Índice de Figuras

Figura 2.1. Movimento Geral-Específico-Geral através da estrutura global do texto (Weissberg & Buker, 1990).....	9
Figura 2.2. Movimento Geral-Específico seguido na seção de contextualização (Weissberg & Buker, 1990).....	21
Figura 2.3. Movimento Específico-Geral seguido na Conclusão (Weissberg & Buker, 1990).....	35
Figura 3.1. Dumping de tela mostrando os quatro módulos do WE (Lansman et al., 1993) ..	40
Figura 3.2. Dumping de tela do módulo <i>Notes View</i> do Writer's Assistant.....	41
Figura 3.3. Dumping de tela do módulo <i>Linear View</i> do Writer's Assistant	42
Figura 3.4. Ilustração da visualização de um esquema estrutural no Composer.....	44
Figura 3.5. Menu principal do Academic Writer (Broady & Shurville, 2000)	45
Figura 3.6. Arquitetura da ferramenta de Referência (Aluísio, 1995).....	51
Figura 3.7. Arquitetura da ferramenta de Suporte (Aluísio, 1995)	54
Figura 3.7. A abordagem de críticas (Fischer et al., 1991)	55
Figura 3.8. A arquitetura da Ferramenta de Crítica (Silva, 1999).	57
Figura 4.1. Representação gráfica do <i>XML Schema</i> (nível 0) para marcação do CorpusDT, mostrando a expansão do elemento <i>CorpoDissertação</i> (nível 1).....	70
Figura 4.2. Representação gráfica do elemento <i>Resumo</i> (nível 2 do <i>XML Schema</i>), mostrando a expansão dos elementos <i>Subcomponente</i> (nível 3) e <i>Estratégia</i> (nível 4) ...	70

Índice de Tabelas

Tabela 3.1. Adequabilidade do tipo de ferramenta quanto ao conhecimento do usuário (Barros, 2000).....	49
Tabela 4.1. Número de trabalhos em cada área de pesquisa	63
Tabela 4.2. Totais de palavras em dissertações e teses.....	63
Tabela 4.3. Totais de palavras divididos por área de conhecimento	64
Tabela 4.4. Dados sobre o tamanho em palavras dos resumos.....	65
Tabela 4.5. Dados sobre o tamanho em palavras das introduções.....	66
Tabela 5.1. Cronograma de atividades previstas	85

Índice de Gráficos

Gráfico 4.1. Respostas de estudantes iniciantes e experientes quanto ao nível de conhecimento da escrita acadêmica.....	72
Gráfico 4.2. Respostas de estudantes iniciantes e experientes quanto ao número de textos acadêmicos já produzidos	72
Gráfico 4.3. Respostas de estudantes iniciantes e experientes quanto ao número médio de revisões em cada texto	73
Gráfico 4.4. Respostas de estudantes iniciantes e experientes quanto ao nível de influência do orientador (ou alguém mais experiente) na produção de um texto.....	73
Gráfico 4.5. Respostas de estudantes iniciantes e experientes quanto ao tipo de texto utilizado como modelo para a escrita do seu próprio texto	73
Gráfico 4.6. Respostas dos estudantes experientes quanto ao grau de dificuldade encontrado na escrita de cada componente	74
Gráfico 4.7. Respostas dos estudantes iniciantes quanto ao grau de dificuldade encontrado na escrita de cada componente	74
Gráfico 4.8. Respostas dos estudantes experientes quanto a relevância das funções de uma ferramenta computacional de auxílio à escrita	75
Gráfico 4.9. Respostas dos estudantes iniciantes quanto a relevância das funções de uma ferramenta computacional de auxílio à escrita	76
Gráfico 4.10. Respostas dos estudantes iniciantes e experientes quanto ao estilo de interação de interação com o usuário	78
Gráfico 4.11. Respostas dos estudantes iniciantes e experientes quanto à relevância do acoplamento de uma ferramenta de auxílio à escrita acadêmica a um processador de textos.....	78

Índice de Quadros

Quadro 2.1. Componentes do plano padrão para textos científicos	9
Quadro 2.2. Modelo de resumo típico (Weissberg & Buker, 1990)	12
Quadro 2.3. Modelo de resumo curto (Weissberg & Buker, 1990)	12
Quadro 2.4. Modelo de introdução CARS (Swales, 1990)	15
Quadro 2.5. Modelo para a estrutura de introduções (Weissberg & Buker, 1990)	16
Quadro 2.6. Modelo para a estruturação de introduções (Booth et al., 2000)	17
Quadro 2.7. Esquema detalhado para introduções (Aluísio & Oliveira Jr., 1996)	18
Quadro 2.8. Modelo simplificado para introduções	20
Quadro 2.9. Modelo de ordenação dos possíveis elementos da seção de materiais e métodos (Weissberg & Buker, 1990)	31
Quadro 2.10. Categorias de materiais (Weissberg & Buker, 1990)	31
Quadro 2.11. Possíveis passos para o início de uma conclusão (Weissberg & Buker, 1990)	34
Quadro 3.1. Exemplos de expressões padrões para introduções (Aluísio, 1995)	50
Quadro 4.1. Modelo de Resumo utilizado para a marcação dos resumos do CorpusDT	67

Capítulo 1

Introdução

Nas duas últimas décadas, várias foram as ferramentas e ambientes de software produzidos para auxiliar o processo de escrita em inglês. Como exemplos de sistemas que apóiam o processo de composição de um texto de forma mais abrangente, isto é, tanto a geração de idéias quanto a escrita do texto, podemos citar o Composer (Pemberton et al., 1996), o Writer's Assistant (Sharples et al., 1994), e o WE (Smith & Lansman, 1988). Outras ferramentas que oferecem um suporte mais específico e que já se encontram comercialmente disponíveis são os corretores gramaticais, como o Grammatik (RSI, 1992) e o Correct Grammar (WTG, 1991), os ambientes de referência e auto-instrução como o HELPDISK! 2.6 e English Language 2.0¹, corretores de estilo como o StyleWriter² e vários dicionários e thesauri eletrônicos. A grande maioria é dedicada ao pós-processamento do texto visando corrigi-lo tanto do ponto de vista gramatical, como da clareza, estilo e concisão das idéias apresentadas, conforme levantamento descrito em Fontana et al. (1993).

Tais ferramentas de pós-processamento oferecem várias informações estatísticas sobre o texto, tais como o número e tamanho médio de parágrafos, sentenças e palavras do texto, e também estatísticas que avaliam o grau de legibilidade do texto. Além disso, oferecem regras para a correção da ortografia, da gramática, de adequação do estilo do texto a certo gênero (comercial, científico, jornalístico), ou a um padrão específico de uma companhia ou editora.

¹ <http://www.mantex.co.uk/software.htm> (Acessado em 18-02-2002)

² <http://www.editorsoftware.com> (Acessado em 18-02-2002)

Embora esse tipo de auxílio seja extremamente útil para usuários com dificuldades lingüísticas, existe uma limitação para o uso dessas ferramentas pois proporcionam somente mecanismos de correção de um texto acabado. Entretanto, nem sempre um usuário é capaz de produzir um texto que possa vir a ser simplesmente melhorado. Além disso, as ferramentas de pós-processamento não resolvem uma dificuldade comum a muitos escritores: a produção do primeiro rascunho. Uma possível estratégia para minimizar esse problema seria a reutilização de sentenças, parágrafos e trechos maiores, como as seções de um artigo, compilados a partir de textos relevantes ao gênero em questão (Aluísio, 1995; Silva, 1999; Barros, 2000).

Uma vertente importante de ferramentas que adotam tal estratégia é aquela formada por ferramentas de auxílio à escrita em inglês para não-nativos, ou seja, usuários que não têm o inglês como língua materna. Um gênero em que é evidente a necessidade de ferramentas como essas para a produção de textos é o acadêmico³, já que a divulgação desses trabalhos se dá prioritariamente em revistas especializadas e mundialmente veiculadas, que adotam o inglês como língua padrão. Um exemplo de ambiente dessa natureza, que trata tanto do processo da escrita de um texto como de seu produto, é o AMADEUS – *Amiable Article Development for User Support* (Oliveira Jr. et al., 1992; Fontana et al., 1993; Aluísio, 1995; Aluísio & Oliveira Jr., 1995; Aluísio & Oliveira Jr., 1996; Aluísio & Gantenbein, 1997; Aluísio et al., 2001).

O AMADEUS é um ambiente de auxílio e ensino da escrita técnica que conta com várias ferramentas inter-relacionadas: Ferramenta de Referência, Ferramenta de Suporte, Ferramenta de Crítica e Ferramenta Tutorial, sendo que cada uma dessas ferramentas se aplica especificamente a um conjunto de problemas caracterizado por diferentes níveis de dificuldade no inglês e grau de experiência na escrita técnica. Outro exemplo de ambiente voltado ao auxílio à escrita para não-nativos em inglês é o sistema Composer (Pemberton et al., 1996; Shurville et al., 1997), composto de ferramentas para o planejamento (ou geração de idéias) e composição de textos que foi fundamentado em teorias cognitivas do processo de escrita. Enquanto o Composer focaliza a organização de idéias, o AMADEUS enfatiza estratégias para a organização do texto escrito. Outra característica do ambiente AMADEUS é que o usuário encontra auxílio durante a criação do texto e após a geração de um produto (esquema detalhado de um texto), que pode ser submetido a uma ferramenta de crítica (Silva et al., 1998; Silva, 1999) para melhorá-lo. Assim, o usuário é apoiado tanto durante o processo de escrita quanto no processo de crítica do produto produzido.

³ Chamamos de “gênero acadêmico” aquele que abrange a divulgação de trabalhos acadêmicos dentro de uma comunidade científica. Exemplos desse gênero são os artigos para conferências ou periódicos, dissertações, teses, relatórios técnicos, etc.

Se, por um lado, as ferramentas destinadas a auxiliar a escrita na língua inglesa tornam-se cada vez mais comuns, ainda não são encontrados similares para a língua portuguesa. Só recentemente têm surgido ferramentas lingüísticas de pós-processamento, como corretores ortográficos, revisores gramaticais, dicionários para consulta, entre outras. Contudo, nenhuma dessas ferramentas, que focalizam o produto, agrupa todas as funcionalidades desejáveis de um ambiente de auxílio à escrita focalizando todo o processo de escrita, como as já existentes para o inglês.

A experiência adquirida pelo grupo atuante no NILC⁴ desde 1993, na pesquisa e desenvolvimento de sistemas (revisores ortográfico e gramatical) e recursos lingüísticos (léxicos, etiquetadores morfossintáticos, corpora, etc), notadamente para o português brasileiro, somada à experiência no desenvolvimento do ambiente de auxílio e ensino da escrita técnica AMADEUS, tem motivado a investigação de ferramentas de auxílio à escrita do português que sejam mais direcionadas a um gênero particular de escrita (ao contrário dos ambientes de processadores de texto), tornando-se potencialmente mais úteis para uma determinada classe de usuários.

Não seria, no entanto, de grande utilidade uma simples adaptação do AMADEUS para a língua portuguesa, por dois fatores principais. Primeiramente porque o AMADEUS visa auxiliar escritores não-nativos enquanto uma ferramenta voltada para escrita em português se destinaria a escritores nativos, que podem ter necessidades diferentes. Em segundo lugar, o AMADEUS trata artigos científicos e é notória a solicitação para que pesquisadores brasileiros publiquem seus trabalhos científicos em revistas internacionais, portanto, em língua inglesa. Contudo, há uma classe de documentos acadêmicos que necessariamente devem ser escritos em português, cuja organização e redação têm sido objetos de estudo constante de profissionais da área de Letras. Trata-se das teses e dissertações acadêmicas.

É inegável que ferramentas de auxílio ao processo (composição) e produto (escrita) desses documentos seriam de grande valia tanto para os estudantes-autores de teses e dissertações, quanto para os orientadores que exigem qualidade naquilo que é o veículo oficial de divulgação de um trabalho científico desenvolvido, em geral, ao longo de vários anos. Além disso, quando o estudante-autor aprende a escrever seu trabalho de forma correta, ele passa a ter maior consciência do processo envolvido na investigação científica, uma vez que a escrita é a explicitação desse processo. Adicionalmente, tais ferramentas têm potencial para facilitar o aprendizado de quem tem a escrita como objeto de estudo, ampliando assim o conjunto de usuários que se beneficiariam delas.

⁴ Núcleo Interinstitucional de Lingüística Computacional (www.nilc.icmc.sc.usp.br)

Reconhece-se que qualquer sistema de auxílio à escrita, e em particular a escrita acadêmica, para ser eficaz, deve levar em conta as particularidades da grande área de conhecimento em que está inserido o trabalho científico — de ciências exatas, biológicas ou humanas —, pois cada uma tem suas idiossincrasias. É bem provável que a natureza de algumas delas ofereça desafios maiores do que as demais, como os textos acadêmicos da área de humanas, que tendem a ser dissertativos, com uma organização não-padronizada ou que obedeça a padrões difíceis de serem identificados ou formalizados para tratamento computacional. Já os textos acadêmicos da área de exatas parecem ter estruturas discursivas bastante padronizadas, em especial os trabalhos que envolvem pesquisa experimental, e têm sido objetos de estudo frequentes dos pesquisadores da escrita em inglês (Weissberg & Buker, 1990; Swales, 1990) e mesmo para o português (Severino, 1996; Feitosa, 1991). No entanto, para a realização de um estudo mais aprofundado, é preciso definir um recorte no domínio de estudo, uma vez que diferentes subáreas das ciências exatas podem apresentar certas particularidades.

Nesse sentido, neste projeto estaremos focalizando o problema da escrita em português de teses e dissertações na área de Ciências da Computação, visando uma modelagem computacional para a criação e avaliação de ferramentas de apoio à criação de textos desse gênero. Seguindo a linha de pesquisa baseada em corpus, será necessária a compilação de um corpus especialmente dedicado a essa pesquisa, com marcações apropriadas, tais como identificação das diferentes seções, função dos diversos trechos, expressões padronizadas, entre outras. Esse corpus servirá tanto para a análise de padrões de escrita específicos da comunidade focalizada como para a realização de uma análise crítica, visando identificar os principais problemas encontrados no corpus e que poderiam ser minimizados através do auxílio de uma ferramenta computacional.

Para que as ferramentas de auxílio à escrita de teses e dissertações sejam mais efetivas, é preciso delimitar seu escopo de ação, uma vez que muito esforço é necessário para o tratamento de cada uma das partes de um texto desse gênero, conforme mostram duas das ferramentas desenvolvidas no projeto AMADEUS, que têm tratado especificamente a seção de introdução. Decidimos tratar neste projeto três seções de texto reconhecidamente importantes e problemáticas, além de serem mais padronizadas, o Resumo, a Introdução e a Conclusão. A escolha dessas seções também se deu devido aos resultados obtidos em uma pesquisa realizada com estudantes de pós-graduação, em que tais seções foram apontadas como as que oferecem maior grau de dificuldade na escrita.

É importante ressaltar que aqui estaremos tentando utilizar toda a tecnologia de análise de corpus e implementação desenvolvida no projeto AMADEUS, no sentido de queimar etapas, porém não descartando a adaptação ou mesmo inovação nos métodos lá utilizados. É preciso ressaltar que, embora o desenvolvimento do projeto AMADEUS tenha sido iniciado há alguns anos, sua implementação utiliza técnicas que estão de acordo com o estado da arte, conforme apontado no Capítulo 3. Também é nosso objetivo alcançar a fase de avaliação da ferramenta no seu contexto de uso, e, para tanto, definir uma metodologia de avaliação baseada em critérios de performance de sistemas de PLN⁵ (Santos, 2000; Di Eugenio, 2000; Spark Jones & Galliers, 1996; King, 1996) e critérios de usabilidade (Dix et al., 1999; Treu, 1994; Nielsen, 1993).

Se a tecnologia computacional aplicada ao AMADEUS deve ser aproveitada ao máximo neste projeto, suas contribuições devem, então, se dar em dois outros eixos:

- (a) na investigação profunda do discurso científico do português, especificamente na área de Ciências da Computação, incluindo a construção de corpus e a extração de modelos e padrões que devem nortear o tratamento computacional da língua nesse contexto; e
- (b) na definição e aplicação de uma abordagem para avaliação das ferramentas propostas, quer seja sob o ponto de vista funcional da mesma, quanto, e principalmente, sob o ponto de vista da ajuda efetiva que se propõem a dar.

Em ambos os casos, acreditamos constituir uma iniciativa inédita em se tratando da escrita científica na língua portuguesa.

Nos dois próximos capítulos é apresentada a revisão bibliográfica que deverá embasar este trabalho, de forma que são discutidos dois tópicos principais: a estruturação de textos científicos em português, discutida no Capítulo 2, e as ferramentas de auxílio à escrita encontradas na literatura, em especial do ambiente AMADEUS, apresentadas no Capítulo 3. No Capítulo 4 são descritas as atividades previstas para o projeto e que já foram realizadas. O Capítulo 5 apresenta a proposta de trabalho, detalhando a metodologia empregada e as próximas tarefas a serem cumpridas. Finalmente, é discutido o cronograma a ser seguido para o desenvolvimento do projeto.

⁵ Processamento de Linguagem Natural.

Capítulo 2

A Estruturação de Textos Científicos¹

“Escrever é parte inerente ao ofício do pesquisador” (Feitosa, 1991). O trabalho do cientista não se esgota nas descobertas que faz. É de sua responsabilidade comunicar os seus resultados, suas descobertas, suas criações. Sendo assim, a escrita científica caracteriza-se como um processo de comunicação muito importante, pois uma descoberta científica torna-se reconhecida através das publicações de seus resultados.

Porém, escrever não implica em, necessariamente, comunicar com eficiência. Um trabalho escrito sem nenhum cuidado, sem a preocupação em guiar o leitor proporcionando-lhe uma leitura esclarecedora e agradável, não está comunicando como deveria, até porque ninguém se dispõe a ler um trabalho confuso e de leitura desestimulante. Dessa forma, para que um trabalho seja bem escrito e, conseqüentemente, comunique bem o seu propósito, é preciso atentar a uma série de fatores, tanto do seu conteúdo quanto da sua estrutura.

Na literatura, muito já foi discutido sobre a pesquisa científica como um processo ordenado e documentado, sendo que a maioria dos autores discute o que o conteúdo de um texto acadêmico deve relatar e as etapas que devem ser seguidas para se chegar à informação que será incluída em um trabalho, discutindo-se a metodologia da pesquisa com alguma informação sobre a estrutura, como em Eco (2000), Tachizawa & Mendes (2000), Vieira (1999), Severino (1996), Feitosa (1991), Castro (1981), Barrass (1979), Solomon (1977), Hegenber (1976), Rey (1972). Porém, poucos trabalhos discutem mais detalhadamente a

¹ Este capítulo consiste de uma versão estendida de Feltrim et al. (2000).

estrutura esquemática de um texto acadêmico em português. Estruturas esquemáticas (ou superestruturas) foram definidas por Kintsch e van Dijk (1978)² como sendo os componentes do discurso que definem sua forma e ajudam a identificar seu gênero.

Alguns autores tratam a estruturação do texto acadêmico mais profundamente, como Huckin & Olsen (1991), Weissberg & Buker (1990), Swales (1990) e Trimble (1985), porém esses trabalhos são voltados para o inglês. Apesar da diferença de língua, a estrutura esquemática discutida nesses trabalhos pode ser aplicada também para o português. Embora se saiba que existem muitos fatores que podem influenciar a composição de um texto, como fatores culturais (Hartley & Paris, 1994), (Linden & Scott, 1995), (Paris & Evans, 1994), estudos como de Taylor e Tingguang (1991) indicam que apesar de existirem diferenças na organização e na elaboração dos componentes da estrutura esquemática, devido a diferenças entre nacionalidades, culturas e disciplinas, os trabalhos acadêmicos compartilham uma mesma estrutura. De fato, analisando-se os trabalhos citados, fica claro que as diferenças de idioma não interferem na estrutura mais global de um texto acadêmico.

O conteúdo apresentado a seguir é uma compilação das observações sobre a estruturação esquemática de textos científicos que foram discutidas nos trabalhos citados anteriormente.

2.1 Estrutura Esquemática Global

Segundo Rey (1972), antes de se começar a escrever deve-se considerar a estrutura do texto que se pretende redigir. Para isso, deve-se estabelecer um esquema que permita expor as idéias de maneira sistemática e lógica, reunindo em cada item assuntos correlatos, sem risco de omitir ou de repetir as mesmas coisas ao longo do trabalho. Um trabalho científico pode enquadrar-se, em geral, dentro de um esquema que já se tornou clássico pela simplicidade, pelo desenvolvimento metódico e por abranger aspectos essenciais de uma comunicação científica desse gênero.

Todos os autores pesquisados (Eco, 2000; Tachizawa & Mendes, 2000; Vieira, 1999; Severino, 1996; Feitosa, 1991; Castro, 1981; Barrass, 1979; Solomon, 1977; Hegenber, 1976; Rey, 1972; Huckin & Olsen, 1991; Weissberg & Buker, 1990; Swales, 1990; Trimble, 1985) concordam quanto à estrutura esquemática que um texto científico deve seguir. Essa estrutura pode ser enunciada como *Introdução – Desenvolvimento – Conclusão*, sendo que o

² Kintsch, W., van Dijk, T.A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, pp. 363-394 *apud* (Aluísio, 1995)

Desenvolvimento pode desdobrar-se nas seções de *Materiais e Métodos* e *Resultados*, ou ainda *Materiais e Métodos*, *Resultados* e *Discussão*. Essa forma de estruturação tem como objetivo apresentar o texto a partir do contexto no qual ele está inserido. Para isso, cada um dos componentes da estrutura desempenha um papel bem definido, que será discutido mais adiante. Em linhas gerais, essa estrutura deve guiar o leitor e fazer com que ele siga o movimento geral-para-específico, realizado na *Introdução*, e específico-para-geral, realizado na *Conclusão*, conforme ilustrado na Figura 2.1.

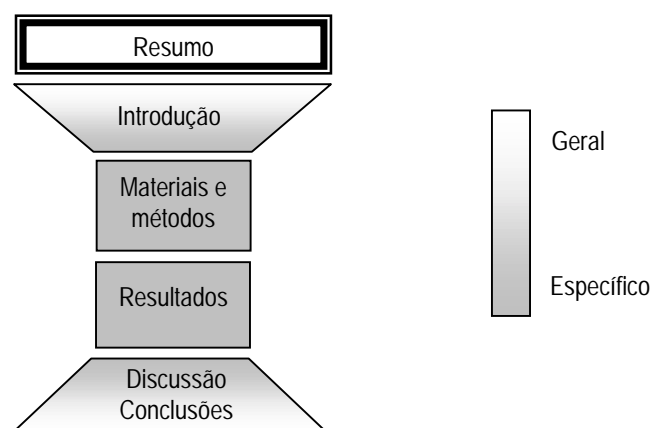


Figura 2.1. Movimento Geral-Específico-Geral através da estrutura global do texto
(Weissberg & Buker, 1990)

Note que a Figura 2.1 apresenta um outro componente chamado *Resumo*. O *Resumo* aparece de uma forma destacada na figura por ser um componente independente do restante da estrutura. Devido ao fato de conter informações relativas a todo o texto, tanto gerais como específicas, o *Resumo* não segue o movimento da estrutura global apresentada. Sendo assim, podemos definir um plano padrão para estruturação esquemática de textos científicos, contendo os seguintes componentes: *Resumo*, *Introdução*, *Materiais e Métodos*, *Resultados*, *Discussão* e *Conclusão*. O Quadro 2.1 indica em linhas gerais o que deve ser incluído em cada um desses componentes.

Quadro 2.1. Componentes do plano padrão para textos científicos

Resumo	<i>Sucinta indicação das principais descobertas</i>
Introdução	<i>O que fez o autor? Por quê?</i>
Materiais e Métodos	<i>Como fez?</i>
Resultados	<i>O que foi encontrado?</i>
Discussão	<i>Interpretação dos resultados</i>
Conclusão	<i>Conclusões do trabalho</i>

É importante ressaltar que, enquanto a estrutura esquemática global é ditada por um esquema mais fixo, a estrutura mais detalhada não é (Smith & Lansman, 1988). Sendo assim, discutiremos nas próximas seções as particularidades de cada um dos componentes e quais os respectivos papéis dentro da estrutura esquemática global.

2.2 Detalhando o Resumo

O resumo é um componente independente, isto é, não interfere no movimento geral-específico-geral que o texto como um todo deve realizar. Trata-se de um componente complementar que pode preceder o texto a fim de dar ao leitor uma visão geral dos principais assuntos tratados. No entanto, essa característica não o torna um componente menos importante, já que muitos leitores se limitam a ler o título e o resumo de um trabalho (Barrass, 1979). Sendo assim, o resumo deve ser redigido com muito cuidado, de forma a ser completo, interessante e informativo, dispensando a consulta ao restante do texto para que o leitor tenha a idéia do que trata o trabalho e, ao mesmo tempo, estimulando o interesse pela leitura do texto completo.

A Sumarização Automática (SA) é uma subárea de PLN que se dedica exclusivamente ao estudo de resumos (tanto os acadêmicos quanto de outros gêneros de texto), de modo que muito já foi discutido sobre essa seção. Um dos aspectos dos resumos que é discutido em SA é a sua informatividade. De acordo com esse aspecto, um resumo pode ser classificado em três categorias (Mani, 2001; Feitosa, 1991; Rey, 1972):

- Informativo (ou autocontido), quando contém todas as informações apresentadas no texto e dispensa a leitura desse último;
- Indicativo, quando não dispensa a leitura do texto completo, pois apenas descreve a natureza, a forma e o propósito do trabalho, cuja matéria não é passível de uma apresentação abreviada. Pode ser visto como um índice, que indica os assuntos tratados no restante do trabalho.
- Crítico, quando formula julgamento sobre o trabalho.

Dependendo da função do resumo, um ou outro tipo é mais adequado. Alguns autores (Feitosa, 1991; Vieira, 1999; Oliveira, 2001) indicam que o resumo constituinte de um texto acadêmico deve ser preferencialmente informativo, descrevendo quais os propósitos, resultados, e conclusões do trabalho em vez de indicativo, dizendo, por exemplo, que o resultado se encontra no texto. Embora outros autores não explicitem o caráter informativo do

resumo de um texto acadêmico, a preferência por um resumo informativo pode ser notada através dos modelos de resumo sugeridos.

Os resumos de quase todas as áreas de estudo são escritos de uma maneira muito similar (Weissberg & Buker, 1990). Os tipos de informação incluídos e a ordem em que aparecem são muito convencionais, de modo que podem ser enunciados como modelos de resumo. Tais modelos objetivam guiar o escritor no sentido do tipo de informação que deve ser incluída em um bom resumo e da ordem que tais informações devem aparecer.

A seguir são apresentados dois exemplos de resumos, um indicativo e outro informativo, respectivamente. Esses exemplos são trechos autênticos extraídos do corpus construído como uma das etapas desse trabalho de doutorado³.

Resumo indicativo:

“Um estudo empírico visando avaliar a eficácia em revelar erros, a dificuldade de satisfação e o custo de aplicação do critério Análise de Mutantes é apresentado neste trabalho. Eficácia e custo também foram avaliados para os critérios Potenciais-Usos, comparando-se assim os resultados obtidos para esses critérios. A especificação e implementação de uma estratégia de minimização de conjuntos de casos de teste adequados ao critério Análise de Mutantes também é apresentada”.

Resumo informativo:

“Este trabalho apresenta um protótipo de ferramenta, a Html2Hip, que proporciona um ambiente de importação e adaptação de documentos descritos segundo o padrão HTML (HyperText Markup Language) para a representação interna do SASHE (Sistema de Autoria e Suporte Hiperídia para Ensino), que se baseia na estruturação de objetos multimídia, segundo a hierarquia de classes proposta pelo MCA (Modelo de Contextos Aninhados). Além disso, este trabalho estendeu a capacidade do editor de nós de informação do tipo texto do protótipo anterior no que concerne ao processamento de arquivos-texto descritos pelo padrão RTF (Rich Text Format). Dessa forma, o SASHE tornou-se capaz de processar e organizar materiais instrucionais preparados em seu próprio ambiente, no ambiente WWW (World-Wide Web), bem como em processadores de texto comuns”.

Vários autores apresentam modelos de resumo, entre eles Booth et al.(2000), Huckin & Olsen (1991) e Weissberg & Buker (1990). Embora cada autor tenha a sua forma de expressar o modelo, existe um consenso sobre os elementos típicos e sua ordem. Apresentaremos o modelo de Weissberg & Buker (Quadro 2.2), por ser o mais detalhado e abranger também os outros modelos citados.

³ Esse corpus, chamado CorpusDT, é apresentado no Capítulo 4.

Quadro 2.2. Modelo de resumo típico (Weissberg & Buker, 1990)

-
1. Alguma informação de contextualização (*background*)
 2. A principal atividade do estudo (seu propósito) e seu escopo
 3. Alguma informação sobre a metodologia usada no estudo
 4. Os resultados mais importantes do estudo
 5. Uma afirmação de conclusão ou recomendação
-

Observando o Quadro 2.2, fica fácil perceber que um resumo conterá informações de outras seções do texto. Dessa forma, é conveniente que o resumo seja elaborado após o término do trabalho (Barrass, 1979). Embora seja o componente que aparece primeiro, geralmente é o último a ser escrito.

Note que os cinco elementos citados no modelo de Weissberg & Buker direcionam para a composição de um resumo informativo, onde todos os elementos principais de um texto são citados. Esses elementos são semelhantes aos encontrados nas introduções, como será visto adiante (Seção 2.3), porém escritos de forma tão concisa quanto possível. Há ocasiões, no entanto, em que só uma escrita concisa não basta, pois existe um limite pequeno de palavras a serem usadas que deve ser respeitado, como o caso de resumos para artigos publicados em revistas. Nesses casos, um tipo mais curto de resumo pode ser escrito através da eliminação ou combinação de alguns dos elementos citados anteriormente, conforme apresentado no Quadro 2.3.

Quadro 2.3. Modelo de resumo curto (Weissberg & Buker, 1990)

-
1. Propósito + metodologia
 2. Resultados principais
 3. Conclusão/Recomendação (opcional)
-

O resumo curto geralmente é escrito usando apenas dois ou três elementos dos cinco citados no modelo anterior, com a ênfase sendo colocada nos resultados do estudo. A informação de *background* nesse caso é retirada, sendo apresentados como primeiros elementos os propósitos e a metodologia utilizada, combinadamente. Em seguida deve ser feito um resumo dos resultados mais importantes do trabalho e finalmente pode ser colocada uma conclusão ou recomendação em uma ou duas sentenças. Essa conclusão ou recomendação é opcional, e pode ser ocultada caso seja necessário.

Algumas diretrizes relativas a atributos sintáticos dos elementos do resumo também foram encontradas na literatura pesquisada. Barrass (1979) sugere que o resumo deve ser escrito preferencialmente na terceira pessoa. Já Weissberg & Buker (1990) e Huckin & Olsen (1991) fazem recomendações sobre os tempos verbais a serem utilizados em cada uma das seções do resumo. São elas:

- A informação de *background* pode ser escrita no presente;
- O propósito do trabalho pode aparecer tanto no presente como no passado;
- A metodologia pode ser escrita no passado;
- Os resultados devem ser escritos no passado;
- As conclusões ou recomendações devem ser escritas no presente.

A variância do tempo verbal utilizado para o propósito será discutida adiante na Seção 2.3.6.4, que trata esse elemento mais detalhadamente.

2.3 Detalhando a Introdução

A introdução é o componente que inicia o movimento geral-específico-geral que o texto deverá seguir até o seu término (Figura 2.1). Ela é a responsável por fazer com que o leitor seja guiado de forma coerente e agradável, partindo-se de um contexto geral até o assunto específico tratado pelo escritor. Dessa forma, ela própria segue o movimento geral-específico. Solomon (1977) diz que a função da introdução é apresentar o assunto do trabalho, colocando o tema abordado pelo escritor dentro de um quadro de referência teórica atualizado. No entanto, escrever uma introdução que realize tal função de maneira bem sucedida não é tarefa fácil, especialmente para escritores acadêmicos iniciantes. Mesmo escritores experientes admitem sentirem mais dificuldade em escrever a introdução do que o restante do trabalho (Swales, 1990).

Sendo assim, grande parte dos autores que discutem a estrutura esquemática de textos científicos dedica uma atenção maior ao componente Introdução. Enquanto alguns autores apresentam observações que podem ser tomadas como diretrizes no momento da escrita, outros apresentam modelos mais completos, descrevendo os elementos típicos encontrados em boas introduções, pois assim como os resumos, as introduções tendem a apresentar um padrão de elementos e de ordenação desses elementos.

Em Feitosa (1991) encontramos uma discussão interessante sobre a estrutura da introdução. A autora argumenta que pode haver situações em que o modelo típico de

introdução deixa de ser adequado e que é importante considerar o papel, ou seja, a função que a introdução deve desempenhar para que o processo de comunicação seja efetivo. Sendo assim, em vez de seguir um modelo pré-estabelecido de introdução, deve-se considerar que ela tem a função de levar o leitor a ler o texto pelo interesse nas informações nele contidas. Assim, a função da introdução é transmitir ao leitor informações que o deixe a par dos assuntos tratados no restante do trabalho e que o entusiasme a continuar a leitura. Dessa forma, nos casos em que um esquema tradicional de introdução for adequado então ele deve ser usado. Caso contrário é melhor procurar construir uma introdução apropriada, para não haver risco de que a introdução se torne um elemento isolado do restante do trabalho.

Um esquema tradicional de introdução pode ser resumido em três elementos ou estágios: Contextualização Geral, Revisão Bibliográfica e Propósito do trabalho. Todos os modelos de introdução estudados baseiam-se nesses elementos, porém diferem no detalhamento de cada elemento e também na inclusão de outros elementos que complementam cada modelo.

Barrass (1979) descreve um modelo simples de introdução focalizando os três elementos citados. Segundo esse autor, a introdução deve trazer uma clara apresentação do problema focalizado, indicando o escopo do trabalho, isto é, estabelecendo o contexto no qual o trabalho se desenvolve (Contextualização Geral). Deve incluir uma breve referência a qualquer nota preliminar e outras investigações relevantes, próprias do autor ou alheias, para mostrar – se for o caso – de que modo o trabalho a ser relatado está baseado em trabalhos anteriores (Revisão Bibliográfica). Deve dizer também o que especificamente é tratado no trabalho e os motivos que levaram a sua realização (Propósito).

Outros autores apresentam modelos de introdução mais detalhados, onde são apresentados não só os elementos ou estágios do modelo, mas também os passos que compõem cada elemento/estágio. Como exemplos podemos citar os modelos de Swales (1990), Weissberg & Buker (1990), Booth et al. (2000) e Aluísio (1995). A seguir apresentamos um breve resumo de cada um desses modelos.

2.3.1 O Modelo de Swales

Em Swales (1990) é apresentado um modelo para a estruturação de introduções chamado CARS – *Create a Research Space*. Esse modelo foi derivado da análise de um corpus de introduções das áreas da física, biologia/medicina e ciências sociais. Foi um dos modelos de introdução mais comentados (Aluísio, 1995; Huckin & Olsen, 1991; Weissberg & Buker,

1990) e é constituído de três movimentos, conforme apresentado no Quadro 2.4, sendo que cada movimento é refinado em outros passos.

Quadro 2.4. Modelo de introdução CARS (Swales, 1990)

–	Movimento 1	Estabelecendo um território
	Passo 1	Mostrando proeminência e/ou
	Passo 2	Generalizando o tópico e/ou
	Passo 3	Revisando pesquisas anteriores
–	Movimento 2	Estabelecendo um nicho
	Passo 1A	Contra-argumentando ou
	Passo 1B	Indicando uma lacuna (<i>gap</i>) ou
	Passo 1C	Colocando questões ou
	Passo 1D	Continuando uma tradição
–	Movimento 3	Ocupando o nicho
	Passo 1A	Indicando propósito(s) ou
	Passo 1B	Descrivendo as principais características da pesquisa
	Passo 2	Indicando principais resultados
	Passo 3	Indicando a estrutura do artigo

2.3.2 O Modelo de Weissberg & Buker

Weissberg & Buker (1990) apresentam um modelo básico, contendo cinco estágios para modelar introduções. Seu modelo também foi derivado da análise de corpus contendo textos das áreas das ciências exatas, humanas e biológicas. Como pode ser observado no Quadro 2.5, cada estágio é novamente subdividido em passos. A numeração dos passos em sequência indica que eles podem ser empregados dessa forma, enquanto que a indicação dos passos com números e letras indica exclusão mútua.

Quadro 2.5. Modelo para a estrutura de introduções (Weissberg & Buker, 1990)

- **Primeiro Estágio:** Fornece um contexto para o problema a ser analisado
 1. Fatos relacionados à área geral da pesquisa
 2. Identificação da subárea
 3. Indicação do tópico
 - **Segundo Estágio:** Revisa aspectos do problema já estudados por outros autores
 1. Citações onde a informação é enfatizada
 2. Citações sobre o trabalho de vários autores
 3. Nível de atividade da pesquisa na área
 4. Citações onde o autor é enfatizado
 - **Terceiro Estágio:** Indica a necessidade de mais investigação na área
 - 1A. A literatura revisada é inadequada, ou
 - 1B. Conflito/Problema não resolvido: um *gap* na teoria ou na prática, ou
 - 1C. Questões ainda não consideradas na área, ou
 - 1D. Extensão/Melhoria do tópico
 - **Quarto Estágio:** Indica o propósito ou objetivo do estudo
 - 1A. Orientação ao artigo, ou
 - 1B. Orientação à pesquisa
 - **Quinto Estágio:** (Opcional) Valor/Justificativas da realização do estudo
-

2.3.3 O Modelo de Booth, Colomb e Willians

Booth, Colomb e Willians (2000) apresentam um modelo menos discutido do que o de Swales e o de Weissberg & Buker, porém bastante interessante, pois confirma os três movimentos básicos da introdução, os quais os autores chamam de pontos de vista. Para Booth et al. uma introdução pode ser vista como um conjunto de três pontos de vista chamados de *Base Comum*, *Ruptura* e *Resolução*, que aparecem quase sempre nessa ordem. O Quadro 2.6. mostra o refinamento desse modelo.

No modelo apresentado, **CONDIÇÃO** representa a má compreensão ou ignorância sobre um problema, isto é, uma lacuna no conhecimento, um conflito inexplicado ou uma discrepância, uma falta de conhecimento ou entendimento. O **CUSTO/BENEFÍCIO** de uma condição pode ser tanto os custos da ignorância da solução de tal problema como os benefícios de sua solução. A “proposição principal” enuncia a solução para a condição. A “proposição de lançamento” promete que a solução será apresentada no decorrer do texto. No entanto, tal proposição deve ir além de simplesmente introduzir o tópico. Ela deve sugerir os esboços conceituais da solução e anunciar um plano (metodologia utilizada). Nesse caso, a proposição principal é apresentada na Conclusão.

Quadro 2.6. Modelo para a estruturação de introduções (Booth et al., 2000)

1 – Base comum:	Tipos de Abertura <ul style="list-style-type: none"> ◦ Uma declaração geral. ◦ Um acontecimento ou caso. ◦ Uma citação ou fato estimulante. Contexto <ul style="list-style-type: none"> ◦ Compreensão compartilhada sobre o estado atual do problema ou antecedentes tidos como certos.
2 – Ruptura:	Objeção: <i>mas, contudo, por outro lado, etc.</i> Declaração do problema <ul style="list-style-type: none"> ◦ CONDIÇÃO de ignorância, pouca compreensão, etc. ◦ CUSTO/BENEFÍCIO de deixar a condição não resolvida ou de solucioná-la.
3 – Resolução:	Declaração da resposta <ul style="list-style-type: none"> ◦ Proposição principal ou proposição de lançamento

Embora a ordem apresentada no modelo seja comum, pode haver casos em que o escritor deseje alterá-la. Os autores fazem três sugestões:

- A *Base Comum* é opcional, e pode ser omitida em casos específicos;
- A *Ruptura* normalmente contém tanto o CUSTO quanto a CONDIÇÃO, mas, se os leitores estão familiarizados com o problema tratado, pode conter apenas um deles;
- A *Resolução* deve declarar uma proposição principal ou uma proposição de lançamento, de preferência a primeira;
- Um planejamento explícito do texto (*Outline*) pode ser colocado depois da proposição.

2.3.4 O Modelo de Aluísio

Aluísio definiu em sua tese de doutorado um esquema detalhado para a estruturação de introduções baseada em dois modelos tradicionais, o de Swales e o de Weissberg & Buker, e na análise de um corpus de introduções de artigos da área da Física Experimental. Como resultado da análise de corpus foi proposto um modelo de oito componentes subcategorizados em estratégias, que cobre todos os componentes dos modelos de Swales e de Weissberg & Buker e adiciona o componente “Metodologia”. Além disso, são deixadas explícitas as várias estratégias utilizadas pelos componentes para desenvolver suas mensagens características. O Quadro 2.7 apresenta o esquema detalhado para introduções da área da Física Experimental descrito em (Aluísio, 1995) e (Aluísio & Oliveira Jr., 1996).

Quadro 2.7. Esquema detalhado para introduções (Aluísio & Oliveira Jr., 1996)

Componente 1 – Contextualização

1. Introduzir o tópico de pesquisa a partir da área de pesquisa
2. Familiarizar termos ou objetos ou processos
3. Argumentar sobre a proeminência do tópico/área

Componente 2 – Revisão da literatura

1. Revisão histórica da área
2. Tendências atuais na área
3. Citações organizadas da área geral para o tópico
4. Progresso na área
5. Requisitos para o progresso na área
6. Estado da arte
7. Citações e *gaps* cíclicos
8. Citações agrupadas por abordagem

Componente 3 – Necessidade de mais investigação/ Tipos de *gaps*

1. Existência de conflitos ou problemas não resolvidos
2. Restrições em trabalhos anteriores
3. Questões ainda não consideradas

Componente 4 – Propósito da pesquisa

1. Indicar o propósito principal
 - 1A. Resolver um conflito entre autores
 - 1B. Apresentar uma nova abordagem ou metodologia ou técnica
 - 1C. Apresentar melhorias/avanços em um tópico da literatura
 - 1D. Apresentar uma extensão de um trabalho anterior do autor
 - 1E. Apresentar uma abordagem alternativa
 - 1F. Apresentar um trabalho comparativo
2. Especificar o propósito
3. Introduzir mais propósitos

Componente 5 – Metodologia

1. Indicar critérios ou condições
2. Descrever materiais e métodos
3. Justificar escolhas por métodos e materiais

Componente 6 – Principais Resultados

1. Apresentação de resultados
2. Comentários sobre os resultados

Componente 7 – Valor da Pesquisa

Componente 8 – Estrutura do trabalho

- 1A. Indicar as seções do artigo
 - 1B. Apresentar a lista dos tópicos a serem abordados
-

No modelo acima, temos estratégias enumeradas em cada componente. Aquelas que podem ser escolhidas mais de uma vez na escrita de uma introdução são identificadas somente com numerais, embora certas combinações possam não ser razoáveis. Os números seguidos de letras indicam exclusão mútua.

Segundo Aluísio (1995), o esquema proposto não impõe uma ordem convencional para os componentes, pois sua função é apresentar as informações que aparecem em uma introdução. Entretanto, é muito razoável que eles apareçam nessa ordem.

2.3.5 Comentários sobre os Modelos

Como se pode perceber, os modelos apresentados são coerentes entre si. Eles apresentam similaridades tanto em relação aos componentes quanto à ordem em que aparecem, embora não exista um mapeamento direto de um modelo para o outro. Apesar da similaridade, algumas particularidades podem ser observadas em cada modelo.

O modelo de Weissberg & Buker, por exemplo, dá uma ênfase maior à contextualização da grande área de trabalho do que os modelos de Swales e de Booth, Colomb e Willians. Esses autores focalizam mais o tópico de pesquisa, evitando uma grande contextualização. Esse padrão se mostra mais adequado no caso de introduções menores ou em introduções destinadas a um público-alvo mais homogêneo em termos de conhecimento da área. Quando o tamanho da introdução não for limitado ou o público alvo for heterogêneo, o modelo de Weissberg & Buker parece ser mais adequado por conter um estágio de contextualização mais abrangente.

O modelo de Aluísio deixa a cargo do escritor a decisão do tipo de estratégia que será usada em cada componente da introdução. No caso da contextualização, pode ser tanto uma estratégia mais abrangente como “Introduzir o tópico de pesquisa a partir da área de pesquisa” quanto uma estratégia mais restrita ao tópico como “Argumentar sobre a proeminência do tópico/área”. Um dos fatores que proporciona essa flexibilidade é o fato de o modelo não ter como objetivo impor uma estrutura, mas simplesmente apresentar os componentes típicos de uma introdução. Outro fator que contribui para que o modelo seja flexível é que ele engloba tanto as estratégias do modelo de Swales quanto as do modelo de Weissberg & Buker.

Um diferencial do modelo de Aluísio é a introdução do componente Metodologia, que não aparece explicitamente nos outros modelos estudados, embora alguns autores, como Huckin & Olsen (1991), sugiram que esse componente exista na introdução. Esse componente é mais característico das introduções de trabalhos que envolvem pesquisa experimental, em que os materiais e métodos utilizados são enfocados com uma relevância maior.

Quanto à ordem dos componentes, tanto Swales quanto Weissberg & Buker observam que existem outras formas de se organizar os componentes em uma introdução. Weissberg & Buker comentam que nem sempre os autores arranjam os estágios de uma introdução na

ordem exata proposta pelo modelo, mas que o plano por eles fornecido é comum e o mais fácil para escritores iniciantes.

2.3.6 Detalhando os Elementos da Introdução

Partindo do modelo proposto em Aluísio (1995), em que uma introdução pode ser estruturada a partir de oito componentes ordenados numa sequência “geral para específico” (Quadro 2.7), e das recomendações feitas em Weissberg & Buker (1990), chegamos a uma versão simplificada de um modelo de introdução dividido em oito estágios, conforme mostrado no quadro abaixo.

Quadro 2.8. Modelo simplificado para introduções

-
- | | |
|-------------------|---|
| Estágio 1. | Estabelece um contexto que ajuda os leitores a entenderem como a pesquisa se situa num campo de estudo maior |
| Estágio 2. | É feita uma revisão bibliográfica, ou seja, são apresentados aspectos do problema que já foram estudados por outros pesquisadores |
| Estágio 3. | Indica a necessidade de mais investigação na área |
| Estágio 4. | Indica os objetivos/propósitos do estudo |
| Estágio 5. | (opcional) Descreve a metodologia utilizada na realização do trabalho |
| Estágio 6. | (opcional) Descreve os principais resultados encontrados |
| Estágio 7. | (opcional) Dá uma justificativa para se empreender o estudo em questão, afirmando o valor do trabalho |
| Estágio 8. | (opcional) Define a estrutura do trabalho, isto é, seu <i>outline</i> |
-

Nas subseções seguintes, detalharemos cada um dos oito componentes citados, listando algumas diretrizes e observações importantes para que o componente realize sua função dentro da introdução de forma satisfatória. É importante destacar que grande parte das informações mais detalhadas foi retirada do trabalho de Weissberg & Buker (1990).

2.3.6.1 Primeiro Estágio: Contextualização

O contexto de uma introdução deve ser escrito de forma a fornecer aos seus leitores o *background* necessário para o entendimento do tópico particular da pesquisa em relação a uma área de estudo geral. Para que isso seja feito, pode-se começar com afirmações óbvias, amplamente aceitas, sobre a área na qual se está trabalhando. Então, passo a passo, o leitor deve ser guiado para ficar mais próximo do tópico de pesquisa específico do trabalho. Isso

pode ser feito em apenas algumas sentenças ou em vários parágrafos, dependendo do tipo de trabalho a ser escrito e do estilo de escrita do autor. Esse estágio pode ser escrito com os verbos no presente. A Figura 2.2 ilustra o movimento “geral para específico” seguido nesse estágio.

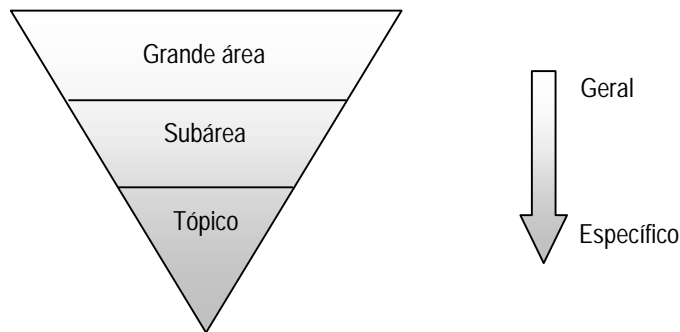


Figura 2.2. Movimento Geral-Específico seguido na seção de contextualização (Weissberg & Buker, 1990)

Weissberg & Buker (1990) apresentam uma analogia interessante, em que o estágio de contextualização é pensado como um processo em que primeiro se estabelece um “universo” para os leitores; então se isola uma “galáxia” dentro desse universo; e finalmente, guia-se os leitores para uma “estrela” na galáxia. Essa “estrela” é o tópico específico de pesquisa. Esse processo pode ser resumido da seguinte maneira:

- Comece com afirmações aceitas do fato relacionado à área geral de pesquisa (seu universo);
- Dentro da área geral, identifique uma subárea que inclua o tópico de pesquisa (sua galáxia);
- Indique o tópico (sua estrela).

Para que o leitor seja guiado suavemente através das idéias apresentadas nesse primeiro estágio, os escritores geralmente conectam as sentenças fazendo uso de informação conhecida e informação nova, alternadamente. Isso é feito colocando-se informação já conhecida pelo leitor no começo das sentenças e deixando a informação nova para o final da sentença. Veja no exemplo a seguir⁴, como a segunda sentença é conectada a primeira através desse mecanismo.

⁴ Todos os exemplos apresentados neste capítulo são trechos autênticos e foram extraídos do CorpusDT.

“Dentre as técnicas de verificação e validação, o teste é, sem dúvida, a atividade mais utilizada. Essa atividade [info. conhecida] apresenta-se bastante onerosa podendo, em alguns casos, consumir 40% dos custos de desenvolvimento do software”. [info. nova]

O exemplo a seguir exemplifica o movimento de contextualização descrito anteriormente. Repare como o leitor vai sendo guiado de um contexto bem geral para um contexto mais específico.

“O crescimento do mercado de software a cada dia acarreta o aumento do uso de técnicas de desenvolvimento, muitas vezes informais. [contexto mais geral] A manutenção de tais softwares torna-se problemática, uma vez que a documentação associada ao software, na maioria das vezes, não está de acordo com o código implementado. Dessa forma, quando diante da manutenção do produto, o engenheiro de software encontra uma documentação informal e incompleta, que não reflete o software existente. [contexto intermediário] Nesse contexto é que se encontra a Engenharia Reversa de Software, com o propósito de recuperar as informações de projeto perdidas durante a fase de desenvolvimento, e de documentar o real estado do software”. [contexto mais especializado]

2.3.6.2 Segundo Estágio: Revisão Bibliográfica

No primeiro estágio da introdução se estabelece um contexto para o tópico de pesquisa em questão. No segundo estágio se revê resultados encontrados por outros pesquisadores e que já tenham sido publicados na área de pesquisa de interesse. Por esta razão, o segundo estágio é comumente chamado de Revisão Bibliográfica. Esse estágio constitui-se essencialmente de uma coleção organizada de referências ou citações de outros trabalhos, os quais estão listados em uma seção separada no final do trabalho (normalmente intitulada *Referências Bibliográficas*).

O estágio de revisão bibliográfica serve basicamente a três funções importantes. Primeiro, esse estágio dá continuidade ao processo iniciado no primeiro estágio, de dar ao leitor informação de *background* necessária para entender o estudo em foco no trabalho. Segundo, ele garante ao leitor que o autor do trabalho está familiarizado com as pesquisas relevantes que têm sido realizadas na sua área de pesquisa. Terceiro, esse estágio estabelece o estudo relatado como um elo em uma corrente de pesquisa que está desenvolvendo e aumentando o conhecimento em um campo de atuação.

Em alguns tipos de trabalhos, é comum existir uma seção de revisão bibliográfica escrita separadamente da introdução (Barrass, 1979). No caso específico de teses e

dissertações, pode ser conveniente incluir um levantamento da literatura relevante anteriormente publicada como uma seção independente (um capítulo ou mais). Isso coloca a tese/dissertação em um quadro de referência e expõe o conhecimento existente que poderá servir de base para futuras investigações.

Tanto as referências bibliográficas apresentadas na introdução, como a seção “Revisão Bibliográfica” (se existir) precisam dar ao leitor o quadro indispensável em que se situa a investigação do trabalho sendo relatado e fornecer os elementos necessários para o desenvolvimento da argumentação. Essa revisão da literatura não deve ser uma seqüência impessoal de resumos de outros trabalhos; mas deve demonstrar que os trabalhos foram examinados e criticados objetivamente (Tachizawa & Mendes, 2000).

Seja na introdução, ou como um capítulo separado, a revisão bibliográfica será composta por um conjunto de citações de outros autores, de forma que é preciso atentar a dois fatores: o foco dado às citações e a ordem em que elas aparecem. A seguir, é discutido qual o enfoque que se deve dar a uma citação e como se pode ordenar um conjunto de citações de forma coerente.

Foco da Citação

Quando se cita trabalhos de outros autores, pode-se escolher focalizar a informação mostrada por aquele autor ou focalizar o próprio autor. O primeiro tipo de foco é chamado de “proeminente na informação”, porque a informação é apresentada com principal importância. O nome do autor e a data da publicação são então colocados entre parênteses no fim da sentença. Dessa forma, fontes de informação mais completas podem ser encontradas na lista ordenada alfabeticamente de referências, no final do trabalho. Um tipo alternativo de citação proeminente na informação usa números entre parênteses ao invés do nome do autor e da data. O número se refere à lista de referências enumerada no final do trabalho. Geralmente, citações proeminentes na informação são utilizadas no início da revisão bibliográfica e referem-se a fatos da área geral do estudo em questão aceitos por toda a comunidade científica. Essas citações podem também aparecer no primeiro estágio (Estabelecimento de Contexto). Nesses casos, pode-se usar o presente como tempo verbal. Veja o exemplo:

“Na engenharia reversa, o sistema geralmente é o ponto inicial do processo” (Chikofsky, 1990).

Conforme a revisão bibliográfica prossegue, as citações passam a se referir a estudos mais proximamente relacionados ao estudo em foco no trabalho. Nesse tipo de citação é dada

maior ênfase ao nome do autor. Nesse caso, o nome do autor passa a ser o foco da sentença, seguido pela data ou pelo número da citação entre parênteses, e então seguido pela informação em si. Esse tipo de citação é chamado de citação “proeminente no autor” e pode ser escrita usando-se o presente ou o passado. Em geral, quando se está reportando resultados obtidos por outros autores em seus trabalhos, utiliza-se o passado enquanto que, quando se reporta teorias, opiniões e recomendações, utiliza-se o presente. Em uma análise superficial do nosso corpus não foram encontradas ocorrências do uso do passado como tempo verbal, e isso deverá ser posteriormente investigado. O exemplo abaixo ilustra o uso de uma citação proeminente no autor feita no presente.

*“Rugaber (1992) **afirma** que a maior parte do esforço de desenvolvimento de software é gasto na manutenção de sistemas existentes e não no desenvolvimento de sistemas novos”.*

Ordem das Citações

O grupo de citações da revisão bibliográfica também deve ser ordenado em uma sequência lógica, visando facilitar o entendimento pelo leitor. Em Weissberg & Buker (1990) são sugeridas três estratégias para o arranjo das citações, sendo que cada estratégia se adapta melhor a um caso específico.

A estratégia mais comumente usada é a de arranjar as citações da revisão bibliográfica na ordem das mais indiretamente relacionadas para aquelas mais diretamente relacionadas ao estudo em questão, seguindo assim a estratégia utilizada no primeiro estágio que guia o leitor do geral para o específico.

Existem casos, porém, em que as outras estratégias são mais eficientes. Em uma revisão bibliográfica na qual se descreve a história da pesquisa em uma determinada área, o arranjo das citações em ordem cronológica pode ser mais indicado. Já no caso de existir um grande número de citações a serem incluídas em uma revisão bibliográfica, como em uma tese ou dissertação, por exemplo, pode-se agrupá-las de acordo com as diferentes abordagens ao problema tomadas pelos diferentes autores. As citações referentes a cada abordagem podem então ser ordenadas cronologicamente ou da geral para a específica, ficando a critério do autor.

2.3.6.3 Terceiro Estágio: Indicação da Lacuna

O terceiro estágio finaliza a revisão bibliográfica apontando uma “lacuna” (*gap*), isto é, uma área de pesquisa importante que não foi investigada por outros autores ou que não tenha sido suficientemente desenvolvida. Normalmente, esse estágio é completado em apenas uma ou duas sentenças e é escrito usando-se o presente como tempo verbal. Abaixo são apresentadas três alternativas para se escrever a(s) sentença(s) do terceiro estágio.

- Pode-se indicar que a literatura previamente estudada, descrita no segundo estágio, é inadequada porque um aspecto importante da área de pesquisa tem sido ignorado por outros autores, isto é, um determinado problema não está resolvido.
- Pode-se indicar que há um conflito não resolvido entre os autores dos estudos prévios relacionados ao tópico de pesquisa em foco, isto é, existe uma controvérsia. Essa controvérsia pode ser um desentendimento teórico ou metodológico.
- Pode-se indicar que o exame da literatura sugere uma extensão do tópico, ou levanta uma nova questão de pesquisa não considerada previamente por outros pesquisadores em seu campo de atuação.

Dessa forma, indicando algum tipo de falha deixada pelos estudos anteriores, o terceiro estágio prepara o leitor para focalizar o estudo em questão no trabalho, e de certa forma justifica a realização do estudo. O exemplo abaixo mostra como o escritor pode indicar a existência de uma lacuna.

“No entanto, existe pouca informação disponível sobre como escrever uma tese ou um artigo em português”.



Indicação da lacuna



Tópico em questão

Repare como é sinalizado no início da sentença que uma lacuna na informação dada até então será anunciada. Nesse exemplo, isso se dá pelo uso do *No entanto*. Outros sinalizadores podem ser utilizados, como os que aparecem no modelo de Booth, Colomb e Willian (*mas, contudo, por outro lado, etc.*)

2.3.6.4 Quarto Estágio: Apresentação do Propósito

O quarto estágio serve para mostrar, o mais concisamente possível, o objetivo específico do trabalho que está sendo escrito. Dessa forma, este estágio de anúncio do propósito segue diretamente o terceiro estágio, pois é a resposta à necessidade de pesquisa adicional na área de estudo em foco, expressa no estágio anterior.

A apresentação do propósito deve estar diretamente ligada à questão da pesquisa na qual está baseado o estudo. Apesar de não ser obrigatória a inserção explícita dessa questão, o propósito deve ser escrito de modo que o leitor possa inferir qual é a questão por trás do estudo sendo relatado.

Caso o trabalho seja um plano de pesquisa ou um trabalho em andamento, deve ser escrito usando-se o presente como tempo verbal. No caso de trabalhos concluídos pode-se usar o passado ou o presente, dependendo da orientação utilizada, conforme comentado a seguir.

A apresentação de propósito pode ser escrita seguindo-se uma de duas alternativas de orientação. As alternativas são:

- A orientação da apresentação do propósito pode ser dirigida ao próprio trabalho, isto é, pode se referir ao artigo, tese, dissertação ou relatório que vai comunicar a informação sobre a pesquisa em questão, conforme mostrado no exemplo abaixo. Note que o tempo verbal utilizado é o presente.

“O principal objetivo desta dissertação é o projeto e implementação de um mecanismo de tratamento de exceções para a construção de sistemas orientados a objetos confiáveis”.

[Orientado ao trabalho]

- A orientação do propósito pode ser dirigida à atividade de pesquisa, em outras palavras, ao próprio estudo em vez do trabalho escrito, como pode ser visto no exemplo abaixo. Note que o tempo verbal utilizado nesse caso é o passado.

“O principal objetivo deste trabalho de mestrado foi a investigação de uma estrutura adequada de hiperdocumento para apoiar a documentação requerida durante o processo de engenharia reversa de software”. [Orientado à pesquisa]

2.3.6.5 Quinto Estágio: Breve Descrição da Metodologia

Embora esse seja um estágio opcional, é recomendável que se diga logo na introdução qual(is) foi (foram) a(s) metodologia(s) usadas para o desenvolvimento do trabalho. No entanto, não se deve descrever toda a metodologia, pois isso deverá ser apresentado na seção Materiais e Métodos, onde a metodologia utilizada é descrita passo a passo. A citação da metodologia empregada no trabalho pode ser escrita usando-se o passado como tempo verbal, conforme exemplificado abaixo.

*“A engenharia reversa **foi** desenvolvida baseando-se no método de engenharia reversa Fusion-RE/I, e os produtos obtidos foram inseridos em uma hiperbase”.*

2.3.6.6 Sexto Estágio: Citação dos Principais Resultados Encontrados

Um trabalho científico não deve criar um suspense sobre os seus resultados, dessa forma é recomendável que se apresente na introdução os principais resultados alcançados, embora esse estágio seja visto como opcional por alguns autores. Essa descrição de resultados pode ser escrita usando-se o passado como tempo verbal, quando se tratar de trabalhos concluídos. Um exemplo de uma citação de resultados é mostrado abaixo.

*“A partir das análises realizadas **foi** estabelecido um conjunto de diretrizes que visam auxiliar o processo de desenvolvimento de uma simulação distribuída conservativa”.*

No caso específico de teses e dissertações, esse estágio pode incluir uma lista concisa das principais contribuições do trabalho para a ampliação do conhecimento da área investigada (Barrass, 1979).

2.3.6.7 Sétimo Estágio: Afirmação do Valor

O sétimo estágio é usado para justificar a pesquisa relatada, baseado em algum possível valor ou benefício que o estudo possa trazer para outros pesquisadores do campo de pesquisa ou para as pessoas que trabalham em situações práticas relacionadas a esse campo.

Essa afirmação de valor não é incluída em todas as introduções. Esse estágio deve aparecer quando se escreve a introdução de uma tese, dissertação ou uma proposta de tese. A afirmação do valor é também comumente incluída em relatórios de pesquisa escritos para

descrever um projeto conduzido com dinheiro proveniente de fontes externas. Em artigos escritos em jornais e revistas técnicas, por exemplo, esse estágio é quase sempre omitido.

Embora existam situações em que se deve colocar uma afirmação de valor, é convenção expressar certa modéstia por parte do autor, ou então não se mostrar tão seguro dos benefícios alcançados, mostrando uma certa cautela, pelo fato de que não se pode ter certeza de todas as futuras aplicações da pesquisa relatada. Esse estágio pode ser escrito usando-se o presente ou o futuro como tempo verbal.

Existem dois pontos de vista a partir dos quais a afirmação de valor pode ser escrita:

- A afirmação de valor pode ser escrita do ponto de vista dos benefícios práticos que podem resultar da aplicação dos resultados da pesquisa.
- Ou pode-se escrever a afirmação de valor para enfatizar a importância teórica do estudo no avanço do estado do conhecimento em uma área de pesquisa específica.

O exemplo abaixo mostra uma afirmação de valor orientada à prática. Não foram encontrados no corpus exemplos de afirmação de valor orientada à importância teórica. Acreditamos que isso se deve ao fato dos textos pertencerem a uma área científico-tecnológica (Ciências da Computação), onde é mais comum destacar a importância prática da pesquisa.

*“O sistema Animbs permite a associação de uma geometria ao MBS sendo simulado e utiliza os dados produzidos pela simulação do SD/FAST para criar uma animação do comportamento do sistema mecânico e, dessa forma, **melhorar a análise de dados feita pelos usuários do SD/FAST**”.* [Orientação Prática]

2.3.6.8 Oitavo Estágio: Estrutura do Trabalho (Outline)

No final da introdução, o autor pode dar ao leitor um roteiro do que será encontrado no restante do trabalho. Esse componente é chamado de *outline* e pode ser escrito no presente. Esse *outline* pode ser tanto uma indicação das seções do artigo (ou capítulos de uma tese/dissertação) como um apanhado dos assuntos abordados no trabalho na ordem em que aparecem. Abaixo segue como exemplo parte de um *outline* de uma dissertação.

“Este trabalho está organizado em 7 capítulos. No Capítulo 2 é feita uma revisão sobre sensores, detalhando os aspectos relacionados à robótica e as técnicas comumente empregadas para medição de distâncias. Uma revisão sobre fusão de sensores é feita no Capítulo 3. (...) No Capítulo 6 são descritos os experimentos realizados, analisados os resultados obtidos com cada algoritmo e comparados os diversos métodos aplicados para a fusão dos sensores. A conclusão deste trabalho é apresentada no Capítulo 7”.

2.3.7 Considerações sobre a Introdução

Conforme pôde ser visto até agora, escrever uma introdução interessante e que comunique ao leitor, de forma agradável, todas as informações que se espera encontrar em uma introdução não é uma tarefa simples, mas que pode ser facilitada se forem seguidas as diretrizes e estratégias apresentadas nos modelos propostos. Entretanto, os modelos não mencionam o que se deve evitar para que o objetivo da introdução não seja prejudicado.

Em Solomon (1977), encontramos alguns exemplos de erros cometidos em introduções e que foram observados pelo autor. São apresentados quatro tipos de erros comumente percebidos e que se deve evitar. São eles:

1. Introduções grandiloquentes, ambiciosas, que incluem intermináveis discursos, considerações marginais e lugares-comuns, como o de afirmar que o tema escolhido é “complexo, interessante e discutido”. É evidente que se não fosse, não valeria a pena ocupar-se dele.
2. Introduções históricas que remetem a questão a seus antecedentes remotos e se demora em sua descrição e análise.
3. Introduções exemplificadoras, em que se formulam exemplos ilustrativos do tema. Às vezes, em áreas como a medicina, se incorre nesse erro, incluindo casos desde a introdução do trabalho.
4. Introduções-soluções, em que se anunciam todos os resultados da investigação, cometendo-se assim duplo erro: psicológico, porque priva o leitor da motivação de encontrar a solução, seguindo o desenvolvimento do trabalho, e lógico, porque se o resultado já foi alcançado, tem pouco sentido o desenvolvimento e a argumentação.

É importante observar com cuidado o quarto tipo de erro apontado por Solomon. De fato, não se deve colocar na introdução toda a descrição dos resultados obtidos no trabalho. No entanto, Vieira (1999) esclarece que trabalho científico também não deve ser escrito como uma novela policial, que deixa o desfecho para o final. O interesse da literatura científica não

é estabelecer um clímax, mas dar ao leitor condições para julgar a qualidade da informação, reproduzir o trabalho e verificar se as conclusões são convincentes. Swales (1990) também argumenta que, em um trabalho científico, o objetivo deve ser esclarecido logo na introdução, sem criar suspense sobre os seus resultados.

2.4 Detalhando os Materiais e Métodos

Normalmente, depois da introdução, é apresentada uma segunda seção de texto chamada de Materiais e Métodos, principalmente em trabalhos que envolvem pesquisa experimental. Nessa seção são descritos os passos que foram seguidos para conduzir o estudo e os materiais que foram usados em cada passo. Essa seção é especialmente útil aos leitores que querem saber como a metodologia utilizada influenciou os resultados, ou aos leitores que estão interessados em replicar ou estender o trabalho descrito.

Segundo Barrass (1979), a seção de materiais e métodos deve incluir detalhes suficientes capazes de assegurar que a repetição da investigação por alguém com experiência na mesma área leve à obtenção de dados similares. Barrass também observa que quando o trabalho científico for uma tese, o autor deve indicar de que modo abordou o problema e qual foi sua contribuição para a resolução desse problema.

A parte principal da seção Materiais e Métodos é a descrição dos passos procedimentais tomados no estudo e dos materiais utilizados em cada passo. Entretanto, existem outros elementos comumente descritos nessa seção. Esses elementos não seguem um padrão fixo, assim como a ordem na qual eles aparecem também não é fixa, com exceção dos procedimentos e materiais, que sempre aparecem e sempre estão nessa ordem. Apesar dessa flexibilidade, existe uma ordem convencional que pode ser seguida e que pode ser um bom modelo, conforme mostrado no Quadro 2.9. Os elementos dessa seção são escritos usando-se como tempo verbal o passado ou o presente.

Embora todos os elementos do Quadro 2.9 possam estar presentes, o enfoque principal se encontra nos procedimentos e materiais. A descrição dos passos (procedimentos) que foram seguidos durante a execução do estudo deve ser escrita de forma clara para que o leitor possa repetir sem problemas os seus passos. Naturalmente, a melhor maneira de descrever um procedimento é passo-a-passo ou cronologicamente.

Quadro 2.9. Modelo de ordenação dos possíveis elementos da seção de materiais e métodos (Weissberg & Buker, 1990).

Possíveis Elementos
Overview do experimento (Projeto)
População / Amostra
Locação
Restrições / Condições Limites
Técnica de Amostragem
Procedimentos
Materiais
Variáveis
Tratamento Estatístico

A descrição dos procedimentos e dos materiais pode ser feita em conjunto ou separadamente. O autor pode escolher descrever os materiais e equipamentos utilizados juntamente com os passos em que esses materiais foram utilizados, alternando-se métodos e materiais. Por materiais entende-se qualquer item usado para dar andamento ao projeto de pesquisa. Weissberg & Buker apresentam uma categorização de materiais, como apresentado no Quadro 2.10.

Quadro 2.10. Categorias de materiais (Weissberg & Buker, 1990).

Materiais
Equipamento de laboratório
Equipamento de campo
Material humano ou animal
Substâncias naturais
Materiais fabricados
Pesquisas, questionários e testes
Modelos computacionais
Modelos matemáticos

Os materiais usados em um estudo também podem ser descritos de forma separada dos métodos. Um exemplo para o uso desse arranjo é quando várias partes diferentes de equipamento convencional de laboratório são usadas para executar um procedimento rotineiro. Mais comumente, no entanto, os materiais e os métodos são descritos em uma forma integrada, geralmente com ambos os elementos mencionados na mesma sentença.

Se os materiais que foram usados no estudo são conhecidos pelos pesquisadores no campo de pesquisa em questão, o convencional é apenas citar os nomes desses materiais para identificá-los. Entretanto, se no experimento forem usados materiais especialmente projetados

para o estudo ou não convencionais é comum se escrever uma descrição detalhada de cada material. Nesse caso, as seguintes informações devem ser incluídas:

1. *Overview*: este passo constitui-se de uma ou duas sentenças dando uma idéia geral do material e do seu propósito na pesquisa.
2. Descrição das partes principais: neste passo, cada parte ou característica principal do material é descrita em uma sequência lógica.
3. Descrição funcional: este último passo mostra como as várias características descritas no passo 2 interagem.

Para se fazer a descrição das principais características do material usado no estudo (Passo 2), Weissberg & Buker apresentam dois planos de organização que podem ser utilizados, dependendo do tipo de material. No Plano 1 é feito um “arranjo espacial”, isto é, descrevem-se as características de cima para baixo, da frente para trás, da esquerda para direita, do centro para fora, ou em alguma outra forma espacial. Esse arranjo é especialmente útil para se descrever equipamentos que consistem de várias partes conectadas. Uma outra forma (Plano 2) é se fazer um “arranjo funcional”, descrevendo as principais características do material na ordem em que elas funcionam, do início para o fim. Esse arranjo é o melhor para descrever partes que operam em uma sequência fixa.

2.5 Detalhando os Resultados

Como o próprio nome sugere, na seção Resultados são apresentados os resultados do estudo e um breve comentário sobre eles. Os resultados do trabalho devem ser apresentados numa ordem lógica – que pode ser diversa da ordem em que foi desenvolvida a investigação (Barrass, 1979). O importante é que o leitor consiga acompanhar a exposição dos dados e entender como os resultados foram alcançados.

Alguns autores chamam essa seção de Resultados e Discussão, indicando assim que são apresentados comentários mais extensos sobre os resultados. Outra forma é deixar tais comentários para próxima seção, chamada de Discussão, e na seção Resultados deixar apenas a apresentação dos resultados juntamente com alguns comentários breves. A seção Resultados deve ser escrita preferencialmente com os verbos no passado. As discussões sobre os resultados podem ser escritas usando-se o presente.

Na apresentação dos resultados é comum o uso de figuras, quadros, tabelas ou outros elementos gráficos seguidos de texto escrito, descrevendo os dados que estão representados. É

importante destacar que todos esses elementos gráficos devem sempre estar acompanhados de um texto explicativo, pois enquanto a demonstração dos resultados por meio do uso de figuras, gráficos e outros apresenta esses resultados de forma completa em termos matemáticos, o texto que os acompanha ajuda o leitor a focar sua atenção nos aspectos mais importantes dos resultados e a interpretá-los. Dessa forma, um não pode vir sem o outro. Outro ponto importante é que sempre deve haver no texto uma referência aos elementos gráficos, relacionando-os diretamente ao tópico/resultado específico que o elemento gráfico representa.

O texto que apresenta os resultados pode ser dividido em três elementos básicos de informação: (1) uma sentença localizando o(s) elemento(s) gráfico(s) onde os resultados podem ser vistos; (2) algumas sentenças que apresentam os aspectos mais importantes dos resultados e (3) outras sentenças que comentam esses resultados. Uma outra forma, mais compacta, de se escrever a apresentação dos resultados seria combinar os elementos (1) e (2). Assim, eles passam a ser apresentados como afirmações que apresentam os resultados mais importantes e indicam entre parênteses o elemento gráfico onde esses resultados podem ser vistos. O elemento (3) continua sendo o mesmo, ou seja, um comentário sobre os resultados.

Quanto ao comentário dos resultados, elemento (3), Weissberg & Buker (1990) sugerem duas formas de ordenação. Numa delas, colocam-se comentários curtos (uma ou duas sentenças) após cada resultado significativo que for mencionado. Outra forma é deixar para fazer todos os comentários no final, após terem sido apresentados todos os resultados. A primeira maneira é chamada “padrão alternado” e a segunda de “padrão seqüencial”. O padrão alternado é melhor quando se tem muitos resultados individuais com comentários específicos sobre cada resultado. O padrão seqüencial é usado quando existem muitos resultados individuais aos quais apenas um comentário geral se aplica. Em alguns casos, nenhum comentário é colocado, deixando para serem todos feitos na seção de discussão. Algumas funções podem ser atribuídas aos comentários dos resultados. As mais comuns são:

- generalizar os resultados;
- explicar possíveis razões para os resultados;
- comparar os resultados com resultados de outros estudos.

2.6 Detalhando a Conclusão

A seção Conclusões é a última do texto e, dessa forma, tem a função de finalizar o assunto. Durante a conclusão, a discussão deve ser uma consideração objetiva dos resultados

apresentados na seção anterior e deve conduzir com naturalidade às suas principais conclusões. A conclusão deve relacionar as diversas partes da argumentação e unir as idéias desenvolvidas anteriormente no trabalho. É por isso que se diz que, em certo sentido, a conclusão é uma volta à introdução, ficando no leitor a impressão de estar diante de um sistema harmônico, acabado em si mesmo (Solomon, 1977).

Assim como na Introdução, os pesquisadores usam essa seção para examinar seu trabalho no contexto maior do seu campo de estudo. No entanto, ao contrário da introdução, essa seção guia o leitor da informação específica apresentada nas seções de materiais e métodos e resultados para uma visão mais geral de como os resultados devem ser interpretados. Dessa forma, podemos dizer que na introdução parte-se do geral para o específico enquanto na discussão parte-se do específico para o geral (Figura 2.3).

A informação que se inclui nessa seção depende em grande parte dos resultados do estudo apresentado, mas o movimento “específico para geral” é uma convenção que a maioria dos escritores segue. Os tipos de informação que se pode incluir nessa seção não são fixos, porém os primeiros elementos são tipicamente aqueles que se referem mais diretamente ao estudo e aos resultados. Weissberg & Buker (1990) apresentam um modelo descrevendo os possíveis elementos de informação que aparecem no início da seção de conclusões, mostrado no Quadro 2.11.

Quadro 2.11. Possíveis passos para o início de uma conclusão (Weissberg & Buker, 1990)

1. Uma referência ao principal propósito e às hipóteses do estudo;
 2. Uma revisão dos resultados mais importantes, mesmo que esses resultados não suportem a hipótese original do estudo ou não concordem com os resultados de outros pesquisadores;
 3. Possíveis explicações sobre os resultados (resumidamente);
 4. Limitações do estudo que restringem a generalização dos resultados.
-

Conforme a seção prossegue, o escritor deve ir distanciando a atenção do leitor dos resultados específicos do estudo e começar a focalizar de forma mais generalizada a importância que o estudo pode ter para outros pesquisadores no campo de estudo, relacionando seu trabalho no contexto maior da sua área de pesquisa. Os últimos elementos de informação na seção de discussão são, portanto, afirmações gerais sobre o estudo, como sugerido por Weissberg & Buker (1990):

5. Implicações do estudo (generalização dos resultados);
6. Recomendações para pesquisas futuras e possíveis aplicações práticas.

É importante ressaltar que essa ordem de elementos da seção Conclusão não é estritamente seguida pelos autores, mas o movimento de elementos mais específicos para elementos mais gerais é convenção. Referências ao que foi feito no trabalho devem ser escritas com o verbo no passado, enquanto que referências a fatos devem ser escritas no presente. A Figura 2.3 apresenta um esquema do movimento específico para geral seguido na Conclusão.

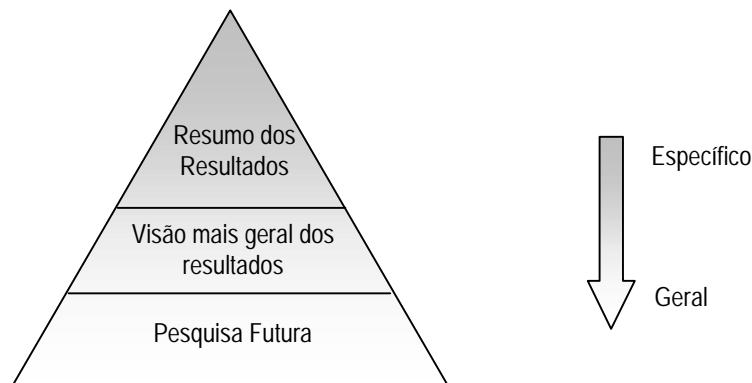


Figura 2.3. Movimento Específico-Geral seguido na Conclusão (Weissberg & Buker, 1990)

Nessa seção, mais do que em qualquer outro lugar do trabalho escrito, os escritores tornam explícitas suas próprias visões sobre o estudo e sobre os resultados. Os escritores podem tomar uma posição com respeito às explicações, implicações, limitações ou aplicações dos resultados. Quando isso acontece, é possível perceber partes de sentenças que expressam a opinião do escritor, conforme mostrado no exemplo abaixo. Note que, embora o autor não explicita o sujeito da sentença, fica clara a sua opinião sobre o tópico.

“Ao final desta tese conclui-se que aplicar corretamente o escalonamento de processos não é uma tarefa fácil”.

↑
Posição do autor

Além de desempenhar o papel de fecho de uma demonstração, a conclusão deve servir para abrir novos horizontes, para apontar caminhos, para despertar novas questões ou dúvidas, enfim, para inserir o trabalho no fluxo da busca científica que o transcende (Feitosa, 1991). Por isso algumas vezes usa-se o título “Conclusões e Trabalhos Futuros”.

2.7 Considerações Finais

Neste capítulo foi feita uma revisão bibliográfica sobre a estrutura de textos científicos. Embora muito pouco se tenha encontrado na literatura sobre a estrutura de textos em português, foram feitas adaptações de referências em inglês, uma vez que outros estudos indicam que a estrutura global dos textos científicos não varia em sua forma, e sim na sua organização na escrita em nível de parágrafo.

Dessa forma, foram apresentadas diretrizes encontradas para a escrita em português e também foi apresentada a estrutura global para textos científicos (Resumo, Introdução, Materiais e Métodos, Resultados, Conclusões e Referências). Segundo recomendações de Weissberg & Buker (1990), foram detalhadas cada uma das etapas desse plano global.

A realização dessa pesquisa serviu para mostrar, entre outras coisas, que a estrutura esquemática de textos acadêmicos em português é muito parecida com a estrutura usada em inglês. Não há como negar que existe uma variação no conteúdo devido às diferenças culturais e da própria língua, porém essas diferenças não se manifestam no nível da estrutura global quando se trata desse gênero de texto.

A realização desse levantamento bibliográfico serviu também para indicar o pouco desenvolvimento de pesquisas na área de estruturação de textos científicos em português, ao contrário do que acontece com textos em inglês, em que se têm modelos/esquemas bem definidos.

Capítulo 3

Ferramentas de Auxílio à Escrita

Um texto será considerado bom se ele servir bem ao seu propósito e estiver de acordo com as expectativas de sua audiência. Os textos científicos possuem ambos os aspectos bem definidos, ou seja, possuem um propósito bem específico – relatar/informar sobre uma determinada pesquisa – e uma audiência também bem específica/homogênea – a comunidade científica (Sharples & Pemberton, 1992).

Uma das expectativas de acadêmicos é encontrar textos claros, que possuam um determinado conjunto de componentes de texto tidos como essenciais para textos científicos. Por essa razão, o texto científico acaba por seguir um padrão rigoroso de estruturação. Com o intuito de esclarecer as particularidades desse gênero, vários autores procuram descrever os padrões encontrados e considerados bons, seja através de modelos que emergiram de análises (descrição) ou de diretrizes (prescrição), conforme foi apresentado no Capítulo 2.

O fato de o texto científico seguir um padrão de estruturação que pode ser reproduzido abriu caminhos para que fossem desenvolvidas ferramentas computacionais que apoiassem tanto o processo de composição do texto quanto o processo de revisão do texto produzido. No entanto, o processo de escrita, seja ela científica ou não, envolve várias etapas, como discutido por Hayes & Flower (1980), de modo que desenvolver um ambiente que suporte o processo como um todo de modo eficiente pode ser bastante complexo. Sendo assim, as ferramentas desenvolvidas até hoje enfocam alguns aspectos desse processo, geralmente

atacando um problema específico ou uma categoria de problemas que pode ser tratada em conjunto.

Grande parte das ferramentas desenvolvidas atualmente enfoca o pós-processamento do texto, isto é, trabalham sobre um texto já construído, geralmente um rascunho, com o objetivo de corrigir erros e melhorar aspectos como clareza e concisão. Nesse grupo incluem-se ferramentas como os corretores ortográficos, gramaticais e estilísticos. Sistemas de pós-processamento são extremamente úteis e de fato ajudam a melhorar a qualidade de um texto em vários aspectos. Entretanto, para se utilizar tais sistemas, o usuário deve ser capaz de produzir um primeiro rascunho. Swales (1990) afirma que mesmo a composição do primeiro rascunho de uma introdução não é fácil até para escritores experientes. Sendo assim, existe uma outra classe de ferramentas que visa apoiar o usuário no processo de composição e não apenas na revisão. Exemplos desses sistemas são o Writer's Assistant (Sharples & Pemberton, 1992; Sharples et al., 1994), o WE (Smith & Lansman, 1988; Lansman et al., 1993), o Composer (Pemberton et al., 1996; Shurville et al., 1997), o Academic Writer (Broady & Shurville, 2000), o Abstract Helper (Narita, 2000a; 2000b) e o AMADEUS (Oliveira Jr. et al., 1992; Fontana et al., 1993; Aluísio, 1995; Aluísio & Oliveira Jr., 1995; Aluísio & Oliveira Jr., 1996; Aluísio & Gantenbein, 1997; Aluísio et al., 2001).

Broady & Shurville (2000) argumentam que sistemas como os citados podem ser pensados como ferramentas cognitivas, uma vez que tentam minimizar a sobrecarga cognitiva imposta ao escritor no momento da escrita de um texto científico. Tal sobrecarga é atribuída ao esforço realizado para a integração dos diferentes níveis de exigências do processo de escrita (Sharples & Pemberton, 1992). Essas exigências podem ser resumidas em quatro classes: exigências “conceituais” (o que escrever e por que), “estruturais” (como as idéias devem ser organizadas), “de expressão” (geração e monitoramento da realização lingüística apropriada) e “estratégicas” (planejar, monitorar e avaliar a realização correta da escrita).

A seguir, apresentamos uma descrição das principais características de cada um dos sistemas, de modo a esclarecer os níveis de auxílio ao usuário e como isso é feito. Por fim, detalhamos o AMADEUS, uma vez que é objetivo deste projeto utilizar a tecnologia desenvolvida nesse projeto.

3.1 Writing Environment – WE

O WE é um sistema que, baseado no modelo de escrita de Hayes & Flowers (1980), tenta simular o processo cognitivo de criação de um texto, fornecendo recursos para que esse

processo seja realizado de maneira mais controlada, favorecendo assim o resultado final, ou seja, produzindo um texto melhor.

Sua principal característica é fornecer suporte para que o usuário faça uma espécie de *brainstorming* no momento do planejamento do texto, atuando como um organizador de idéias. Uma vez que o usuário tem o suporte do sistema para esse planejamento, diminuindo assim a sua sobrecarga cognitiva, espera-se que ele sinta-se mais livre para a geração e a organização das idéias.

O WE baseia-se num processo de composição dividido em 3 estágios: planejamento, escrita e revisão, sendo que o sistema possui módulos específicos para cada etapa do processo. No total, o WE possui quatro módulos, sendo que dois deles são dedicados ao planejamento do texto, um à edição e um à revisão. Esses módulos são respectivamente chamados de *Network Mode*, *Tree Mode*, *Edit Mode* e *Text Mode*.

O *Network Mode* permite que o usuário mapeie suas idéias em nós e faça as ligações entre elas, relacionando esses nós. Nesse módulo não há preocupação quanto à organização estrutural do texto, mas quanto ao relacionamento existente entre as idéias. Outro módulo do sistema se encarrega de organizar os nós criados no *Network Mode* numa forma hierarquizada, o *Tree Mode*. Nesse módulo é representada a organização estrutural dos nós, definindo os tópicos do texto. O *Tree Mode* funciona como um *outliner*, dando uma visão dos tópicos do texto como um todo. A transição do *Network Mode* para o *Tree Mode* não é automática, ou seja, é o usuário que monta a estrutura de tópicos, arrastando os nós de uma janela para outra.

Estando o texto planejado e organizado hierarquicamente, o usuário então preenche cada nó utilizando o módulo *Edit Mode*. Basicamente, esse módulo aborda o segundo estágio do processo de composição de um texto, a escrita. Nesse módulo o usuário vai preencher a estrutura de tópicos feita no *Tree Mode*, produzindo assim o primeiro rascunho do texto. Finalmente, no *Text Mode*, o usuário tem uma visão linear do texto, podendo, por exemplo, corrigir erros de coesão entre parágrafos. Com esse último módulo o sistema aborda a última etapa do processo de composição, isto é, a revisão do texto completo. A Figura 3.1 mostra como ficaria a tela do usuário enquanto estivesse usando o WE. O *Network Mode* é mostrado no lado superior esquerdo, com o *Tree Mode* logo abaixo. O *Edit Mode* é mostrado no lado superior direito e abaixo dele está o *Text Mode*.

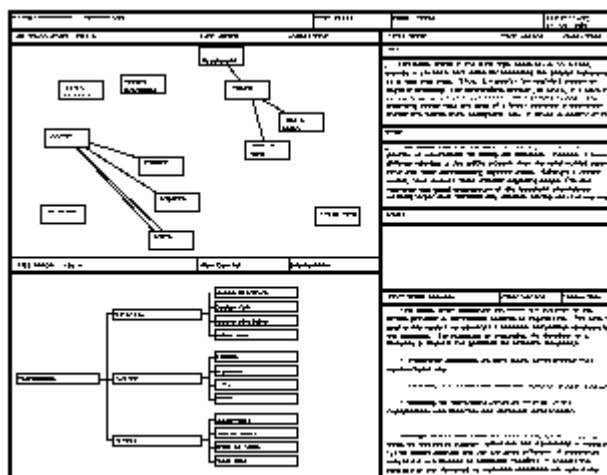


Figura 3.1. Dumping de tela mostrando os quatro módulos do WE (Lansman et al., 1993)

Em experimento realizado com o WE (Lansman et al., 1993), os usuários destacaram como principal ponto positivo o fato dos módulos serem apresentados em janelas separadas e paralelas, de modo que a pessoa pode escrever em uma janela enquanto visualiza o *outline* do documento em outra. Outra vantagem destacada foi a facilidade de se agrupar os nós para montar o *outline* no *Tree Mode*. Como principal desvantagem foi apontada a falta de recursos dos editores do *Edit Mode* e do *Text Mode*. Os usuários sentiram falta das facilidades de edição oferecidas pelos processadores de texto comerciais. Alguns usuários acharam os módulos *Network* e *Tree* redundantes, mas não souberam indicar qual dos dois é mais relevante e qual poderia ser removido.

3.2 Writer's Assistant

Assim como o WE, o Writer's Assistant também é baseado em modelos de escrita que consideram os aspectos cognitivos do processo. Dessa forma, este ambiente também é dividido em módulos que se especializam em cada etapa do processo, com o objetivo de diminuir a sobrecarga cognitiva imposta ao escritor.

O Writer's Assistant é um protótipo que combina um organizador de idéias, um editor de estrutura e um editor de texto, chamados de *Notes View*, *Structure View* e *Linear View*, respectivamente. No módulo *Notes View* o usuário pode colocar suas idéias na forma de nós e interligá-los através de arcos que especificam o tipo de relacionamento existente, formando assim uma rede de idéias. Trechos de texto podem ser anexados aos nós da rede já nesse módulo. Além dos nós estarem ligados entre si, alguns pontos da rede também são ligados a pontos que correspondem à estrutura geral do texto. Dessa forma, a rede de idéias já está diretamente relacionada à estrutura final do texto. A Figura 3.2 apresenta um *dumping* de tela

mostrando uma rede de idéias no *Notes View*. Note que dois pontos da rede da figura estão ligados à estrutura do texto (*body_element* e *ending*) através de arcos conectados à coluna que aparece do lado esquerdo da figura.

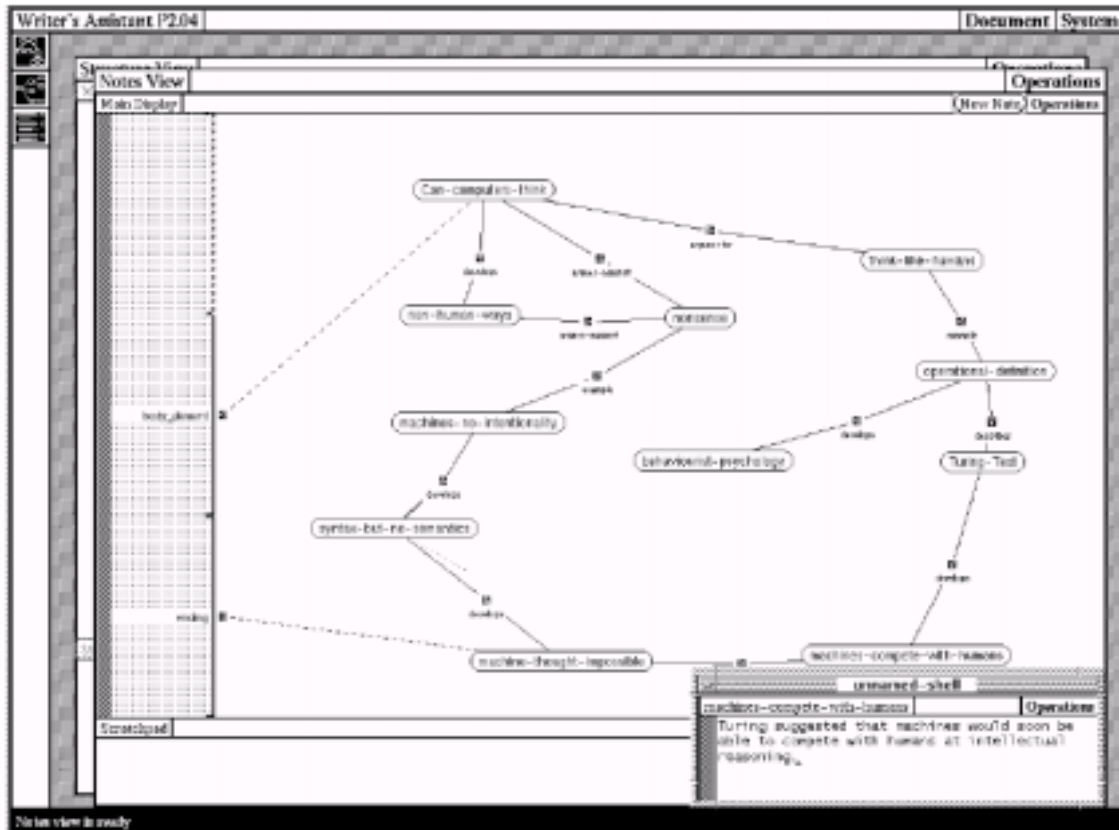


Figura 3.2. Dumping de tela do módulo *Notes View* do *Writer's Assistant*¹

O módulo *Structure View* apresenta uma visão da estrutura hierárquica da rede de idéias editada anteriormente, dando ao usuário a oportunidade de visualizar melhor a estrutura do texto. Uma visão do texto completo é apresentada no módulo *Linear View*.

A característica mais importante do *Writer's Assistant* é a facilidade de movimentação entre os três módulos do ambiente, isto é, entre as três visões de um texto. O ambiente utiliza um algoritmo que lineariza a rede de idéias construída pelo usuário criando automaticamente um primeiro rascunho do texto. Para a linearização é respeitada a prioridade dos tipos de ligações estabelecidas entre os nós e o posicionamento de partes da rede a uma ou outra posição do texto através da ligação feita com a estrutura (Sharples et al., 1994).

Em Sharples et al. (1994) é relatado um experimento feito para testar o algoritmo de linearização. O algoritmo foi aplicado a doze redes de idéias de tamanho médio (vinte nós).

¹ As Figuras 3.2 e 3.3 foram encontradas na página do projeto *Writer's Assistant*, disponível em <http://www.cogs.susx.ac.uk/users/mike/wa/wa.html> (Acessado em 10-03-2002)

Os textos linearizados foram avaliados quanto à organização e classificados numa escala de 5 pontos. A média obtida para os textos linearizados automaticamente foi de 3.5, ficando entre “algo organizado” e “aceitavelmente organizado”. Esse resultado demonstrou que o uso desse algoritmo é aceitável para a criação de um primeiro rascunho, embora melhorias ainda sejam necessárias. A Figura 3.3 mostra a janela do módulo *Linear View* apresentando um texto linearizado.

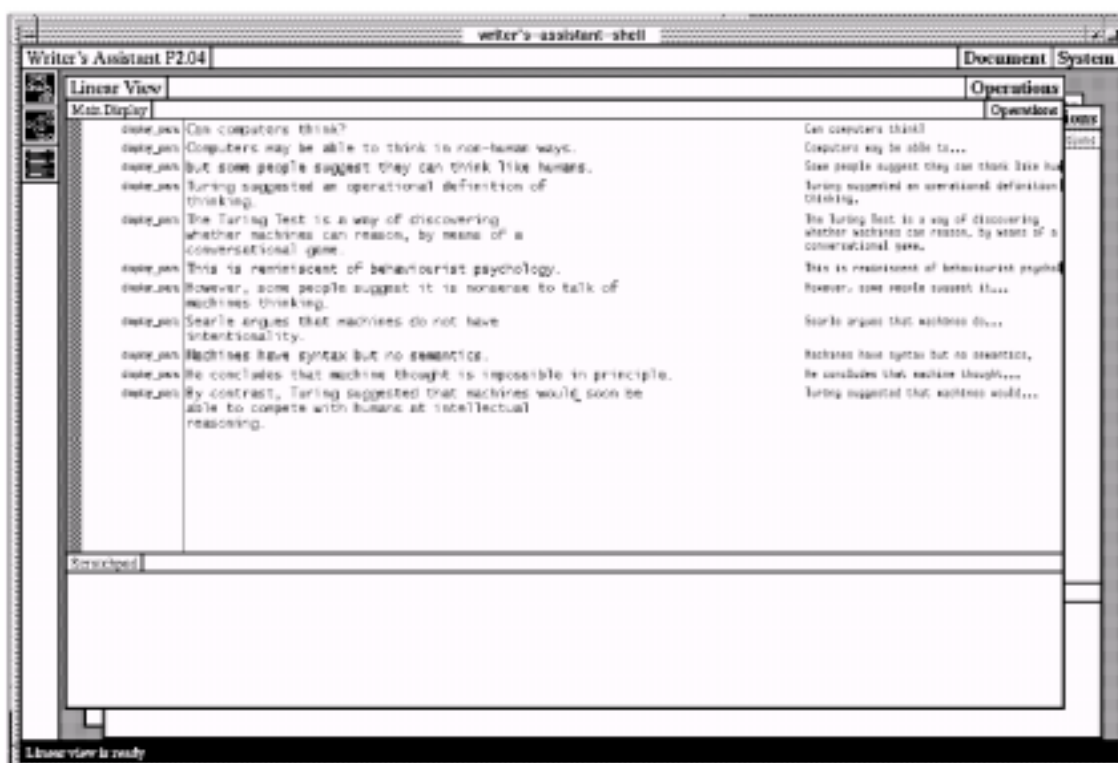


Figura 3.3. Dumping de tela do módulo *Linear View* do Writer's Assistant

3.3 Composer

O Composer é um projeto de ambiente de escrita técnica para não-nativos da língua inglesa que começou a ser desenvolvido na Universidade de Brighton, Inglaterra, mas que não chegou a ser finalizado. Esse projeto pretendia reunir em um único ambiente a abordagem de diferentes visões implementada no Writer's Assistant com estratégias mais específicas de apoio a composição de textos técnicos, focalizando em especial as necessidades de escritores não-nativos.

Embora o projeto não tenha sido finalizado, referências encontradas na literatura especializada (Pemberton et al., 1996; Shurville et al. 1997) descrevem os estudos

preliminares, a metodologia de desenvolvimento e parte da funcionalidade esperada, o que nos dá uma boa idéia do que o sistema viria a ser.

O Composer foi projetado para a escrita de tipos específicos de textos – relatórios de experimentos, artigos científicos e propostas de projetos. Para especificar o tipo de ajuda que o sistema daria, foi feita uma pesquisa junto aos estudantes não-nativos daquela universidade e aos seus professores. Essa pesquisa foi baseada em questionários e seu objetivo era descobrir os problemas particulares da escrita numa segunda língua e que poderiam ser minimizados com o suporte de um sistema computacional. Os resultados permitiram que fossem elicitados vários requisitos do sistema e também apontaram várias características que deveriam ser evitadas, tanto relacionadas à funcionalidade quanto ao estilo de interação.

A principal característica apontada como desejável na pesquisa quanto à funcionalidade foi que o sistema fornecesse modelos de textos *on-line*, com uma variedade de formas e de propósitos, para que os usuários pudessem reproduzir a estrutura e o estilo. Os usuários apontaram não só a necessidade de modelos estruturais, mas também a necessidade de instâncias de tais modelos. Embora se tenha demonstrado interesse por modelos pré-estabelecidos, foi apontado que não se deveria impor uma seqüência de passos rígida para o processo. Quanto ao estilo de interação, ficou clara a preferência por sistemas que auxiliam o processo e permitem ao usuário aprender a escrever melhor, ao invés de sistemas de pós-processamento.

Baseado nesses requisitos, o Composer foi projetado para ser um ambiente que trabalhasse as três visões do processo de escrita abordadas pelo Writer's Assistant e que também permitisse acesso *on-line* a esquemas estruturais específicos para cada propósito, diretrizes para a boa escrita relacionadas a cada modelo e textos-exemplo. O modo como essa visão estrutural viria a ser apresentada no Composer é exemplificado em (Pemberton et al., 1996) e reproduzido na Figura 3.4.

A Figura 3.4 representa um esquema estrutural para artigos científicos, mostrando as seções típicas desse tipo de texto e a ordem em que as seções aparecem. Além disso, a figura possui informações associadas em dois eixos, vertical e horizontal, que podem ser percebidas pelo formato dos elementos que representam cada seção. A extensão de cada elemento na horizontal representa o grau de generalidade das informações contidas naquela seção. Por exemplo, a Introdução (*Introduction*) deve começar com termos gerais e então, progressivamente, ir usando termos mais específicos, como é sugerido pelo formato da figura. No eixo vertical, estão indicadas a ordem dos elementos para esse gênero de texto e a extensão relativa que cada seção pode apresentar. O título de cada seção provê acesso a

menus que disponibilizam informações relativas à seção, como: informações sobre o papel da seção dentro do texto, textos bem escritos exemplificando a seção, principais erros cometidos quando se escreve tal seção e também conselhos sobre como refinar o texto. Um exemplo de menu é mostrado para a seção Resultados (*Results*).

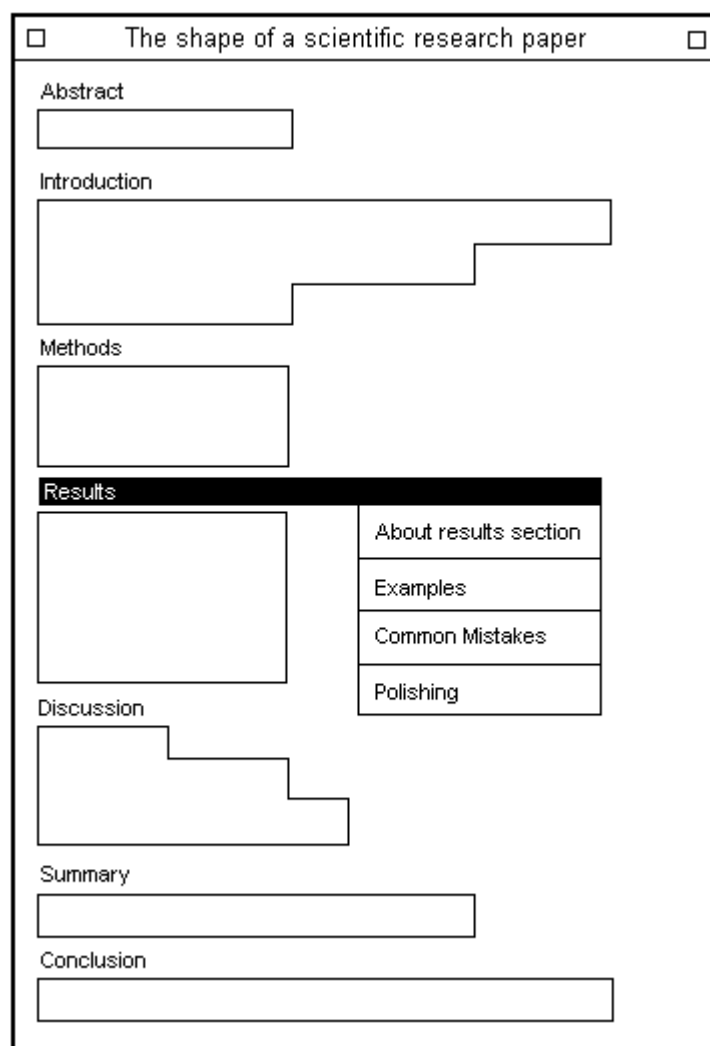


Figura 3.4. Ilustração da visualização de um esquema estrutural no Composer
(Pemberton et al., 1996)

3.4 Academic Writer

O Academic Writer é um ambiente de escrita *web-based* que nasceu do projeto Composer, e que ainda está em desenvolvimento. Seu objetivo é dar suporte a escritores iniciantes e/ou não-nativos em inglês acadêmico. Parte de seus requisitos foi herdada do projeto Composer, porém sua implementação foi simplificada e codificada em páginas Web.

Esse sistema é mais voltado para o ensino do gênero científico do que os sistemas apresentados anteriormente, pois grande parte da sua arquitetura é baseada em hipertextos explicativos do gênero. O sistema focaliza principalmente três aspectos da escrita científica: (1) a argumentação, encorajando um *brainstorming* estruturado que guia o aluno do tópico para o problema para a definição de propósito; (2) a estruturação, indicando a organização adequada do conteúdo; e (3) a realização lingüística, descrevendo as características estilísticas e lingüísticas da escrita acadêmica.

O Academic Writer trabalha com sete módulos, chamados de áreas, enfocando diferentes aspectos do processo de escrita: Estratégias, Gênero, Idéias, Foco, Estrutura, Revisão e Língua, sendo que cada área é implementada em uma página. Os recursos oferecidos pelo sistema são classificados como diretrizes, descrições e exemplos, e são organizados de forma a dar suporte ao entendimento de um determinado gênero (p.e. dissertação).



Figura 3.5. Menu principal do Academic Writer (Broady & Shurville, 2000)

O sistema também mantém um banco de dados onde são armazenadas as idéias de cada usuário (respostas às perguntas feitas pelo sistema) na área *Ideas* de modo que possam ser recuperadas posteriormente. Cada usuário mantém uma senha de acesso para proteger seus

dados. A Figura 3.5 apresenta a página que mostra o menu principal do Academic Writer, de onde se pode acessar qualquer uma das sete áreas. Cada área é representada por um ícone seguido de um texto explicativo indicando o tipo de ajuda que o usuário irá encontrar.

Na primeira área, relativa a *Strategies*, o usuário encontra cinco estratégias de escrita que sugere cinco formas diferentes de se acessar as páginas do Academic Writer. Para cada estratégia, o sistema sugere que as outras áreas sejam acessadas numa determinada ordem, porém essa ordem não é imposta.

Na área *Genre* o usuário seleciona o tipo de gênero e a área nos quais ele está escrevendo. Essa escolha irá determinar as informações que serão mostradas em outras áreas, como por exemplo, a área *Strategies*.

As áreas *Ideas* e *Focus* são independentes de gênero e focalizam a argumentação. Na área *Ideas* o usuário responde a uma lista de perguntas do sistema, chamada de *structured brainstorming*, com o intuito de revelar, através de suas respostas, seu tópico de pesquisa e o problema enfocado. A área de *Focus* faz o refinamento das idéias trabalhadas anteriormente, visando focalizar o propósito do trabalho que está sendo escrito e expressá-lo de maneira correta. As informações registradas nessas duas áreas são armazenadas no banco de dados para posterior recuperação.

Na área *Structure* são indicadas as possíveis seções que comporão o texto do gênero e área escolhidos. Essa escolha deve ter sido feita previamente na área *Genre*. Nessa área, referente à estrutura, o usuário tem acesso a descrições de cada seção, indicando, por exemplo, sua função dentro do texto como um todo. A partir das descrições o usuário pode então acessar exemplos de como cada seção pode ser realizada na prática.

Na área *Language* o usuário encontra um guia de referência na forma de um hipertexto, para características lingüísticas e estilísticas de textos científicos. Esse guia pode ser navegado de duas formas. Uma delas é por meio de *links* referentes a três máximas da escrita científica (“seja lógico”, “seja objetivo” e “seja aberto a dúvidas”) e duas características chaves do estilo acadêmico (“use marcadores discursivos” e “escreva de forma clara e simples”). A partir desses *links*, o usuário tem acesso às outras páginas que mostram como esses princípios podem ser realizados lingüisticamente. Outra forma de navegar pela área *Language* é por meio de um índice organizado de A-Z com os pontos relevantes da escrita acadêmica (p.e., o uso de tempos verbais), também ligados a exemplos de textos autênticos.

A área *Revision* ainda não está totalmente desenvolvida, mas o objetivo é ajudar o usuário a revisar seu texto oferecendo várias *checklists*. No momento, o sistema provê uma

checklist originada das respostas dadas nas áreas *Ideas* e *Focus* e das máximas relacionadas ao estilo acadêmico que aparecem na área *Language*.

Numa avaliação inicial, em experimento realizado com quatro estudantes em nível de mestrado na área de *Media-Assisted Language Teaching*, sendo dois nativos e dois não-nativos (um chinês e um espanhol), o Academic Writer se mostrou adequado a seus propósitos, isto é, os recursos oferecidos pelo sistema foram considerados proveitosos pelos avaliadores, entretanto, vários aspectos do sistema ainda necessitam de melhorias. Os resultados completos desse experimento são reportados em Broady & Shurville (2000).

3.5 Abstract Helper

O Abstract Helper é uma ferramenta para auxiliar a escrita de resumos em inglês (*abstracts*) por escritores não-nativos, em especial os japoneses. Ao contrário das outras ferramentas apresentadas neste capítulo, o Abstract Helper não se preocupa com o processo de geração e organização de idéias, concentrando-se na estruturação e na realização lingüística do *abstract*. A motivação para construir esse sistema, segundo Narita (2000a), foi tentar solucionar o problema de escrita de textos científicos em inglês enfrentado por escritores japoneses, proveniente não só das diferenças de idioma, mas também da própria organização do texto.

A ferramenta implementa a abordagem de reutilização de textos autênticos como modelos organizacionais e estilísticos para a escrita de textos novos, tendo em vista três objetivos: disponibilizar um corpus paralelo Inglês/Japonês, anotado, de exemplos de *abstracts* bem organizados; permitir ao usuário encontrar um bom modelo para seu *abstract* ou sua sentença; permitir que o usuário tenha acesso rápido aos recursos lingüísticos para conseguir informação relevante no contexto do seu texto.

Para atender os objetivos, foi coletado um corpus de 539 *abstracts* provenientes de publicações relevantes na área da Ciência da Computação, considerados bons exemplos de organização e estilo. Em seguida, foi preparada a tradução para o japonês de cada *abstract*, visando facilitar o acesso pelos usuários. Os *abstracts* foram traduzidos sentença a sentença e, posteriormente, foram alinhados manualmente.

Esse corpus paralelo foi anotado em SGML utilizando um conjunto de etiquetas-padrão, em dois níveis: *abstract* como um todo e sentença. No primeiro nível foram marcados o tipo do abstract (p.e. proposta de um novo sistema/modelo/algoritmo) e o esquema organizacional do abstract, que foi classificado em quatro tipos: (1) *abstracts* que iniciam com a sentença tópico; (2) *abstracts* com a sentença tópico no meio; (3) *abstracts* que terminam

com a sentença tópico; (4) *abstracts* multi parágrafos. No nível sentencial, foram marcados os papéis de cada sentença, sendo que foram definidos sete papéis: (1) sentença introdutória; (2) sentença tópico; (3) sentença explanatória; (4) sentença de verificação; (5) sentença suplementar; (6) sentença de conclusão; (7) sentença de fechamento. Além dessa marcação em dois níveis, o corpus foi anotado com algumas informações relativas aos verbos utilizados, que foram agrupadas em uma base de dados lexicais, juntamente com as frequências de ocorrência no corpus. Outro recurso disponível nesse sistema é uma base de colocações em inglês extraídas do corpus e checadas manualmente.

O Abstract Helper pode ser visto como um grande repositório de recursos lingüísticos que podem ser acessados através de quatro tipos diferentes de busca: por *abstract*, por sentença, por padrão de sentença e por colocação. Não existe uma ordem pré-estabelecida para o acesso aos recursos, porém é esperado que o usuário acesse o sistema num procedimento *top-down*: buscando primeiramente um bom modelo de *abstract*, em seguida buscando por exemplos de sentenças e, finalmente, buscando informações sintáticas e léxicas.

Narita (2000a) reporta um experimento feito com usuários acadêmicos e o resultado mostrou uma atitude positiva dos avaliadores em relação ao sistema. Entretanto, algumas melhorias foram requisitadas como diversificação do domínio do corpus, maior número de exemplos e produção de uma versão do sistema baseada em janelas.

3.6 AMADEUS

O AMADEUS² é um ambiente de auxílio e ensino da escrita técnica em inglês, voltado para escritores não-nativos que sofrem influência da língua materna quando escrevem em inglês. É composto por quatro ferramentas inter-relacionadas: ferramenta de referência, ferramenta de suporte, ferramenta de crítica e ferramenta tutorial, sendo que as três primeiras já foram implementadas e a quarta está em fase de projeto.

O AMADEUS foi fundamentado nos estudos de alunos brasileiros em pós-graduação no exterior feito por Fontana et al. (1993) cujos resultados mostraram que algumas deficiências na escrita dos alunos estão relacionadas ao mau uso ou omissão de expressões mais ou menos convencionais que desempenham funções específicas no texto científico. Uma solução para esse problema, e que foi implementada no AMADEUS, é a reutilização de expressões padrões utilizadas em textos reconhecidos como bons para que, no caso de dúvida no momento da escrita, o escritor tenha uma base de bons exemplos a qual ele possa recorrer.

² *Amiable Article Development for User Support*

As três ferramentas atualmente implementadas no ambiente AMADEUS trabalham com essa abordagem, embora cada uma focalize um nível diferente de auxílio. A Tabela 3.1 apresenta o relacionamento entre as características do usuário relativas ao conhecimento de inglês e de escrita técnica e as quatro ferramentas previstas para o ambiente.

Tabela 3.1. Adequabilidade do tipo de ferramenta quanto ao conhecimento do usuário
(Barros, 2000)

Inglês	Escrita Técnica	Boa Experiência	Alguma Experiência	Nenhuma Experiência
Boa recepção		<i>Ferramenta de Referência</i>		
Problemas de coesão em nível de parágrafo			<i>Ferramenta de Suporte</i>	
Problemas na escrita para um propósito e audiência específicos			<i>Ferramenta de Crítica</i>	
Problemas de coesão em vários níveis				<i>Ferramenta Tutorial</i>

As ferramentas de Referência e Suporte foram desenvolvidas como parte do trabalho de doutorado de Aluísio (1995) e foram implementadas para o domínio da Física Experimental. A ferramenta de Crítica foi desenvolvida durante o trabalho de mestrado de Silva (1999) e trabalha com o domínio específico da comunidade de HCI³, com artigos publicados na CHI'96⁴.

As seções seguintes detalham cada uma das três ferramentas do ambiente, indicando contexto de uso, formalismos computacionais utilizados, recursos lingüísticos acessados e outros detalhes de implementação.

3.6.1 Ferramenta de Referência

Como o próprio nome sugere, essa ferramenta tem como objetivo servir de referência para o escritor não-nativo, provendo uma base de expressões padrões que podem ser utilizadas no contexto da escrita científica, nesse caso específico, artigos científicos. Estudos como os descritos no Capítulo 2 desta monografia indicam a existência de padrões que podem ser reproduzidos, mesmo no nível de sentenças, de forma que um número relativamente pequeno de padrões pode ser suficiente para cobrir uma grande proporção das expressões usadas em uma área de conhecimento específica. Para tal, foi compilada uma base de expressões

³ Human-Computer Interaction

⁴ Conference on Human Factors in Computing Systems realizada em 1996.

autênticas de artigos científicos da física, através da análise de corpus, categorizando essas expressões de acordo com os componentes e subcomponentes da estrutura esquemática de um artigo. As expressões padrões são apresentadas juntamente com lacunas, que devem ser preenchidas com o material factual da pesquisa do escritor.

Quadro 3.1. Exemplos de expressões padrões para introduções (Aluísio, 1995)

- a) *Importance of the field, general interests, etc.*
There has been substantial interest in the fabrication of ...
 - b) *Description of an effect, phenomenon, etc.*
The phenomenon of ... induced by ... has not only provided a sensitive and convenient probe for monitoring ... (membrane breakdown) but has also revealed the irreversible changes that can occur during ...
 - c) *Previous reports on related work.*
Several papers have reported measurements aimed at obtaining evidence for, and insight into, ... processes in ...
 - d) *What is lacking in the field.*
Although significant advances have been made in the understanding of how ... (something) influences ... (another), very little further attention appears to have been given to the ...
 - e) *Difficulties faced in a particular analysis*
Further difficulties arise from the limited ... available and the requirement for a ...
 - f) *Improvements on previous works, models, etc.*
This is an improvement over the existing practice of simply inserting (calculating, determining, etc) ...
 - g) *What the present work does.*
The purpose of the work reported here was to study the influence of ... on the ...
 - h) *Relevance of this work to the field or other areas*
The surface properties of ... apart from the pure physical chemical interest will help to elucidate the role of ...in many ... phenomena.
 - i) *Layout or Outline of the paper*
The organisation (outline) of the (this) paper is as follows. In Section II we describe ... The ... is presented in section III. In Sec. II we solve the ... equation giving expressions for ... This is necessary for the work of Sec. III, in which the extended ... equation is derived. Numerical results of the theory are given in Sec. IV, together with a comparison with ... and ... calculations.
-

Note que as expressões apresentadas no Quadro 3.1 variam em tamanho, de períodos simples (entrada a) e compostos (entrada d) a vários períodos (entrada i). A idéia subjacente é fornecer, para alguns casos, os conectivos lógicos necessários para a conexão entre períodos. Além de apresentar a realização adequada de certas funções retóricas, as expressões fornecem também substantivos, adjetivos e verbos adequados à descrição de processos, comparação de dados e figuras e outros. Essa característica sugere que a base de expressões seja específica a cada área de pesquisa/conhecimento.

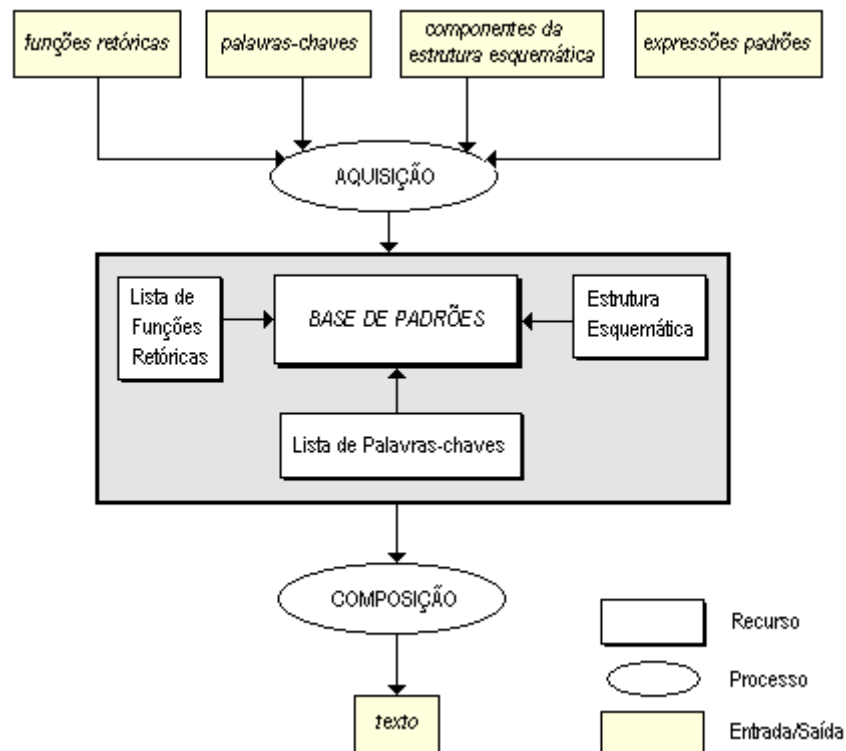


Figura 3.6. Arquitetura da ferramenta de Referência (Aluísio, 1995)

A Figura 3.6 ilustra a arquitetura da ferramenta de Referência, enfatizando dois processos principais, composição e aquisição, e os recursos disponibilizados pela ferramenta. A ferramenta utiliza quatro recursos/bases:

1. a base de expressões padrões, chamada de Base de Padrões;
2. uma lista de componentes e subcomponentes da estrutura esquemática para artigos proposta por Deyes (1982)⁵, com ponteiros para a Base de Padrões;
3. uma lista de funções retóricas (p.e. definições, exemplificações, etc) que também apontam para as entradas correspondentes na Base de Padrões;
4. uma lista de palavras-chave características do gênero (p.e. *report*, *paper*, *model*, *technique*, etc) que permite acesso rápido às entradas da Base de Padrões através de uma busca simples que retornará todas as expressões contendo a palavra-chave pesquisada.

Sendo assim, durante o processo de composição o usuário pode acessar a Base de Padrões de três formas: a partir da estrutura esquemática, da lista de funções retóricas e da lista de palavras-chave. O processo de aquisição diz respeito à inserção de novo material lingüístico à ferramenta, isto é, inserção de novas expressões padrões, funções retóricas,

⁵ Deyes, T. (1982). Discourse, Science and Scientific Discourse (The Raw Material of Comprehension in ESP). Working Paper 6, Brazilian ESP Project, PUC/SP, São Paulo *apud* (Aluísio, 1995).

componentes da estrutura esquemática e palavras-chaves. Assim o usuário pode personalizar a ferramenta, inserindo o material lingüístico que desejar.

Testes realizados com a ferramenta de Referência junto a alunos de pós-graduação mostraram uma avaliação positiva. Algumas características favoráveis apontadas foram a familiarização com a estrutura retórica de artigos científicos e melhorias na organização do texto. A ferramenta também se mostrou útil para superar o problema do bloqueio inicial com relação a escrita em uma língua não materna.

É importante destacar que a ferramenta de Referência é voltada para escritores não-nativos que possuem um bom conhecimento tanto da língua inglesa quanto do gênero científico (conforme apontado na Tabela 3.1), uma vez que a ferramenta trabalha com construções gramaticalmente sofisticadas e não oferece nenhum auxílio direto para a organização no nível de parágrafo ou de seção. Esse tipo de auxílio é oferecido pela ferramenta de Suporte, que é apresentada a seguir.

3.6.2 Ferramenta de Suporte

A ferramenta de Suporte é voltada para usuários que, embora tenham um conhecimento razoável da língua inglesa, não possuem muita experiência em escrita científica, encontrando dificuldades em organizar seu texto de forma adequada ao gênero, produzindo textos sem coesão. Essa ferramenta é destinada à escrita de introduções de artigos curtos (*letters*) da área da Física experimental e utiliza o método de resolução de problemas conhecido como Raciocínio Baseado em Casos (RBC) (Mantaras, 1995).

Visando diminuir o problema da falta de coesão dos textos escritos em inglês por não-nativos, Aluísio (1995) propôs uma abordagem baseada em casos para a construção de uma ferramenta para a escrita de introduções que utiliza dois mecanismos principais: instanciamento de um esquema detalhado de introdução com material lingüístico de textos autênticos e aplicação de adaptações interativas que flexibilizam as opções do esquema instanciado. O primeiro mecanismo fornece a estrutura retórica do texto real para que o escritor veja como os mecanismos de coesão são expressos na língua. O segundo adapta essa estrutura às necessidades atuais do novo escritor.

A abordagem proposta pode ser resumida em três princípios:

1. pré-compilação do conhecimento do gênero em um esquema detalhado que é utilizado para mapear introduções em casos e como fonte de conhecimento na fase de adaptação dos casos;

2. reutilização do material lingüístico não factual dos textos reais, ajudando a aumentar a fluência dos textos e servindo de fonte de conhecimento na fase de adaptação; e
3. utilização de raciocínio baseado em casos, RBC, como modelo. Esse princípio dita as fases da interação e os recursos utilizados, que são:
 - Uma Base de Casos, gerada através da instanciação do esquema detalhado do texto com material lingüístico de textos autênticos, sendo que o índice de cada caso é a sua própria estrutura retórica;
 - Métricas de Similaridade, utilizadas para recuperar os melhores casos dada a estrutura retórica do texto a ser escrito;
 - Regras de Revisão, utilizadas na adaptação interativa de um caso autêntico para outras necessidades.

A ferramenta de Suporte é um protótipo que implementa tal abordagem. Ela foi originalmente implementada em Sicstus Prolog com a interface em Xview ToolKit e Netscape Browser, mas posteriormente foi migrada para o ambiente Windows. A Figura 3.7 apresenta a arquitetura da ferramenta, destacando os processos e recursos utilizados.

Como pode ser observado na figura, existem três fontes de conhecimento na ferramenta:

- a **Base de Casos**,
- a **Base de Regras de Revisão** e
- as **Regras e Medidas de Similaridade**.

A **Base de Casos** foi obtida do processo de análise de corpus de introduções de artigos da área da Física Experimental. Essa base contém cinquenta e quatro instâncias de estruturas retóricas de introduções autênticas, descrevendo os componentes do caso e as estratégias dos componentes. Por sua vez, cada estratégia contém uma lista de mensagens associadas. Cada predicado Prolog (entrada de caso) tem associado um arquivo com o texto da introdução anotado com suas características retóricas (componentes e estratégias). É a partir desse repositório que os casos são recuperados.

A **Base de Regras de Revisão** também foi obtida em parte (estrutura esquemática) da análise de corpus e é composta por regras de adaptações genéricas, pois servem para adaptar qualquer uma das trinta estratégias e suas quarenta e cinco mensagens associadas. São quatro os tipos de adaptações possíveis: inserção de mensagens, remoção de mensagens, mudança de material lingüístico da mensagem e troca de mensagem/estratégia por uma similar.

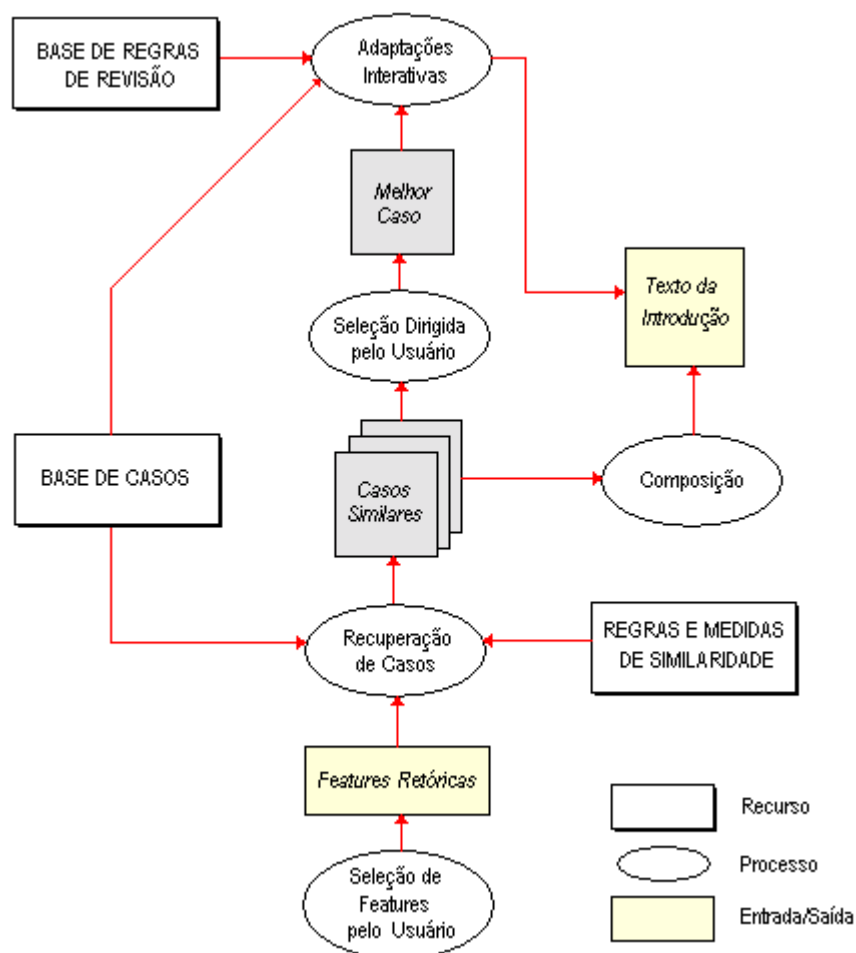


Figura 3.7. Arquitetura da ferramenta de Suporte (Aluísio, 1995)

As **Regras e Medidas de Similaridade** são baseadas em regras de similaridade entre listas (casamento de padrões) e na medida de similaridade conceitual chamada *nearest neighbors matching* (Kriegsman & Barletta, 1993). Essas regras são utilizadas no processo de recuperação dos casos, realizando a busca de acordo com a requisição (estrutura retórica) feita pelo usuário.

Os processos representados na Figura 3.7 mostram como o usuário interage com a ferramenta. Através do processo *Seleção de Features pelo Usuário* o usuário monta a sua requisição, que alimentará o processo *Recuperação de Casos* que é realizado pela ferramenta. Esse processo toma as *Features Retóricas* fornecidas (requisição do usuário), as estruturas retóricas de cada caso da Base de Casos e as Regras e Medidas de Similaridade e devolve ao usuário os casos mais próximos a sua requisição. Tendo os melhores casos recuperados, o usuário escolhe o melhor caso através do processo *Seleção Dirigida pelo Usuário* e assim pode iniciar as *Adaptações Interativas*. Através das adaptações e da entrada de material lingüístico relativo a sua pesquisa o usuário compõe sua introdução. É importante ressaltar que o usuário é livre para escolher se quer fazer as adaptações interativas ou não, podendo

específicas foi chamada de Estrutura Dual (Silva et al., 1998; Silva, 1999). A Figura 3.8 apresenta a arquitetura da ferramenta, destacando os processos (realizados pelo crítico e pelo usuário) e os recursos utilizados.

Como está ilustrado na figura, o processo de crítica inicia-se assim que o usuário apresenta um produto para a ferramenta. Para criticá-lo, a ferramenta deve obter o(s) objetivo(s) do usuário. Isto pode ser feito através de dois processos mutuamente exclusivos: *Reconhecimento de Objetivos*, que reconhece o objetivo do usuário através das escolhas utilizadas no produto, ou *Aquisição de Objetivos* que utiliza informações explícitas fornecidas pelo usuário.

Tendo os objetivos definidos, o modelo utiliza dois tipos de analisadores para avaliar o produto: *Analítico* e *Diferencial*. O *Analizador Analítico* faz uma análise do produto baseado em um conjunto de *guidelines* (regras heurísticas que ajudam o escritor a adequar seu artigo à comunidade em questão, nesse caso a comunidade da CHI) enquanto que o *Analizador Diferencial* utiliza um caso recuperado da Base de Casos através da abordagem de Raciocínio Baseado em Casos (RBC) como modelo para ser comparado com o produto do usuário. O módulo que faz a recuperação dos casos segue as especificações do trabalho de Aluísio (1995) e recupera o caso mais similar ao produto do usuário. O *Analizador Analítico* é responsável por checar o conteúdo (quais componentes devem estar presentes na introdução) e o *Diferencial* a organização (a ordem mais provável desses componentes, se alguns dos componentes podem ser opcionais, ou aparecer mais de uma vez).

A saída final dos analisadores são *fatos* a respeito das diferenças entre o produto do usuário e o caso recuperado, e da análise do produto em relação às *guidelines*. Esses fatos são reportados ao usuário da forma mais adequada possível, tendo como base as Estratégias de Críticas (p.e. crítica ativa, crítica passiva) e as Formas de Crítica (elogios, críticas diretas, indiretas, sugestões diretas e instruções e sugestões impessoais para melhorias). Com essas informações, o usuário pode então gerar uma nova versão do produto e iniciar novamente o processo.

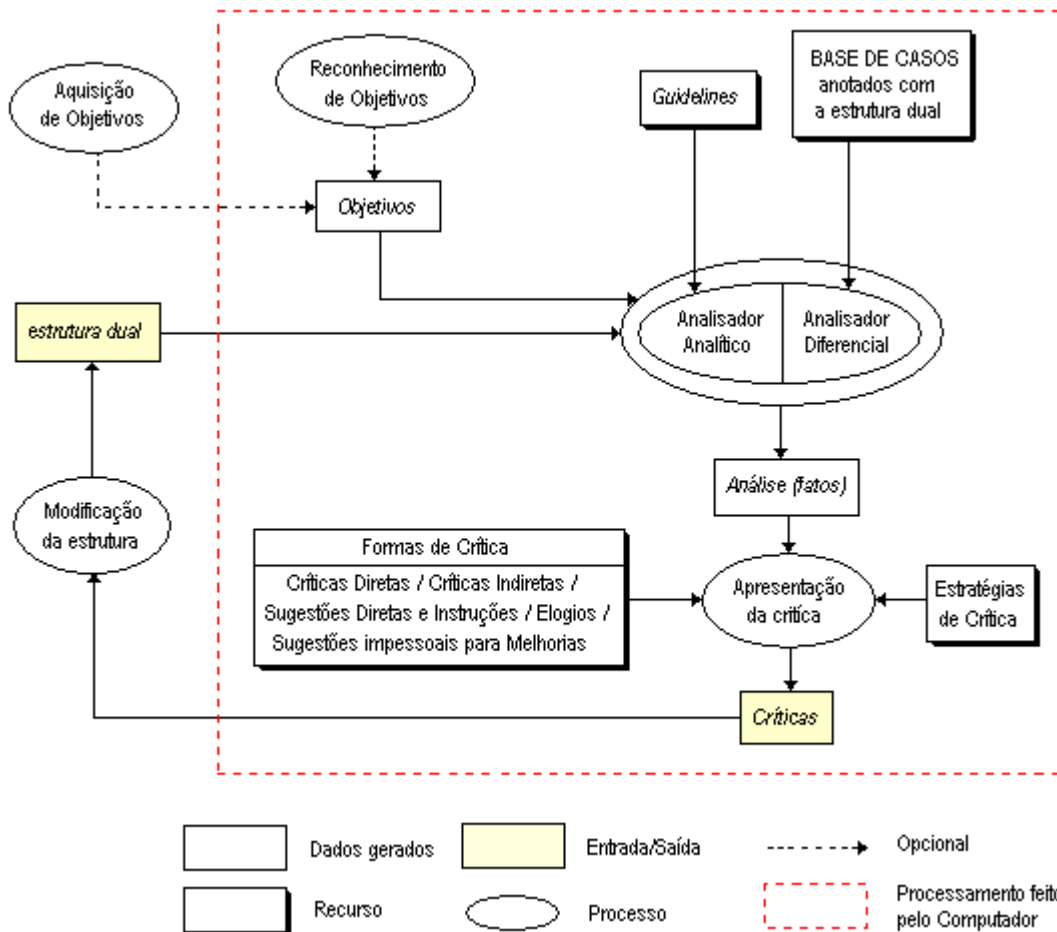


Figura 3.8. A arquitetura da Ferramenta de Crítica (Silva, 1999).

A ferramenta de Crítica foi avaliada em dois aspectos: operacionalidade e praticidade. Para testar a operacionalidade do sistema foram realizadas simulações com três classes de usuários: principiante, intermediário e especialista. Os usuários das três classes se beneficiaram da ferramenta, pois através da interação com as críticas fornecidas pelo sistema produziram estruturas melhores do que as produzidas inicialmente. Para a análise da praticidade foram observados quatro aspectos: (i) Extensibilidade de *guidelines*, de Casos e Estruturas de novas seções ou de novos tipos de artigos; (ii) Personalização; (iii) Portabilidade para uma nova comunidade de pesquisa; e (iv) Custo de implementação. Como resultado concluiu-se que os três primeiros aspectos são possíveis e podem ser implementados, porém a um alto custo. Detalhes dessa avaliação são encontrados em (Silva, 1999).

3.6.4 Ferramenta Tutorial

A ferramenta tutorial é um módulo que foi idealizado para o ambiente AMADEUS, mas que ainda se encontra em fase de projeto. Ela difere das outras ferramentas do ambiente por ser

adequada quando o usuário é completamente novato ou quando existe algum tópico que ele desconheça inteiramente. Nesse caso, é desejável uma interação tutor/aprendiz, pois o sistema tem mais conhecimento que o usuário. Uma tecnologia que poderá ser empregada na ferramenta tutorial é a de agentes inteligentes, de modo que a ferramenta possa gerenciar um agente colaborador e um agente tutor. Essa abordagem colaborador/tutor já foi utilizada por Chan & Baskin (1990)⁶.

3.7 Comentários sobre as Ferramentas

As ferramentas descritas neste capítulo foram apresentadas de acordo com a abordagem que utilizam para auxiliar escritores acadêmicos, indo da mais geral para a mais específica, isto é, enquanto as primeiras ferramentas abordam o processo desde a geração de idéias as últimas focalizam o produto (texto).

As primeiras ferramentas descritas, o WE e o Writer's Assistant, partem do princípio de que, diminuindo-se a sobrecarga cognitiva do escritor durante o processo de escrita, por meio do suporte à organização e hierarquização de idéias, ele se sentirá mais livre para gerenciar todas as restrições embutidas no processo (Sharples & Pemberton, 1992), sendo maiores as chances de produzir bons textos. Essa abordagem focaliza o processo, visando melhorar o produto. No entanto, essas ferramentas não oferecem auxílio para dificuldades como o bloqueio na geração do primeiro rascunho ou a falta de conhecimento da retórica do gênero acadêmico ou mesmo o seu uso inadequado.

Preocupações com essas dificuldades foram demonstradas no projeto Composer, que era direcionado para o auxílio de escritores não-nativos que tinham que escrever em inglês. Uma abordagem adotada para resolver ambas as dificuldades foi a utilização de textos autênticos como modelos de boa escrita, de modo que o usuário pudesse se familiarizar com a retórica observando sua realização e, ao mesmo tempo, quebrar o bloqueio inicial da escrita reutilizando padrões observados nos textos-exemplo. O Academic Writer, ferramenta derivada do projeto Composer, implementa essa abordagem e também utiliza outro recurso para auxiliar na geração das idéias iniciais: um *brainstorming* dirigido, isto é, uma seção de perguntas que tenta fazer com que o usuário identifique as idéias principais que devem ser esclarecidas em seu texto (p.e., qual é sua “*research question*”). Uma vez que essa ferramenta

⁶ Chan, T.W., Baskin, A.B. (1990). Learning Companion Systems. In *Intelligent Tutoring Systems: At the Crossroads of Artificial Intelligence and Education*, C. Frasson and G. Gauthier (eds.), Norwood: Ablex apud (Silva, 1999)

é dirigida para usuários que escrevem em uma segunda língua, também oferece apoio lingüístico, como padrões de escrita (*collocations*) e listas de falsos cognatos.

Outras ferramentas voltadas para escritores não-nativos focaram diretamente o produto e assim se aprofundaram nos aspectos mais superficiais do texto, como a estruturação esquemática e suas estratégias de realização. Esse é o caso do Abstract Helper e do ambiente AMADEUS. Essas ferramentas também utilizam a abordagem de reutilização de exemplos autênticos, como o Academic Writer, porém oferecem mais detalhes quanto à função retórica dos elementos. O AMADEUS, em especial, oferece ao usuário um conhecimento bastante detalhado, uma vez que, além de determinar componentes estruturais, identifica também um grupo de estratégias retóricas para cada componente que pode ser empregado para a sua realização.

É importante destacar que à medida que o conhecimento utilizado pela ferramenta vai sendo mais aprofundado, um recorte no seu domínio de aplicação também é feito. Enquanto ferramentas que abordam aspectos mais gerais, como o WE, o Writer's Assistant e o Academic Writer, abrangem o texto como um todo, ferramentas que oferecem uma ajuda mais especializada e direcionada à estrutura esquemática, como o Abstract Helper e o AMADEUS, são específicas para uma seção ou outra do texto (o Abstract Helper aborda o resumo e o AMADEUS a introdução).

3.8 Considerações Finais

Neste capítulo foi apresentada uma revisão das ferramentas de auxílio à escrita encontradas na literatura que se ocupam de duas fases do processo de escrita: planejamento e escrita. As ferramentas que tratam da fase de revisão, também chamadas de ferramentas de pós-processamento (p.e., corretores ortográficos, gramaticais e estilísticos), não foram abordadas por não serem o foco deste trabalho. Outra classe de ferramentas que não foi abordada nessa revisão bibliográfica são os hipertextos explicativos, que funcionam como sistemas de referência e auto-instrução, devido a sua funcionalidade óbvia.

Essa revisão bibliográfica teve por objetivo verificar quais os problemas de escrita científica atualmente tratados por ferramentas computacionais e quais as abordagens utilizadas por essas ferramentas para solucionar esses problemas. Os principais problemas atacados pelas ferramentas estudadas foram:

- Sobrecarga cognitiva no momento da geração e organização das idéias;
- Bloqueio para escrita do primeiro rascunho;

- Falta de conhecimento da retórica do gênero acadêmico ou uso inadequado da mesma;
- Falta de familiaridade com padrões de escrita utilizados para o gênero acadêmico.

Para tentar minimizar esses problemas, as seguintes abordagens foram implementadas pelas ferramentas:

- Apoio à organização das idéias através do uso de redes de nós e elos e de *outliners*;
- Uso de *brainstorm* dirigido;
- Uso de hipertextos explicativos do gênero acadêmico;
- Uso de textos autênticos como exemplos de boa escrita;
- Uso de possíveis modelos de estruturas esquemáticas;
- Indicação de estratégias retóricas para a realização de componentes da estrutura esquemática;
- Uso de bases de padrões de escrita (p.e. *collocations*, texto reutilizável).

O levantamento desses problemas e a identificação das abordagens utilizadas para minimizá-los, juntamente com as características de projeto do sistema AMADEUS, deverão nortear o projeto da ferramenta de auxílio à escrita de resumos e introduções de dissertações e teses, que será um dos produtos deste trabalho.

Capítulo 4

Atividades Desenvolvidas

Neste capítulo são apresentadas algumas das atividades de maior porte que foram desenvolvidas até o momento. Na primeira seção é descrito o CorpusDT, um corpus de dissertações e teses que foi compilado especialmente para a análise desse gênero. Na Seção 4.2 é descrito o conjunto de etiquetas utilizadas para a marcação estrutural desse corpus. Em seguida, são apresentados os resultados extraídos de um questionário aplicado a potenciais usuários de uma ferramenta de auxílio à escrita de textos acadêmicos.

4.1 O CorpusDT¹

Esta seção descreve o CorpusDT, destacando os motivos que nos levaram à sua formação, uma descrição quantitativa que reflete a distribuição dos textos contidos no corpus e uma breve análise estrutural sobre o seu conteúdo.

4.1.1 *Por que Formar um Novo Corpus?*

Para a investigação da estrutura esquemática e dos padrões de escrita utilizados por escritores brasileiros em teses e dissertações, fez-se necessária a formação de um corpus específico para

¹ Esta seção consiste de uma versão de Feltrim et al. (2001).

esse propósito, ou seja, um corpus formado por dissertações e teses em português. No entanto, ainda era preciso restringir o corpus a uma área de conhecimento específica, pois é conhecido que existem diferenças estruturais entre diferentes áreas (por exemplo, textos das Ciências Humanas são pouco estruturados, ao passo que textos da área das Ciências Exatas possuem uma estrutura bem definida). Sendo assim, decidimos pela formação de um corpus de dissertações e teses em Ciências Exatas.

É claro que quanto mais específico for o propósito de um corpus, maior a chance de esse corpus ser representativo dentro do seu propósito. Sendo assim, percebemos que a área das Ciências Exatas ainda era muito ampla para o estudo pretendido, de modo que decidimos direcionar a pesquisa a dissertações e teses em Ciências da Computação, e a formação do corpus a textos dessa área. A escolha por essa área específica se deveu principalmente ao fato deste trabalho estar sendo desenvolvido em um departamento de Ciência da Computação.

A seguir são descritos alguns aspectos quantitativos dos textos coletados para a formação do CorpusDT.

4.1.2 Os Textos Coletados

Conforme comentado na subseção anterior, escolhemos coletar especificamente teses e dissertações de Ciências da Computação e, devido ao fato de trabalharmos em um departamento dessa área, a coleta dos textos foi de certa forma facilitada. Todos os trabalhos foram coletados já em forma eletrônica. Alguns arquivos foram encontrados no repositório público de teses e dissertações da USP², outros foram encontrados na *homepage* do LaSDPC³, que possui uma área para repositório de produção científica. O restante dos arquivos foi coletado diretamente com seus autores ou com seus respectivos orientadores.

Foram encontradas algumas dificuldades na manipulação dos arquivos. Devido ao fato de os arquivos estarem em vários formatos, sendo alguns deles não manipuláveis por ferramentas para manipulação de corpus (por exemplo, arquivos *postscript*), foi necessário um grande esforço para a conversão correta desses arquivos. Além da conversão em si, também foi necessário extrair as figuras, gráficos e tabelas dos textos, substituindo cada um desses elementos por uma etiqueta que os identificasse. O conteúdo textual dos trabalhos foi mantido inalterado, incluindo possíveis erros ortográficos/gramaticais. Os trabalhos estão hoje armazenados em arquivos do tipo texto sem formatação (extensão TXT), sendo que o

² www.teses.usp.br

³ Laboratório de Sistemas Distribuídos e Programação Concorrente – ICMC – USP (<http://lasdpc.icmc.sc.usp.br>)

tamanho do arquivo compactado contendo o corpus é de 2.972 Kbytes. Atualmente, esse corpus não se encontra disponível via Web.

Até o momento, foram coletados trabalhos de várias áreas de pesquisa da Ciência da Computação, num total de 52 arquivos, sendo 49 dissertações de mestrado e 3 teses de doutorado. Os trabalhos datam de 1994 a 2001, sendo que a grande parte data dos últimos três anos. A Tabela 4.1 mostra a distribuição dos trabalhos entre as áreas de pesquisa reconhecidas no corpus.

Tabela 4.1. Número de trabalhos em cada área de pesquisa

Área de Pesquisa	Dissertação	Tese
Banco de Dados	3	-
Inteligência Computacional	7	-
Engenharia de Software	15	1
Hipermídia	12	-
Sistemas Digitais	1	-
Sistemas Distribuídos e Programação Concorrente	10	2
Computação Gráfica e Processamento de Imagens	1	-
Total	49	3
	52	

Cada dissertação ou tese está armazenada em um único arquivo texto, sendo que foram mantidas as identificações iniciais de cada trabalho como primeiro elemento dentro do texto. Os arquivos estão identificados de acordo com o seu tipo (dissertação ou tese) e de acordo com a área de pesquisa a que pertencem. Os nomes de arquivos iniciados em “t_” identificam teses de doutorado, os restantes identificam dissertações de mestrado. A sigla que segue “t_” ou que nomeia o arquivo identifica a área de pesquisa do trabalho. Cada arquivo é identificado dentro da sua área por um número. Desse modo, o arquivo `t_es1.txt` identifica uma tese de doutorado da área de Engenharia de Software.

As Tabelas 4.2 e 4.3 apresentam alguns números que dão uma idéia da distribuição do corpus com relação ao seu tamanho. A Tabela 4.2 mostra uma comparação entre os tipos de texto, dissertação e tese. São apresentados os totais de palavras no corpus de cada tipo (dissertações e teses), a média de palavras em dissertações e em teses, e também o número de palavras da maior e da menor dissertação e tese presente no corpus.

Tabela 4.2. Totais de palavras em dissertações e teses

	Total	Média	Maior	Menor
Teses	173.579	57.860	64.473	54.497
Dissertações	1.429.323	29.170	47.071	13.653

A Tabela 4.3 mostra os totais de palavras distribuídos pelas áreas da ciência da computação contempladas no corpus. Para a soma desses totais foram consideradas tanto as dissertações quanto as teses. É apresentado para cada área: o número de textos da área, o total de palavras nos textos daquela área e a média de palavras por texto. Também é mostrado o total de palavras global do corpus (1.602.902 palavras) e a média total de palavras, isto é, se tomados todos os textos de todas as áreas.

Tabela 4.3. Totais de palavras divididos por área de conhecimento

Área de Pesquisa	Nº_textos	Nº_palavras	Média
Banco de Dados	3	79.561	26.520
Inteligência Computacional	7	181.757	25.965
Engenharia de Software	16	556.441	34.778
Hipermídia	12	312.117	26.010
Sistemas Digitais	1	26.894	26.894
Sistemas Distribuídos e Programação Concorrente	12	425.139	35.428
Computação Gráfica e Processamento de Imagens	1	20.993	20.993
Total	52	1.602.902	30.825

4.2 Análise Superficial da Estrutura Esquemática no CorpusDT

Com o objetivo de levantar certos padrões esquemáticos (componentes e subcomponentes) presentes no corpus, foi realizada uma análise superficial dos textos, visando estabelecer um recorte das partes dos textos que seriam mais aprofundadas no momento da marcação.

Feita uma primeira análise, notou-se que os componentes de mais alto nível relatados na bibliografia pesquisada de fato encontravam-se nos textos do corpus. Os macrocomponentes presentes nas dissertações e teses analisadas são: Resumo, *Abstract*⁴, Introdução, Revisão da Literatura, Metodologia Utilizada, Resultados Obtidos, Discussão dos Resultados e Conclusões.

Desses componentes, os que aparecem nos textos de forma bem definida são o Resumo, a Introdução e as Conclusões. Esse fato foi constatado na análise do corpus e já era esperado, pois esses componentes exercem papéis extremamente importantes em um texto acadêmico. O componente Revisão da Literatura aparece tanto de forma definida, em uma única seção, como “misturada” a outros componentes como, por exemplo, a Metodologia. Em

⁴ Neste trabalho, não citaremos o *abstract* como um componente, por se tratar de uma parte de texto escrita em inglês que, normalmente, é uma tradução do resumo.

geral, isso depende do assunto abordado na dissertação ou tese. Quando a dissertação (ou tese) trata de um assunto específico, é comum que a Revisão da Literatura seja escrita como uma seção única. Quando a dissertação (ou tese) trata vários assuntos (por exemplo, trabalhos que envolvem mais de uma subárea de pesquisa da Ciência da Computação e trabalhos multidisciplinares), esse componente tende a ser dividido em mais de uma seção, geralmente acompanhando as partes da Metodologia.

Os outros componentes observados, Metodologia Utilizada, Resultados Obtidos e Discussão dos Resultados, aparecem no corpus seguindo diferentes padrões de combinação. Em geral, os três são combinados em diferentes seções de texto, seguindo a ordem metodologia-resultado-discussão dentro de cada seção. No entanto, esse padrão pode ser cíclico, isto é, pode ficar se repetindo dentro da seção. Isso faz com que os componentes fiquem “embaralhados”, dificultando a identificação de cada um. Outro padrão comum é a combinação dos Resultados Obtidos com a Discussão dos Resultados em uma ou mais seções. Nesse caso, também pode ocorrer a repetição cíclica do padrão dentro de uma mesma seção. Em alguns casos, cada um desses componentes aparece isoladamente, mas isso não se mostrou comum no corpus.

Focalizando os componentes Resumo e Introdução, coletamos alguns dados numéricos que demonstram de forma geral como esses componentes ocorrem no CorpusDT. A Tabela 4.4 apresenta a média de palavras dos resumos, mostrando também o tamanho, em palavras, do maior e do menor resumo encontrados no corpus. Vale destacar que o total de resumos analisados é de 49 e não 52, que é o número de textos do corpus. Isso se deve ao fato de alguns textos terem sido coletados sem o resumo.

Tabela 4.4. Dados sobre o tamanho em palavras dos resumos

Média de palavras	Maior	Menor
189	425	65

A Tabela 4.5 apresenta a média de palavras das introduções em geral, incluindo dissertações e teses, a média de palavras das introduções de dissertações e a média de palavras das introduções de teses. Nessa tabela também consta o tamanho, em palavras, da maior e da menor introdução encontradas no corpus.

Tabela 4.5. Dados sobre o tamanho em palavras das introduções

Média Dissertação	Média Tese	Média Total	Maior	Menor
1369	1981	1404	2652	359

Sendo os componentes Resumo, Introdução e Conclusão o foco deste trabalho, uma análise mais detalhada é necessária e está em andamento, na qual o objetivo é identificar subcomponentes e estratégias retóricas. Essa tarefa foi iniciada com a análise dos resumos que se encontra em fase de finalização. Para obtermos algum grau de certeza quanto aos subcomponentes e estratégias retóricas marcadas, uma marcação manual foi realizada, separadamente, por três pessoas: pela aluna e autora desta monografia, Valéria Delisandra Feltrim, e pelas professoras Maria das Graças Volpe Nunes e Sandra Maria Aluísio, respectivamente orientadora e co-orientadora desse trabalho de doutorado. Para que essa tarefa seja finalizada, falta ser observada a concordância entre as três marcações, serem discutidas as discordâncias e ser feita a etiquetagem eletrônica dos resumos. O Quadro 4.1 mostra o modelo utilizado para a marcação dos resumos. Trechos autênticos exemplificando cada subcomponente e estratégias retóricas associadas são mostrados no Apêndice A.

Esse modelo foi feito a partir da revisão da literatura e de uma primeira análise dos resumos, em que foi observada uma dificuldade na marcação do subcomponente Resultados. Acreditamos que tal dificuldade se deve ao fato do corpus ser de uma área científico-tecnológica, em que muitas vezes o resultado do trabalho vem embutido no componente Propósito. É importante destacar que, embora o corpus seja composto por dissertações e teses defendidas, elas não passaram por uma avaliação rigorosa, como é o caso, por exemplo, de artigos publicados em revistas e em conferências de prestígio. Dessa forma, a marcação tem sido feita com um olhar crítico, e uma análise crítica mais completa deverá ser realizada ao término da fase de marcação. O mesmo procedimento deverá ser realizado com as introduções e conclusões.

Com a marcação dos resumos, pudemos observar, entre outras coisas, que muitos textos apresentam estrutura e, algumas vezes, a própria realização lingüística semelhantes, principalmente entre textos produzidos por alunos sob a mesma orientação. Tais semelhanças também foram observadas entre textos de autores do mesmo grupo de pesquisa. Sendo assim, é possível que o corpus seja estendido, incorporando principalmente teses e dissertações de outras instituições (atualmente, 2 textos do corpus não são do ICMC). Dessa forma, esperamos ter uma visão mais abrangente do gênero acadêmico em português.

Quadro 4.1. Modelo de Resumo utilizado para a marcação dos resumos do CorpusDT

⇒	Contexto
	C1. Declarar proeminência do tópico (argumentação)
	C2. Familiarização de termos
	C3. Introduzir a pesquisa a partir da grande área
⇒	Gap
	G1. Problemas/dificuldades
	G2. Necessidades/requisitos
	G3. Ausência ou pouca pesquisa anterior
⇒	Propósito
	P1. Indicar o propósito principal
	P2. Re-frasear/detalhar/especificar o propósito
	P3. Introduzir mais propósitos
⇒	Metodologia
	M1. Listar critérios ou condições
	M2. Descrever materiais e métodos
	M3. Justificar a escolha pelos materiais e métodos
⇒	Resultado
	R1. Descrição do artefato
	R2. Apresentar os resultados
⇒	Conclusão
	Co1. Recomendação
	Co2. Comentários/discussão (generalizar/explicar/comparar) dos resultados
	Co3. Valor do trabalho

Atualmente, existe uma versão do CorpusDT com todas as marcações de *layout*. A marcação do corpus com as etiquetas de *layout* foi de certa forma trabalhosa, pois, embora tenha tido auxílio computacional para a verificação da boa formação dos documentos etiquetados, as etiquetas foram colocadas manualmente. O conjunto de etiquetas utilizado é discutido na próxima seção e o detalhamento é apresentado no Apêndice B.

4.3 A Etiquetação do Corpus

Para que o corpus possa ser manipulado e “compreendido” pelas ferramentas pretendidas neste projeto é necessário que ele esteja eletronicamente marcado com um conjunto de etiquetas apropriadas. Para a realização dessa marcação, optamos por utilizar o XML, um padrão de marcação para documentos, aprovado pelo W3C⁵, que já vem sendo utilizado para a marcação de corpus. Embora projetos importantes relacionados à marcação de corpus

⁵ World Wide Web Consortium - <http://www.w3.org> (Acessado em 21-02-2002)

utilizem o SGML⁶ como linguagem de marcação, como o TEI⁷, preferimos utilizar o XML principalmente pela facilidade de visualização, visto que um documento XML pode ser visualizado em *browsers* comuns, como o Microsoft[®] Internet Explorer. Exemplos de projetos que utilizam o XML para a marcação de corpus são o MATE⁸ e GNOME⁹.

O XML – *Extensible Markup Language*, é subconjunto do SGML que foi criado para a manipulação de documentos na Internet. Ele incorpora 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 marcação, uma vez que o XML não especifica como os seus elementos devem ser apresentados. Na verdade, não especifica nem mesmo se eles devem ser apresentados.

Uma forma de apresentação de documentos XML que tem sido utilizada consiste em transformar esses documentos em documentos HTML¹⁰, possibilitando assim que sejam visualizados em qualquer *browser* (XML, 2001). A tecnologia que permite a transformação de um documento XML em HTML é conhecida como XSL – *Extensible Stylesheet Language*, que é uma linguagem para a definição de *stylesheets*¹¹. Utilizando XSL podemos especificar quais partes do documento XML serão apresentadas e quais serão os seus formatos, o que possibilita a criação de diversas visões a partir de um mesmo documento. Embora ainda não tenhamos definido o modo como o corpus será apresentado, a possibilidade de podermos ter diferentes visões desse corpus é desejável. O uso do XSL deverá ser futuramente explorado com mais cuidado.

A definição de um documento XML pode ser feita através de um DTD – *Document Type Definition*. No DTD são definidas as etiquetas (*tags*) que serão utilizadas para estruturar os dados. Entretanto, recentemente o W3C aprovou um novo padrão para a definição de documentos XML, o *XML Schema*, que deverá substituir o DTD. Como é nosso objetivo utilizar conceitos atuais neste projeto, optamos por utilizar o *XML Schema* como linguagem de definição.

Tendo nos decidido pelo uso do XML, propomos um conjunto de etiquetas para a marcação do corpus em quatro níveis:

- marcação do layout da dissertação (identificação, capítulos, etc);

⁶ *Standard Generalized Markup Language*

⁷ <http://www.tei-c.org/> (Acessado em 10-03-2002)

⁸ <http://mate.nis.sdu.dk/> (Acessado em 04-03-2002)

⁹ <http://www.herc.ed.ac.uk/~gnome> (Acessado em 10-07-2001)

¹⁰ *Hypertext Markup Language*

¹¹ *Stylesheet* ou “folha de estilo”, neste contexto, consiste de uma série de instruções declarativas que define como o documento XML será apresentado em outro formato, por exemplo, em HTML.

- marcação de subcomponentes dos componentes Resumo e Introdução;
- marcação de estratégias retóricas de cada subcomponente;
- marcação de padrões de textos que possam ser reutilizados nos Resumos e Introduções.

Esse conjunto de etiquetas foi definido em um *XML Schema*, que é apresentado no Apêndice B. É possível que esse conjunto seja alterado (ou aumentado) de acordo com o prosseguimento da atividade de marcação, contudo a versão até agora utilizada tem-se mostrado estável. As etiquetas foram propostas com base na literatura pesquisada sobre a estruturação esquemática de textos acadêmicos e também na marcação utilizada nas bases de casos do sistema AMADEUS. A Figura 4.1 foi extraída da ferramenta *XML Spy 4.0*¹², que vem sendo utilizada para a validação dos arquivos etiquetados, e dá uma idéia da estrutura hierárquica definida para a marcação do corpus.

É importante destacar que o esquema definido é bastante flexível no nível dos subcomponentes e das estratégias retóricas, pois os nomes dos subcomponentes e estratégias não foram especificados no *XML Schema*. Sendo assim, haverá sempre um modelo de subcomponentes e estratégias que definirá esses nomes, como o exemplo apresentado no Quadro 4.1 da seção anterior. Isso nos dá mais liberdade no caso de serem necessárias modificações no modelo de subcomponentes e estratégias no momento da marcação. Esse esquema de marcação é exemplificado na Figura 4.2. Essa figura, também extraída da ferramenta *XML Spy*, apresenta uma representação gráfica das etiquetas definidas para a marcação dos resumos.

¹² www.xmlspy.com (Acessado em 21-02-2002)

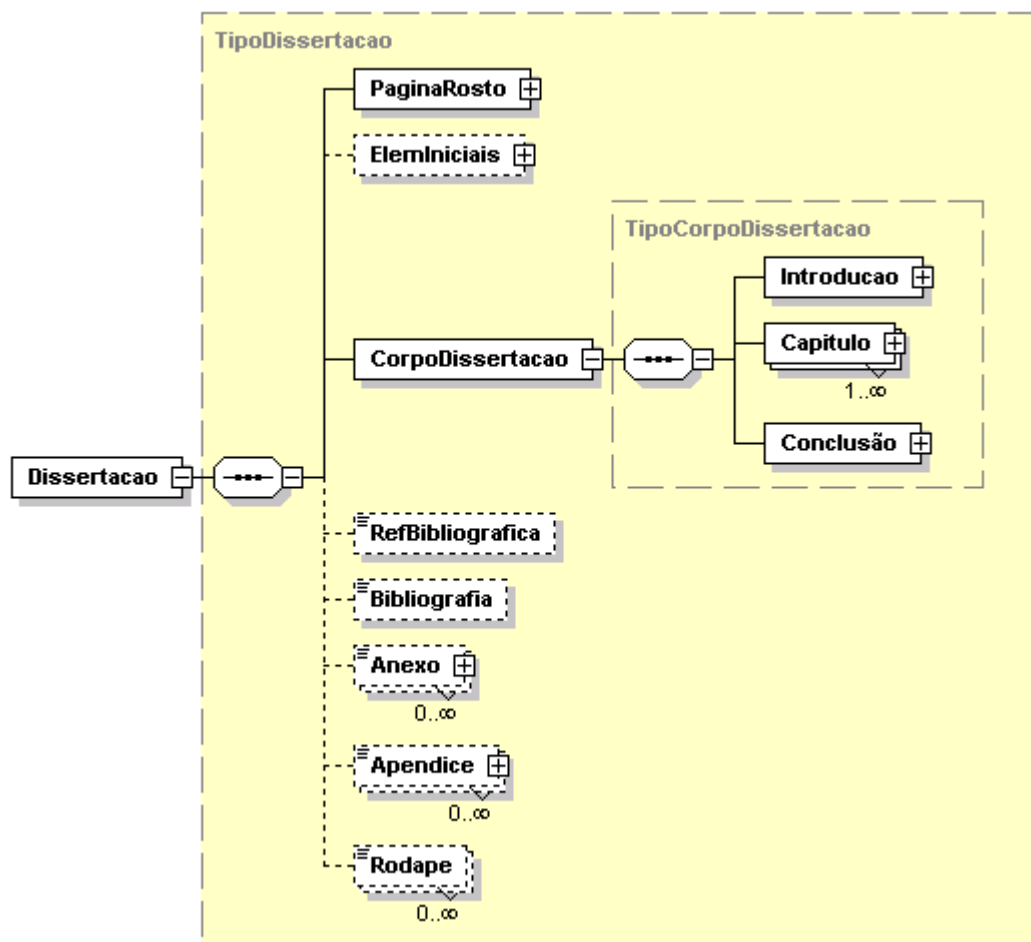


Figura 4.1. Representação gráfica do XML Schema (nível 0) para marcação do CorpusDT, mostrando a expansão do elemento CorpoDissertação (nível 1)

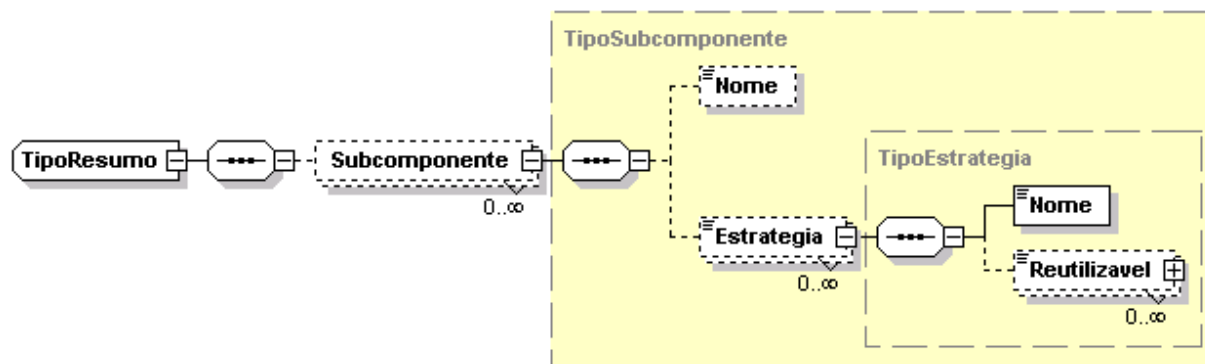


Figura 4.2. Representação gráfica do elemento Resumo (nível 2 do XML Schema), mostrando a expansão dos elementos Subcomponente (nível 3) e Estratégia (nível 4)

4.4 Pesquisa Realizada com Usuários Alvo

Como foi apresentado no Capítulo 3, existem várias ferramentas de auxílio à escrita acadêmica já desenvolvidas, ou ainda em desenvolvimento. No entanto, todas são voltadas para a escrita de textos acadêmicos em inglês, seja para autores nativos ou não. Não foram encontradas referências a ferramentas dessa natureza voltadas para a escrita acadêmica em português. Dessa forma, o projeto de uma ferramenta computacional de auxílio à escrita de textos acadêmicos em português é uma iniciativa inédita nessa área.

Para nos guiar em tal projeto, utilizamos a seguinte abordagem: primeiramente investigamos a literatura especializada sobre sistemas computacionais dessa natureza, mesmo que para outra língua, e levantamos quais os problemas que esses sistemas atacam e como esses problemas são abordados. Numa segunda etapa, analisamos um corpus de textos acadêmicos escritos em português (o CorpusDT, apresentado na Seção 4.1), visando identificar os principais problemas, como estruturação inadequada, uso de léxico inadequado, uso de tempo verbal inadequado, entre outros, que poderiam ser minimizados com o auxílio de uma ferramenta computacional especializada.

Essa análise nos deu uma visão dos principais problemas de escrita e das possíveis soluções que poderiam ser implementadas para minimizá-los. Também é importante saber o que os usuários potenciais esperam encontrar em um sistema computacional como esse, uma vez que não existe semelhante disponível. Para tal, aplicamos um questionário a um grupo de 28 estudantes de pós-graduação do ICMC, sendo 22 da área da Ciência da Computação, 2 da Matemática Computacional, 2 da Matemática e 2 da Estatística. Os estudantes da área da Ciência da Computação foram selecionados de modo que fossem entrevistadas pelo menos duas pessoas de cada uma das áreas de concentração da pós-graduação do ICMC. Dessa forma, evitamos uma amostra viciada em sujeitos que escrevam influenciados pelo mesmo orientador ou pelos mesmos modelos. Esses estudantes foram divididos em dois grupos: 4 iniciantes e 24 experientes. Chamamos de iniciantes os sujeitos sem experiência em escrita acadêmica (estudantes que ingressaram no programa de mestrado e ainda não apresentaram a monografia de qualificação), e de experientes os que já tenham produzido algum texto desse tipo (estudantes que estão finalizando a dissertação de mestrado ou que fazem parte do programa de doutorado). Vale destacar que o questionário também foi aplicado a 4 linguistas (3 experientes e 1 iniciante), e os resultados foram semelhantes aos do grupo descrito anteriormente.

O questionário, iniciado com um pequeno texto explicativo, abordou três aspectos: (1) o conhecimento pessoal sobre a escrita de textos acadêmicos, (2) as funcionalidades de uma

possível ferramenta e (3) o modo de interação da ferramenta com o usuário, totalizando 21 questões, que foram elaboradas como questões de múltipla escolha ou como questões em que os itens devem ser classificados de acordo com uma dada escala. Também foram elaboradas duas questões abertas para que os usuários pudessem fazer sugestões. O questionário completo é apresentado no Apêndice C.

Os gráficos mostrados a seguir demonstram os resultados obtidos para cada grupo, em percentual. Os gráficos 4.1 a 4.7 enfocam a escrita acadêmica de modo geral, sem a interferência de aspectos computacionais. Os gráficos 4.8 e 4.9 enfocam as características funcionais de uma possível ferramenta computacional e os gráficos 4.10 e 4.11 enfocam preferências quanto ao modo de interação com essa ferramenta.

Pelo Gráfico 4.1 podemos notar que a maioria dos estudantes experientes considera seu conhecimento sobre a escrita acadêmica “bom” ou “razoável”, ainda que apenas 25% já tenha escrito mais de 10 textos acadêmicos (Gráfico 4.2). Mesmo entre os estudantes iniciantes, 50% consideram ter um conhecimento “razoável”, embora parte dessa porcentagem nunca tenha escrito um texto acadêmico individualmente. Em contrapartida, mesmo sem ter sido concluída uma análise completa do corpus, foi possível notar problemas como estruturação inadequada de subcomponentes, uso de léxico inadequado, uso de tempo verbal inadequado e outros problemas menos específicos, como erros ortográficos e gramaticais. Isso nos leva a crer que, embora os estudantes digam ter um conhecimento bom ou razoável do gênero acadêmico, existem dificuldades na realização textual desse gênero.

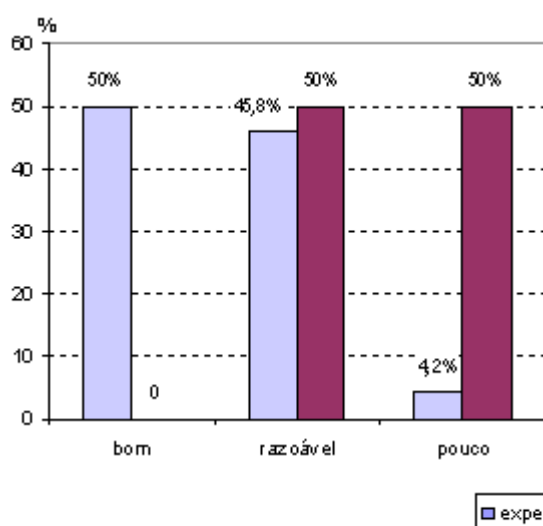


Gráfico 4.1. Respostas de estudantes iniciantes e experientes quanto ao nível de conhecimento da escrita acadêmica

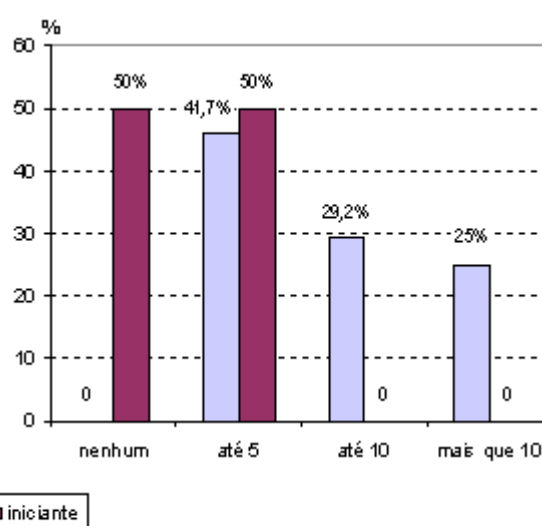


Gráfico 4.2. Respostas de estudantes iniciantes e experientes quanto ao número de textos acadêmicos já produzidos

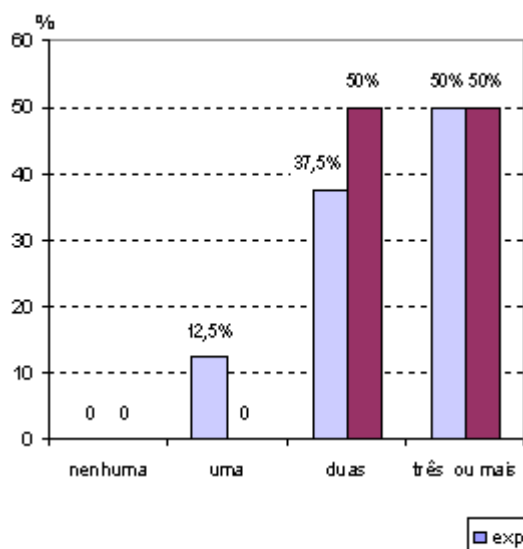


Gráfico 4.3. Respostas de estudantes iniciantes e experientes quanto ao número médio de revisões em cada texto

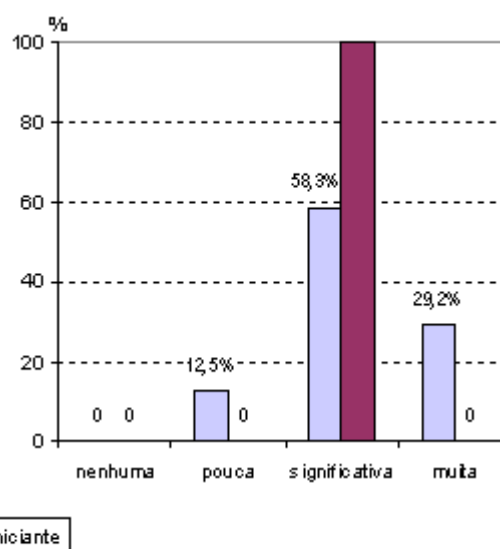


Gráfico 4.4. Respostas de estudantes iniciantes e experientes quanto ao nível de influência do orientador (ou alguém mais experiente) na produção de um texto

O Gráfico 4.3 mostra que tanto estudantes experientes como iniciantes fazem várias revisões em seus textos antes de considerá-los prontos. É sabido que grande parte das revisões é feita pelo próprio orientador, daí sua grande influência na produção do texto final (Gráfico 4.4). Essa influência também foi observada no corpus. Textos produzidos por estudantes sob a mesma orientação tendem a ser semelhantes tanto na estrutura como nas escolhas lexicais. No entanto, essa semelhança também advém dos textos utilizados pelos estudantes como modelos no momento da escrita do seu próprio texto, conforme apresentado no Gráfico 4.5.

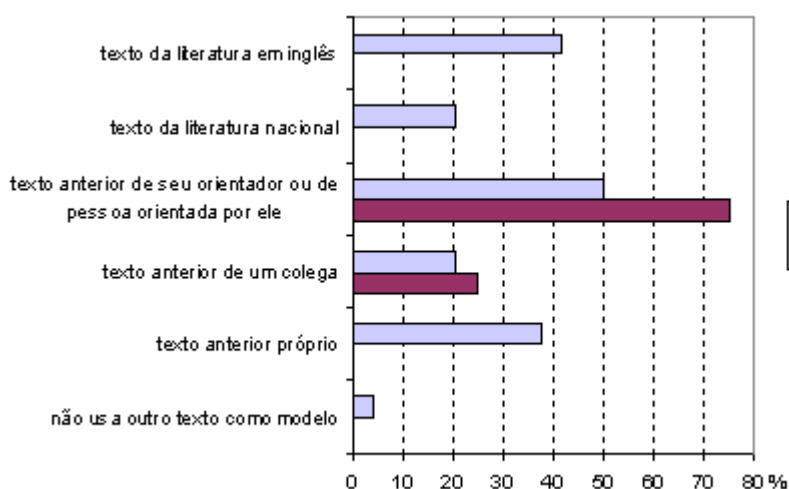


Gráfico 4.5. Respostas de estudantes iniciantes e experientes quanto ao tipo de texto utilizado como modelo para a escrita do seu próprio texto

Como pode ser visualizado no Gráfico 4.5, 50% dos estudantes experientes utilizam como modelo um texto anterior produzido por seu orientador ou por outro estudante orientado por ele. Entre os estudantes iniciantes essa porcentagem é ainda maior. Esse é outro fator que contribui para a semelhança dos textos do corpus. Ainda sobre os modelos, vale destacar que 41,7% dos estudantes experientes utilizam textos da literatura em inglês. De fato, conforme discutido no Capítulo 2, a estrutura esquemática do gênero acadêmico não apresenta diferenças em suas realizações nas línguas inglesa e portuguesa. Portanto, textos em inglês podem ser utilizados como bons modelos, e de fato isso ocorre.

Em uma das questões do questionário foi pedido que os estudantes indicassem o grau de dificuldade encontrada na escrita de cada componente padrão de um texto acadêmico, utilizando uma escala de 1 a 5, em que o 1 indicaria maior grau de dificuldade. O Gráfico 4.6 mostra a porcentagem das respostas dos estudantes experientes, e podemos observar que a Conclusão foi considerada o componente mais problemático (29,2%), seguida da Introdução (20,8%). O componente Revisão Bibliográfica foi considerado o menos problemático (25% grau 4 e 16,7% grau 5). O Gráfico 4.7 mostra a porcentagem das respostas dos alunos iniciantes, em que a Introdução foi considerada o componente mais problemático (75%), seguida da Conclusão (50%). O componente Resultados foi considerado o menos problemático para os estudantes iniciantes (50%).

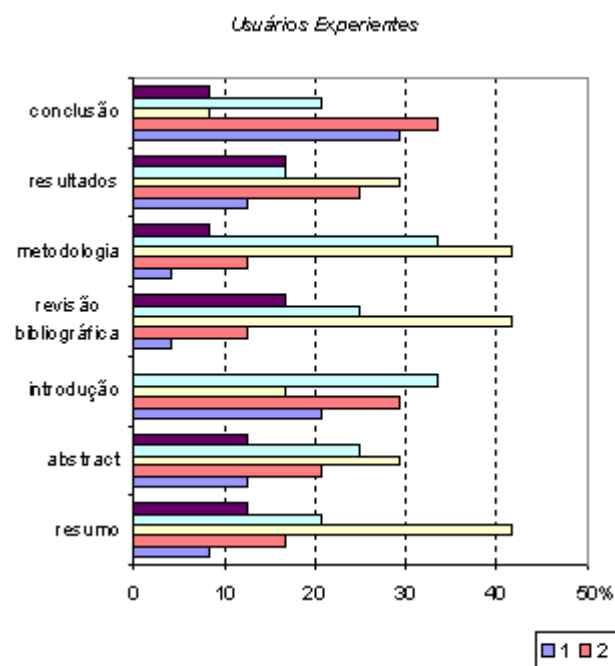


Gráfico 4.6. Respostas dos estudantes experientes quanto ao grau de dificuldade encontrado na escrita de cada componente

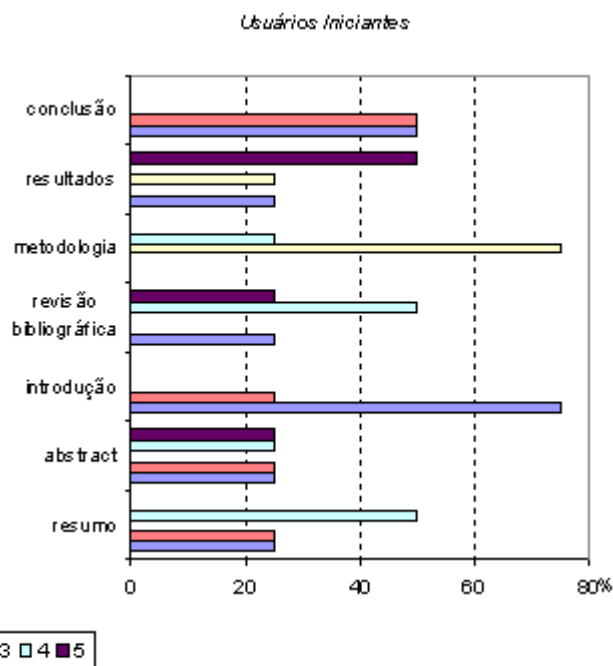


Gráfico 4.7. Respostas dos estudantes iniciantes quanto ao grau de dificuldade encontrado na escrita de cada componente

Numa segunda etapa do questionário, foi apresentada uma lista de funções que poderiam ser implementadas em uma ferramenta computacional de auxílio à escrita acadêmica. Foi pedido aos estudantes que classificassem essas funções utilizando uma escala com quatro níveis: “indispensável”, “desejável”, “interessante” e “dispensável”. A descrição das funções e da escala, tal qual foi apresentada aos estudantes, pode ser vista no Apêndice C.

Conforme mostrado no Gráfico 4.8, duas funções foram destacadas como indispensáveis pelos estudantes experientes: “correção ortográfica” (95,8%) e “correção gramatical e de estilo” (91,7%). Acreditamos que o fato dessas funções já serem conhecidas e utilizadas pelos estudantes provocou um resultado tão expressivo em relação às outras funções. A função “revisão do texto final” também se destacou como indispensável (50%). Com exceção dos dois tipos de *brainstorming* sugeridos (ambos considerados indispensáveis por apenas 16,7%), o restante das funções obteve classificações semelhantes. Ao contrário do que aconteceu com as funções de correção ortográfica, gramatical e de estilo, acreditamos que o fato de os estudantes nunca terem utilizado os outros tipos de funções dificultou a classificação, de modo que não houve resultados mais expressivos.

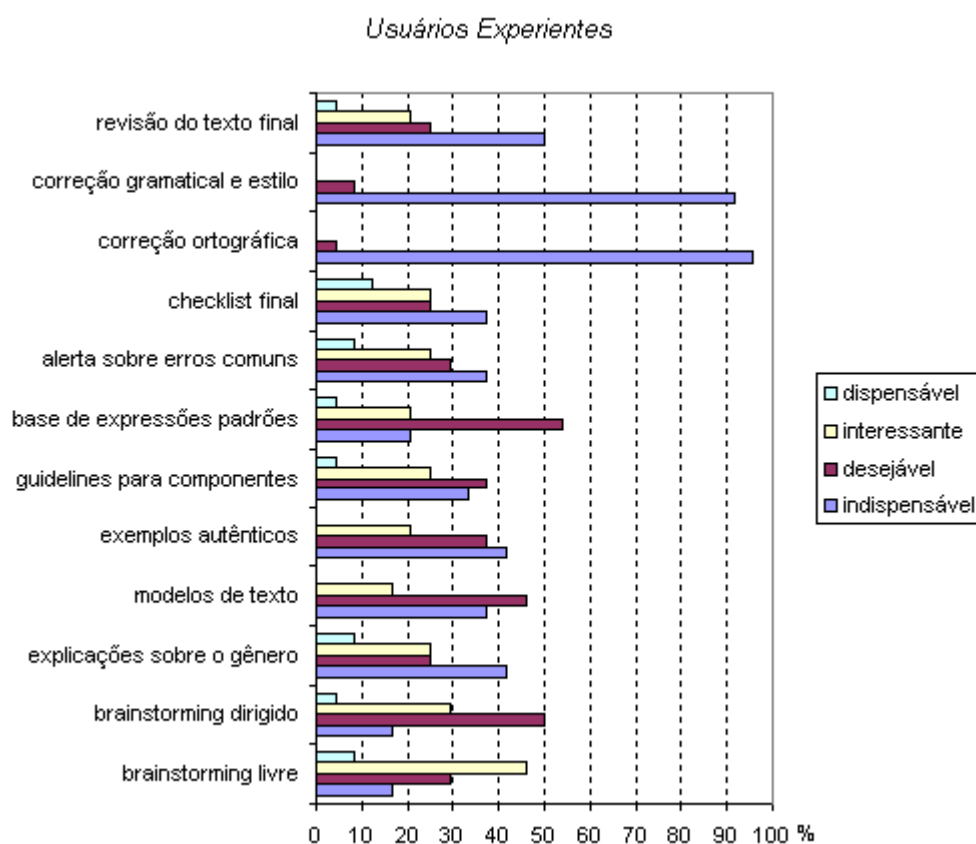


Gráfico 4.8. Respostas dos estudantes experientes quanto a relevância das funções de uma ferramenta computacional de auxílio à escrita

Os estudantes iniciantes também destacaram as funções “correção ortográfica” e “correção gramatical e de estilo” como indispensáveis (ambas com 75%). Diferentemente dos estudantes experientes, 75% dos iniciantes julgaram a função “*guidelines* para a escrita de componentes” indispensável. A função “uso de exemplos autênticos” não foi julgada indispensável, no entanto, 100% dos estudantes iniciantes a classificaram com desejável. Uma vez que esses estudantes têm pouca ou nenhuma experiência na escrita acadêmica, é natural que se mostrem mais interessados em *guidelines* e modelos autênticos do que os estudantes experientes. Como pode ser visto no Gráfico 4.9, nenhuma função foi julgada como dispensável pelos iniciantes.

É importante recordar que a amostra de estudantes iniciantes (4) é bem menor do que a amostra de estudantes experientes (24), por isso os dados apresentados no Gráfico 4.9 podem parecer mais expressivos que os do Gráfico 4.8, mas na realidade, os estudantes iniciantes apresentaram as mesmas dificuldades sentidas pelos estudantes experientes em classificar as funções dentro da escala sugerida.

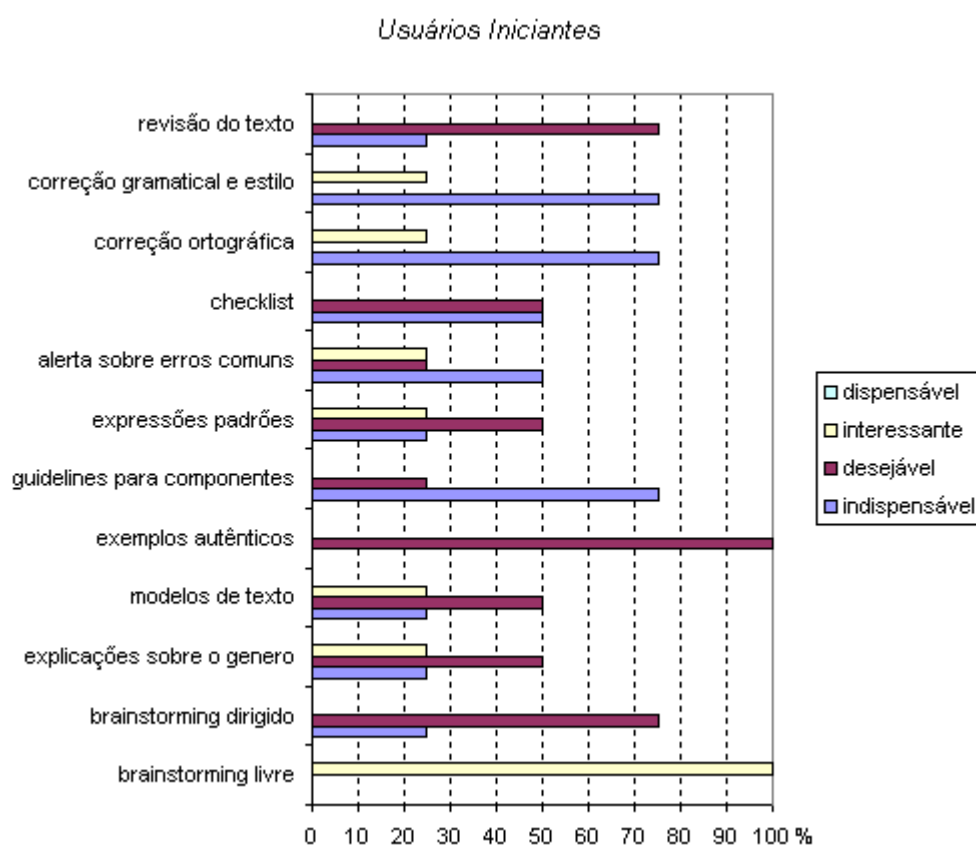


Gráfico 4.9. Respostas dos estudantes iniciantes quanto a relevância das funções de uma ferramenta computacional de auxílio à escrita

Foi dado aos estudantes um espaço para que sugerissem outras funções que julgassem interessantes, diferentes das listadas no questionário. Alguns estudantes (2 iniciantes e 8 experientes) fizeram as seguintes sugestões:

- lista de palavras comuns ao gênero acadêmico com os respectivos significados e sinônimos;
- lista de palavras/expressões que devem ser evitadas no gênero acadêmico, com sugestões de substituição;
- manutenção de um histórico dos textos feitos por um usuário na ferramenta e inserção desses textos numa base para posterior recuperação e reutilização;
- dicionário português/inglês on-line;
- geração automática de *abstract* a partir do resumo;
- apoio de uma ferramenta de tradução automática;
- apoio de uma ferramenta de sumarização automática;
- montagem de uma checklist personalizada pelo usuário;
- apoio ao controle de versões;
- criação de um espaço para anotações, de forma que essas anotações fossem ligadas ao texto original através de *links*;
- verificação da coerência entre Introdução e Conclusões;
- verificação da coerência entre as Considerações Iniciais e Finais de cada capítulo.

Quanto ao modo de interação com o usuário (Gráfico 4.10), ficou clara a preferência dos estudantes, tanto experientes quanto iniciantes, por um ambiente misto, isto é, em que o usuário seja livre para acionar as funções que desejar e que o sistema também interfira quando achar conveniente, porém sem a intensidade de um tutor. Foi sugerido por dois estudantes experientes que o nível de interferência do sistema possa ser personalizado, ou seja, que o usuário possa configurar itens em que o sistema pode interferir ou não, e que esse perfil seja armazenado para posteriores utilizações.

O Gráfico 4.11 evidenciou que o não acoplamento de uma ferramenta de auxílio à escrita acadêmica a um processador de textos não altera a motivação da maioria dos estudantes, tanto experientes (70,8%) como iniciantes (75%), em usá-la, desde que os arquivos manipulados pela ferramenta sejam compatíveis com os manipulados pelos processadores de texto.

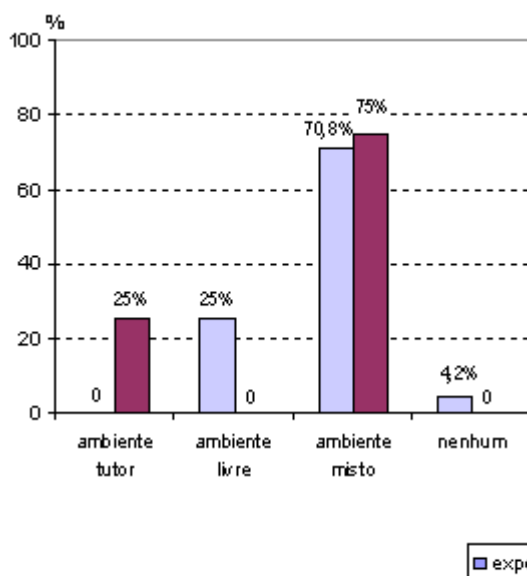


Gráfico 4.10. Respostas dos estudantes iniciantes e experientes quanto ao estilo de interação de interação com o usuário

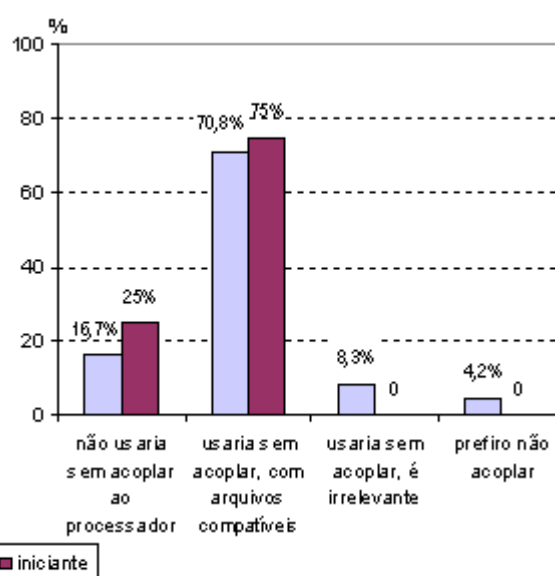


Gráfico 4.11. Respostas dos estudantes iniciantes e experientes quanto à relevância do acoplamento de uma ferramenta de auxílio à escrita acadêmica a um processador de textos

Como citado no início desta seção, o intuito dessa pesquisa foi investigar o que potenciais usuários esperam de uma ferramenta de auxílio à escrita acadêmica. Essa experiência foi muito importante, pois nos deu um cenário mais completo das dificuldades dos usuários e do tipo de auxílio que esperam encontrar. Juntamente com o levantamento dos problemas de escrita e as possíveis soluções investigadas, os resultados obtidos deverão nortear o projeto da ferramenta computacional proposta neste trabalho.

4.5 Considerações Finais

Neste capítulo foram descritas as atividades de grande porte desenvolvidas até o presente momento e seus respectivos resultados. Foi descrito o CorpusDT, que deverá constituir um dos produtos deste projeto, enfocando aspectos quantitativos e qualitativos. Também foi introduzido o esquema de marcação criado para este projeto. Visando estabelecer requisitos para o projeto dos protótipos propostos, foi realizada uma pesquisa junto aos potenciais usuários de ferramentas de auxílio à escrita, por meio de um questionário. Os resultados dessa pesquisa, juntamente com a análise do corpus, deverão determinar os principais problemas de escrita dos autores nativos, que tendem a ser diferentes dos sentidos por não-nativos, pois são mais localizados na realização de um novo gênero do que no domínio da língua.

Capítulo 5

Proposta e Metodologia

O uso de ferramentas computacionais que prestam algum auxílio ao processo de escrita é hoje uma realidade que faz parte do cotidiano da grande maioria dos usuários de computadores. O tipo de auxílio prestado varia entre os mais simples (p.e. dicionários eletrônicos gerais ou especializados, corretores ortográficos) até os mais sofisticados (p.e. para composição de textos), passando por *thesauri*, revisores gramaticais, entre outros.

Para a língua inglesa, cujo tratamento computacional ocorre há mais tempo e por uma comunidade maior – por conta inclusive de ser a língua franca em diversos domínios de conhecimento –, os trabalhos de pesquisa e desenvolvimento de ambientes mais sofisticados de auxílio são mais freqüentes, tendo produzido sistemas comerciais de qualidade.

Para a língua portuguesa, a realidade é outra. Somente recentemente percebe-se um aquecimento nas pesquisas acadêmicas de PLN no Brasil que visam ao desenvolvimento de sistemas robustos, eficazes e eficientes, além de um aparente esforço dedicado à construção de recursos lingüísticos volumosos, imprescindíveis nesse caso. O NILC é um bom exemplo de grupo acadêmico que tem conseguido expressivos resultados, contribuindo efetivamente para o avanço do processamento do português brasileiro. Não se tem notícia, no entanto, de iniciativas que visam à investigação de sistemas de auxílio à escrita do português que, em nível de complexidade e auxílio prestado, vão além dos *thesauri* e revisores gramaticais.

É objetivo desse trabalho contribuir nesse sentido: avançar (se não iniciar!) as investigações sobre ferramentas de auxílio à escrita científica em português, culminando na

implementação de protótipos que ofereçam recursos para a fase de criação e composição de documentos do tipo tese ou dissertação. Necessitando delimitar a área de conhecimento desses documentos para a extração de padrões convencionais, escolhemos as Ciências Exatas, muito provavelmente as mais padronizadas nesse sentido. Como decorrência, por este trabalho estar sendo desenvolvido num departamento de Ciências da Computação, decidimos delimitar os textos a esse domínio, para facilitar a coleta do material lingüístico necessário, bem como esperamos encontrar facilidade para avaliar os protótipos com usuários finais.

5.1 Metodologia

A metodologia adotada deverá seguir, de forma geral, os passos adotados na construção do ambiente AMADEUS. Em particular, deverá ser proposta uma metodologia de avaliação dos protótipos gerados, sendo essa uma das contribuições esperadas desse trabalho. Deve-se ressaltar que, embora tendo o AMADEUS como base e, portanto, queimando-se etapas significativas, as diferenças de língua, de tipo de texto e de usuários-autores fazem deste um novo trabalho que vem exigindo muito esforço na coleta, análise e marcação de corpus, sendo o produto desse esforço outra significativa contribuição que esperamos desse trabalho.

A seguir delimitamos as atividades previstas no desenvolvimento desse trabalho, sendo que a apresentação seqüencial não corresponde necessariamente à ordem de execução das mesmas, conforme mostra a tabela do cronograma proposto, mais adiante.

(1) Pesquisa Bibliográfica.

A pesquisa bibliográfica vem sendo desenvolvida desde o início do projeto e é uma atividade que deve manter-se constante durante todo o andamento da pesquisa, visando dar subsídios às tomadas de decisão. Várias áreas de pesquisa contribuem para esse trabalho e têm sido os alvos da pesquisa bibliográfica: lingüística de corpus, técnicas de escrita científica, ferramentas de auxílio à escrita, técnicas de avaliação de sistemas interativos e de sistemas de PLN, entre outras.

O Capítulo 2 apresentou o resultado da pesquisa bibliográfica realizada na área das técnicas de escrita acadêmica, em que se destacam os trabalhos de Weissberg & Buker (1990) e de Swales (1990), ambos abordando a escrita acadêmica em inglês. As referências encontradas sobre a escrita acadêmica em português são voltadas para o público geral, de modo que apresentam diretivas (*guidelines*) para a boa escrita, não formalizando a estrutura esquemática de um texto acadêmico. No entanto, essas diretivas

indicam muitas semelhanças com a estrutura esquemática detalhada apresentada para o inglês. Dessa forma, decidimos que as referências bibliográficas encontradas que tratavam da estrutura esquemática de textos em inglês seriam aplicadas neste trabalho, embora trabalhemos com o português.

O Capítulo 3 mostrou o resultado da pesquisa bibliográfica sobre ferramentas de auxílio à escrita. Não foi encontrado nenhum material que relatasse pesquisas envolvendo ferramentas de auxílio à escrita acadêmica em português. Isso já era esperado, uma vez que este trabalho pretende iniciar tais pesquisas. Foram descritas seis ferramentas, todas voltadas para a escrita acadêmica em inglês, incluindo o ambiente AMADEUS. Cada uma das ferramentas descritas apresenta particularidades e vantagens (discutidas na Seção 3.7). Esse conjunto de características, tanto funcionais como de interação, deverão servir como guia no projeto dos protótipos resultantes deste projeto.

Outro aspecto deste projeto para o qual foi feita uma pesquisa bibliográfica foi a marcação de corpus. Foram pesquisados trabalhos gerais sobre lingüística de corpus (McEnery, 1997; Sinclair, 1991) e também alguns projetos que trabalham com corpora marcados em XML, como o MATE, cujo objetivo é criar um padrão para a marcação de diálogos, e o GNOME, que trabalha com um corpus etiquetado com informações sintáticas e semânticas para a geração automática de expressões nominais (Poesio, 2000). Essa pesquisa foi realizada para embasar a construção do corpus e a criação do esquema de marcação. A linguagem de marcação XML também foi estudada e para tal utilizamos os documentos disponíveis no *site* do W3C.

Outro assunto que deverá ser investigado são os marcadores discursivos para o português, pois pretendemos identificar e marcar esses elementos nos resumos, introduções e conclusões do corpus. Acreditamos que, sendo essa uma informação importante para que o texto tenha uma boa coesão (Kock, 1998), ela será bastante útil para os futuros usuários das ferramentas propostas neste projeto. No entanto, a identificação dos marcadores discursivos não será uma tarefa trivial, pois a formalização e criação de tipologias para o português ainda está na sua infância. Em (Paizan, 2001), é proposta uma classificação dos marcadores discursivos para o inglês, sendo que também foi feita uma lista contendo os correspondentes marcadores para o português. Em (Dias da Silva, 2002) também é proposta uma tipologia para os marcadores discursivos do português. Dessa forma, a anotação dos marcadores discursivos no CorpusDT deverá ser iniciada tendo por orientação esses trabalhos.

Conforme citado inicialmente, a pesquisa bibliográfica deverá continuar até o término deste projeto, visando o embasamento teórico de cada etapa a ser cumprida.

- (2) Coleta de textos acadêmicos do tipo tese e dissertação na área de Ciências da Computação, compondo o corpus de estudo.

Essa atividade já foi concluída e o corpus resultante foi descrito na Seção 4.1.2 desta monografia.

- (3) Análise Superficial do Corpus.

Essa atividade já foi concluída e visou ao levantamento de padrões (componentes e subcomponentes) a partir do corpus para que fosse feito um recorte das partes desses padrões que seriam mais aprofundadas e, portanto, os alvos de maior atuação das ferramentas. Frente aos dados levantados, apresentados na Seção 4.2, decidimos que os componentes alvos de uma análise detalhada (identificação de subcomponentes, estratégias retóricas e padrões de escrita) seriam o Resumo, a Introdução e a Conclusão.

- (4) Proposta de um conjunto de marcas (*tags*).

Para que os casos do corpus sejam manipulados pelo sistema, é necessário etiquetá-los de forma a associar-lhes informações úteis para o raciocinador de casos. Para tal, foi proposto um conjunto de marcas, em XML, visando à marcação do corpus em quatro níveis – layout, subcomponentes, estratégias retóricas e padrões de escrita – conforme descrito na Seção 4.3.

- (5) Marcação do Corpus.

Essa atividade consiste da marcação manual do corpus, e já foi em parte realizada. Conforme citado na Seção 4.2, a marcação de *layout* já foi realizada nos textos completos. A marcação de subcomponentes e estratégias retóricas já foi realizada nos resumos e em breve será completada. O mesmo processo de marcação deverá ser realizado para as introduções e conclusões. Numa etapa posterior, serão marcados padrões de escrita (trechos de texto que podem ser reutilizados, isto é, material não-factual) e marcadores discursivos.

(6) Análise Crítica do Corpus.

Essa atividade deverá ser realizada à medida que a atividade (5) for sendo completada. Tal análise crítica deverá seguir uma metodologia semelhante à empregada por Liddy (1991) para a análise de *abstracts*. Além disso, será criada uma rubrica própria para este trabalho, baseada em rubricas conhecidas, como as utilizadas para os exames GMAT¹ e TWE². O resultado dessa análise deverá mostrar os erros e os acertos encontrados no corpus e deverá determinar os textos que formarão a base de casos dos protótipos a serem implementados.

(7) Projeto do Sistema Computacional.

Nessa atividade, deveremos definir o conjunto de funções que farão parte das ferramentas, baseados na revisão bibliográfica, apresentada no Capítulo 3, e nos resultados do questionário, apresentados no Capítulo 4. Também deverão ser levados em conta os resultados do questionário quanto ao estilo de interação com o usuário.

O ponto de partida deverá ser o ambiente AMADEUS, portanto deverá ser realizado um levantamento das arquiteturas implementadas em cada uma das ferramentas do ambiente. Embora essas arquiteturas sejam comentadas em trabalhos anteriores (Aluísio, 1995; Silva, 1999; Barros, 2000), é muito provável que o código contenha particularidades da implementação que não estão documentadas. Também através dessa análise, será possível avaliar a possibilidade de reaproveitamento de partes do código utilizado nas ferramentas do AMADEUS para as ferramentas deste projeto.

É importante destacar que, embora tomemos o AMADEUS como ponto de partida e pretendamos reutilizar as abordagens nele implementada, como o raciocínio baseado em casos, não temos o compromisso de reproduzir sua arquitetura tal como está. Somente ao longo do desenvolvimento do projeto computacional saberemos o grau de reutilização da implementação do AMADEUS.

(8) Implementação e Teste dos Protótipos.

A técnica de prototipagem deverá ser utilizada no desenvolvimento desse trabalho. Ela é apropriada na medida em que seu caráter iterativo permite avaliações parciais rápidas e incrementais. Dessa forma, esperamos ter uma versão estável do sistema computacional mais rapidamente, de forma que possa ser submetido à avaliação final (atividade (10)).

¹ *Graduate Management Admissions Test* – <http://www.gmat.org> (acessado em 05-03-2002)

² *Test of Written English* – <http://www.toefl.org> (acessado em 05-03-2002)

(9) Proposta de uma Metodologia de Avaliação das Ferramentas de Auxílio à Escrita.

A questão da avaliação de sistemas dessa natureza deverá merecer atenção especial nesse trabalho. A metodologia proposta deverá cobrir os variados aspectos envolvidos nesse tipo de sistema: a abrangência e balanceamento do corpus, a eficácia no processo de auxílio, o processo de julgamento do texto final³, questões de usabilidade da interface, entre outras. Para tanto, serão investigados trabalhos relacionados à avaliação de sistemas de PLN (Santos, 2000; Di Eugenio, 2000; Spark Jones, 1996; King, 1996) e à avaliação de usabilidade (Dix et al., 1999; Treu, 1994; Nielsen, 1993). Questões operacionais, como o número e tipos de usuários que participarão dos experimentos e o método de análise dos resultados, serão igualmente consideradas.

(10) Avaliação das Ferramentas de Auxílio à Escrita.

Trata-se da aplicação da metodologia proposta em (9) e conseqüente análise e discussão dos resultados.

(11) Publicação dos Resultados Obtidos.

A publicação dos resultados parciais e/ou finais desse trabalho em meios conceituados de divulgação científica será um objetivo constante, assim como sua apresentação em eventos científicos da área. Um primeiro resultado que deverá ser submetido à publicação será o resultado da atividade (6). Um resumo deste trabalho, focalizando o CorpusDT, deverá ser apresentado no 12º InPLA – Intercâmbio de Pesquisas em Linguística Aplicada – que será realizado em abril deste ano na PUC/SP.

(12) Redação da Tese.

Os capítulos da tese deverão ser redigidos à medida que os resultados forem sendo obtidos.

(13) Defesa da Tese.

5.2 Cronograma de Trabalho

A tabela a seguir apresenta o cronograma pretendido para as atividades serem cumpridas, divididas em semestres (4 anos, 8 semestres).

³ Esse julgamento deverá basear-se na rubrica desenvolvida como parte da atividade (6)

Lista Resumida de Atividades a Serem Cumpridas:

- (1) Pesquisa Bibliográfica
- (2) Finalização da Marcação do Corpus
- (3) Análise Crítica do Corpus
- (4) Projeto do Sistema Computacional
- (5) Implementação e Teste dos Protótipos
- (6) Proposta de uma Metodologia de Avaliação das Ferramentas de Auxílio à Escrita
- (7) Avaliação das Ferramentas de Auxílio
- (8) Publicação dos resultados obtidos
- (9) Redação da Tese
- (10) Defesa da Tese

Tabela 5.1. Cronograma de atividades previstas

Atividades \ Período	2002		2003		2004
	1º sem.	2º sem.	1º sem.	2º sem.	1º sem.
(1)					
(2)					
(3)					
(4)					
(5)					
(6)					
(7)					
(8)					
(9)					
(10)					

 Período Previsto

5.3 Viabilidade e Recursos Disponíveis

Este projeto será realizado no NILC, que dispõe de ótimas condições de laboratório computacional, recursos lingüísticos do português, dada sua experiência no desenvolvimento de sistemas e recursos lingüísticos para essa língua desde 1993, mas principalmente de recursos humanos, pois conta com uma grande equipe interdisciplinar formada por cientistas da computação e lingüistas. Vale ressaltar a participação da Profa. Sandra Aluísio, autora do AMADEUS e muito experiente nessa área, como co-orientadora desse trabalho.

Referências Bibliográficas

Aluísio, S.M. (1995). *Ferramentas para Auxiliar a Escrita de Artigos Científicos em Inglês como Língua Estrangeira*. Tese de Doutorado, IFSC-USP, 228 p.

Aluísio, S.M.; Oliveira Jr., O.N. (1995). A Case-Based Approach for Developing Writing Tools Aimed at Non-native English Users. In *Lecture Notes in Artificial Intelligence* 1010, pp. 121-132.

Aluisio, S.M.; Oliveira Jr., O.N. (1996). A Detailed Schematic Structure of Research Papers Introductions: An Application in Support-Writing Tools. In *Revista de la Sociedad Espanyola para el Procesamiento del Lenguaje Natural*, n. 19, pp. 141-147. Also available in <<http://www.cica.es/sepln96/sepln96.html>>

Aluísio, S.M.; Gantenbein, R.E. (1997). Towards the Application of Systemic Functional Linguistics in Writing Tools. In *Proceedings of International Conference on Computers and their Applications*, Arizona, pp. 181-185.

Aluisio, S.M.; Barcelos, I.; Sampaio, J.; Oliveira Jr., O. (2001). 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*, Madison/Wisconsin, pp. 257-260.

Barrass, R. (1979). *Os Cientistas Precisam Escrever: Guia de Redação para Cientistas, Engenheiros e Estudantes*. Ed. da Universidade de São Paulo, São Paulo.

Barros, R.C. (2000). *Modelagem de Usuários para Sistemas de Auxílio à Escrita Técnica*. Dissertação de Mestrado, ICMC-USP, 100p.

Booth, W.C; Colomb, G.G.; Williams, J.M. (2000). *A Arte da Pesquisa*. Martins Fontes, São Paulo.

Broady, E.; Shurville, S. (2000). Developing Academic Writer: Designing a Writing Environment for Novice Academic Writers. In E. Broady (ed.), *Second Language Writing in a Computer Environment*, CILT, London, pp. 131-151.

Castro, C.M. (1981). *A Prática da Pesquisa*. Ed. McGraw-Hill do Brasil, São Paulo.

Di Eugenio, B. (2000). On the Usage of Kappa to Evaluate Agreement on Coding Tasks. In *Proceedings of LREC'2000 Second International Conference on Language Resources and Evaluation*, Athens. (CD-ROM)

Dias da Silva, B. C.; Oliveira, M.F. (2002). Inclusão de Informação Pragmático-Discursiva na Base Lexical de um Thesaurus Eletrônico. *Estudos Lingüísticos*, 31, Cd-rom 6p. (no prelo).

Dix, A.; Finlay, J.; Abowd, G.; Beale, R. (1999). *Human-Computer Interaction*. 2ed., Prentice Hall Europe.

Eco, U. (2000). *Como Se Faz uma Tese*. Ed. Perspectiva, São Paulo.

Feitosa, V.C. (1991). *Redação de Textos Científicos*. Papirus Editora. Campinas.

Feltrim, V.D.; Aluísio, S.M.; Nunes, M.G.V. (2000). *Uma Revisão Bibliográfica sobre a Estruturação de Textos Científicos em Português*. Relatório Técnico, Série Computação, n.120, ICMC-USP, São Carlos.

Feltrim, V.D.; Nunes, M.G.V.; Aluísio, S.M. (2001). *Um corpus de textos científicos em Português para a análise da Estrutura Esquemática*. Série de Relatórios do NILC. NILC-TR-01-4.

Fischer, G.; Lemke, A.C.; Mastaglio, T. (1991). Critics: An Emerging Approach to Knowledge-Based Human-Computer Interaction. In *International Journal of Man-Machine Studies* 35, pp. 695-721.

- Fontana, N.; Caldeira, S.M.A.; De Oliveira, M.C.F.; Oliveira Jr., O.N. (1993). Computer Assisted Writing - Applications to English as a Foreign Language. In *CALL*, vol.6 (2), pp. 145-161.
- Hartley, A.; Paris, C. (1996). *Multilingual Document Production: From Support for Translating to Support for Authoring*. ITRI Technical Report Series, ITRI-96-17, Brighton.
- Hayes, J.R.; Flower, L.S. (1980). Writing as Problem Solving. In *Visible Language*, XIV (4).
- Hegenber, L. (1976). *Etapas da Investigação Científica*. Ed. Epu, São Paulo.
- Huckin, T.N.; Olsen, L.A. (1991). *Technical Writing and Professional Communication For Nonnative Speakers of English*. McGraw-Hill. New York.
- King, M. (1996). Evaluating natural language processing systems. In *Communications of the ACM* 39, pp. 73-79.
- Kock, I.V. (1998). *A Coesão Textual*. Coleção “Repensando a Língua Portuguesa”. 10ed., Editora Contexto, São Paulo.
- Kriegsman, M.; Barletta, R. (1993). Building a Case-based Help Desk Application. In *IEEE Expert*, December, pp. 18-26.
- Lansman, M.; Smith, J.B.; Weber, I. (1993). Using the WRITING ENVIRONMENT to Study Writer’s Strategies. In *Computers and Composition*, v.10, n. 2.
- Liddy, E.D. (1991). The Discourse-Level Structure of Empirical Abstracts: An Exploratory Study. In *Information Processing & Management*, v. 27, n. 1, pp. 55-81.
- Linden, K.V.; Scott, D. (1995). *Raising the Interlingual Ceiling with Multilingual Text Generation*. ITRI Technical Report Series, ITRI-95-12, Brighton.
- Mani, I. (2001). *Automatic Summarization*. John Benjamins Publishing Company. Amsterdam.

Mantaras, R.L.; Plaza, E. (1995). Case-Based Reasoning. In *The Newsletter of the European Network of Excellence in ML*, Special Issue, pp. 29-37.

McEnery, T.; Wilson, A. (1997). *Corpus Linguistics*. Edinburg University Press, Edinburg.

Narita, M. (2000a). Corpus-based English Language Assistant to Japanese Software Engineers. In *Proceedings of MT-2000 Machine Translation and Multilingual Applications in the New Millennium*. pp. 24-1 – 24-8.

Narita, M. (2000b). Constructing a Tagged E-J Parallel Corpus for Assisting Japanese Software Engineers in Writing English Abstracts. In *Proceedings of LREC'2000 Second International Conference on Language Resources and Evaluation*, Athens, pp. 1187-1191.

Nielsen, J. (1993). *Usability Engineering*. Ap Professional, Boston.

Oliveira Jr., O.N.; Aluísio Caldeira, S.M.; Fontana, N. (1992) Chusaurus: A Writing Tool Resource for Non-Native Users of English. In Ricardo Baeza-Yates and Udi Manber (eds.) *Computer Science: Research and Application*, New York: Plenum Press, pp. 63-72.

Oliveira, S.L. (2001). *Tratado de Metodologia Científica*. Pioneira, São Paulo.

Paizan, D.C. (2001). *O Uso da Linguagem da Internet na Produção de um Módulo para o Ensino de Leitura de Inglês Instrumental*. Dissertação de mestrado, FCLar–UNESP, 177p.

Paris, C.; Evans, R. (1994). *Preliminary Proposal for the GIST Architecture*. ITRI, Brighton.

Pemberton, L.; Shurville, S.; Hartley, A. (1996). Motivating the Design of a Computer Assisted Environment for Writers in a Second Language. In *Proceedings of CALICE'96*, pp.141-148.

Poesio, M. (2000). Annotating a Corpus to Develop and Evaluate Discourse Entity Realization Algorithms: Issues and Preliminary Results. In *Proceedings of LREC'2000 Second International Conference on Language Resources and Evaluation*, Athens. (CD-ROM)

Reference Software International (1992). *The #1 Grammar & Style checker Gram.Mat.Ik 5 (For Windows) User's Guide*. San Francisco, CA.

Rey, L. (1972). *Como Redigir Trabalhos Científicos*. Ed. da Universidade de São Paulo, São Paulo.

Santos, D. (2000). *Evaluation of Natural Language Processing Systems*. Tutorial presented at the Joint International Conference IBERAMIA/SBIA'2000, Atibaia, São Paulo, Brazil.

Severino, A.J. (1996). *Metodologia do Trabalho Científico*. Cortez Editora.

Sharples, M.; Goodlet, J.; Clutterbuck, A. (1994). A comparison of algorithms for hypertext notes network linearization. In *International Journal of Human-Computer Studies*, n.40, v.4, pp.727-752.

Sharples, M.; Pemberton, L. (1992) Representing writing: external representations and the writing process. In P.O. Holt and N. Williams (eds.) *Computers and Writing: State of the Art*. Intellect, Oxford, pp.319-336.

Shurville, S.; Hartley, A. F.; Pemberton, L. (1997). A Development Methodology for Composer: a Computer Support Tool for Academic Writing in a Second Language. In E-M. Jakobs, & D. Knorr (eds.), *Textproduktion in elektronischen Umgebungen [Text Production in Electronic Environments]*, Peter Lang Verlag, Frankfurt, pp. 171-182.

Silva, M.H.B. (1999). *A Abordagem de Críticas para a Construção de Sistemas de Aprendizagem da Escrita Técnica*. Dissertação de Mestrado. ICMC-USP, 130p.

Silva, M.H.B.; Pellizoni, J.M.; Aluísio, S.M. (1998). Uma abordagem híbrida baseada em críticas e casos para a construção de ferramentas colaborativas de ensino da escrita de artigos científicos. In *Anais do IX Simpósio Brasileiro de Informática na Educação*, Fortaleza. (CD-ROM)

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford University Press.

Smith, J.B.; Lansman, M. (1988). *A Cognitive Basis for a Computer Writing Environment*. Technical Report, n.87-032, Chapel Hill.

Solomon, D.V. (1977). *Como Fazer uma Monografia: Elementos de Metodologia do Trabalho Científico*. Ed. Interlivros, Belo Horizonte.

Spark Jones, K.; Galliers, J.R. (1996). Evaluating Natural Language Processing Systems: an Analysis and Review. In *Lecture Notes in Artificial Intelligence*, n. 1083, Springer.

Swales, J.M. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge applied linguistics series.

Tachizawa, T.; Mendes, G. (2000). *Como Fazer Monografia na Prática*. Ed. Fundação Getúlio Vargas, Rio de Janeiro.

Taylor, G.; Tingguang, C. (1991). Linguistic, Cultural, and Subcultural Issues in Contrastive Discourse Analysis: Anglo-American and Chinese Scientific Texts. In *Applied Linguistics*, v.12, n.3, pp.319-336.

Treu, S. (1994). *User Interface Evaluation: a Structured Approach*. Plenum Press. New York.

Trimble, L. (1985). *English for science and technology: a discourse approach*. Cambridge University Press.

Vieira, S. (1999). *Como Escrever uma Tese*. Ed. Pioneira, São Paulo.

XML (2001). XML Schema Part 0: Primer. *W3C Recommendation*. Disponível em <<http://www.w3.org/TR/xmlschema-0/>>, acessado em 10-07-2001.

Weissberg, R.; Buker, S. (1990). *Writing up Research: Experimental Research Report Writing for Students of English*. Prentice Hall.

Writing Tools Group, Inc. (1991). *Using Correct Grammar For Windows*. Sausalito, CA.

Apêndice A

***Trechos de Resumos do CorpusDT Exemplificando
os Subcomponentes e suas Estratégias Retóricas***

Os trechos de texto apresentados neste apêndice foram extraídos dos resumos do CorpusDT, com a finalidade de exemplificar os subcomponentes e as estratégias retóricas indicados no Quadro 4.1 do Capítulo 4. Alguns subcomponentes/estratégias retóricas presentes no quadro não são exemplificados por não terem sido encontradas ocorrências no corpus.

⇒ **Contexto**

C1. Declarar proeminência do tópico (argumentação)

“Tarefas envolvendo Reconhecimento de Padrões vêm se tornando mais freqüentes em diferentes domínios de aplicação. A maioria destas tarefas tem sido eficientemente tratada através da utilização de Redes Neurais Artificiais”.

C2. Familiarização de termos

“Robôs móveis dependem de dados provenientes de sensores para ter uma representação do seu ambiente”.

C3. Introduzir a pesquisa a partir da grande área

“Muitos dos sistemas computacionais atuais de apoio ao ensino podem ser considerados parte de uma evolução que tem enfatizado a exploração de sistemas hipermídia em geral, e da Web em particular”.

⇒ **Gap**

G1. Problemas/dificuldades

“Porém, os sensores geralmente fornecem informações incompletas, inconsistentes ou imprecisas”.

G2. Necessidades/requisitos

“Diante da diversidade das técnicas seletivas, fazem-se necessários estudos empíricos para avaliar e comparar a aplicação dessas técnicas”.

G3. Ausência ou pouca pesquisa anterior

“Observa-se que existem poucos trabalhos relacionados à injeção de defeitos de software na literatura, assim como modelos de defeitos e métodos de injeção relacionados”.

⇒ **Propósito**

P1. Indicar o propósito principal

“Este trabalho apresenta um estudo sobre a viabilidade de utilização do protocolo conservativo CMB para sincronização de simulações distribuídas em diferentes plataformas MIMD com memória distribuída, utilizando granulosidade grossa com poucos processos paralelos”.

P2. Re-frasear/detalhar/especificar o propósito

“A meta deste estudo não é apontar qual método é o mais eficiente em termos gerais, mesmo porque acreditamos que isto não seja possível. Queremos, sim, observar o desempenho de ambos os métodos quanto ao problema determinado, visando assim uma maior integração entre eles, aproveitando suas melhores potencialidades”.

P3. Introduzir mais propósitos

“Estratégias para evoluir e refinar um conjunto essencial para diferentes domínios de aplicação também são investigadas”.

⇒ Metodologia

M1. Listar critérios ou condições

Não foram encontrados exemplos no corpus.

M2. Descrever materiais e métodos

“Usando como estudo de caso as regras gramaticais da crase, tomamos como exemplo de modelo tradicional de implementação o revisor gramatical ReGra, e de outro lado, implementamos dois modelos de redes neurais (um modelo backpropagation e um modelo Elman), para detectar erros com relação ao uso da crase, tanto em casos de presença incorreta, quanto de ausência”.

M3. Justificar a escolha pelos materiais e métodos

Não foram encontrados exemplos no corpus.

⇒ Resultado

R1. Descrição do artefato

“Esse procedimento compreende um processo para a criação de uma base de conhecimento (Processo de Aquisição de Conhecimento IPAIA, instanciado para o domínio de engenharia reversa) e diretrizes para a utilização dessa base de conhecimento na construção de visões funcionais do sistema”.

R2. Apresentar os resultados

“A maior precisão geral obtida foi a do MXPOST - 89,66%”.

⇒ Conclusão

Co1. Recomendação

Não foram encontrados exemplos no corpus.

Co2. Comentários/discussão (generalizar/explicar/comparar) dos resultados

“A precisão geral sofreu a influência do tamanho do corpus manualmente etiquetado disponível para treinamento, do conjunto de etiquetas e dos tipos de texto utilizados”.

Co3. Valor do trabalho

“Com a ajuda dessas diretrizes e a disponibilidade da extensão ParSMPL espera-se auxiliar o usuário a desenvolver aplicações eficientes utilizando o paradigma da simulação distribuída conservativa, sem exigir desse usuário o conhecimento das características e particularidades do protocolo de sincronização utilizado”.

Apêndice B

XML Schema para a Anotação do CorpusDT

<!-- edited with XML Spy v4.0.1 U (<http://www.xmlspy.com>) by Valéria Feltrim (USP - NILC) -->

<xsd:schema xmlns:xsd="http://www.w3.org/2001/XMLSchema">

<xsd:annotation>

<xsd:documentation xml:lang="pt">

XML Schema para anotacao de teses e dissertacoes (01-09-01)

Copyright 2001 Valeria D. Feltrim. All rights reserved.

</xsd:documentation>

</xsd:annotation>

<xsd:element name="Dissertacao" type="TipoDissertacao"/>

<xsd:complexType name="TipoDissertacao">

<xsd:sequence>

<xsd:element name="PaginaRosto" type="TipoPaginaRosto"/>

<xsd:element name="ElemIniciais" type="TipoElemIniciais" minOccurs="0"/>

<xsd:element name="CorpoDissertacao" type="TipoCorpoDissertacao"/>

<xsd:element name="RefBibliografica" type="xsd:string" minOccurs="0"/>

<xsd:element name="Bibliografia" type="xsd:string" minOccurs="0"/>

<xsd:element name="Anexo" type="TipoAnexo" minOccurs="0" maxOccurs="unbounded"/>

<xsd:element name="Apendice" type="TipoApendice" minOccurs="0" maxOccurs="unbounded"/>

<xsd:element name="Rodape" type="xsd:string" minOccurs="0" maxOccurs="unbounded"/>

</xsd:sequence>

<xsd:attribute name="Nivel" type="TipoNivel" use="required"/>

<xsd:attribute name="Area" type="TipoArea" use="required"/>

</xsd:complexType>

<xsd:simpleType name="TipoArea">

<xsd:restriction base="xsd:string">

<xsd:enumeration value="ES"/>

<xsd:enumeration value="IC"/>

<xsd:enumeration value="BD"/>

<xsd:enumeration value="HIP"/>

<xsd:enumeration value="CGPI"/>

<xsd:enumeration value="SDG"/>

<xsd:enumeration value="SDPC"/>

</xsd:restriction>

</xsd:simpleType>

<xsd:simpleType name="TipoNivel">

<xsd:restriction base="xsd:string">

<xsd:enumeration value="MS"/>

<xsd:enumeration value="DO"/>

</xsd:restriction>

</xsd:simpleType>

```
<xsd:complexType name="TipoPaginaRosto">
```

```
  <xsd:all>
```

```
    <xsd:element name="Instituicao" type="xsd:string" minOccurs="0"/>
```

```
    <xsd:element name="Titulo" type="xsd:string" minOccurs="0"/>
```

```
    <xsd:element name="Autor" type="xsd:string" minOccurs="0"/>
```

```
    <xsd:element name="Orientador" type="xsd:string" minOccurs="0"/>
```

```
    <xsd:element name="Comentario" type="xsd:string" minOccurs="0"/>
```

```
    <xsd:element name="Local" type="xsd:string" minOccurs="0"/>
```

```
    <xsd:element name="Data" type="xsd:string" minOccurs="0"/>
```

```
  </xsd:all>
```

```
</xsd:complexType>
```

```
<xsd:complexType name="TipoElemIniciais">
```

```
  <xsd:all>
```

```
    <xsd:element name="Citacao" type="xsd:string" minOccurs="0"/>
```

```
    <xsd:element name="Dedicatoria" type="xsd:string" minOccurs="0"/>
```

```
    <xsd:element name="Agradecimentos" type="xsd:string" minOccurs="0"/>
```

```
    <xsd:element name="Sumario" type="xsd:string" minOccurs="0"/>
```

```
    <xsd:element name="ListadeFigs" type="xsd:string" minOccurs="0"/>
```

```
    <xsd:element name="ListadeTabs" type="xsd:string" minOccurs="0"/>
```

```
    <xsd:element name="ListadeGraficos" type="xsd:string" minOccurs="0"/>
```

```
    <xsd:element name="ListadeAbrev" type="xsd:string" minOccurs="0"/>
```

```
    <xsd:element name="Resumo" type="TipoResumo" minOccurs="0"/>
```

```
    <xsd:element name="Abstract" type="TipoResumo" minOccurs="0"/>
```

```
    <!-- "Citacao" pode ser uma frase, uma poesia, um conto, etc -->
```

```
  </xsd:all>
```

```
</xsd:complexType>
```

```
<xsd:complexType name="TipoAnexo" mixed="true">
```

```
  <xsd:sequence>
```

```
    <xsd:element name="Elemento" type="TipoElemento" minOccurs="0" maxOccurs="unbounded"/>
```

```
  </xsd:sequence>
```

```
</xsd:complexType>
```

```
<xsd:complexType name="TipoApendice" mixed="true">
```

```
  <xsd:sequence>
```

```
    <xsd:element name="Elemento" type="TipoElemento" minOccurs="0" maxOccurs="unbounded"/>
```

```
  </xsd:sequence>
```

```
</xsd:complexType>
```



```

<xsd:complexType name="TipoCorpoDissertacao">
  <xsd:sequence>
    <xsd:element name="Introducao" type="TipoIntroducao"/>
    <xsd:element name="Capitulo" type="TipoCapitulo" maxOccurs="unbounded"/>
    <xsd:element name="Conclusão" type="TipoConclusao"/>
  </xsd:sequence>
</xsd:complexType>

<xsd:complexType name="TipoIntroducao">
  <xsd:sequence>
    <xsd:element name="Titulo" type="xsd:string"/>
    <xsd:element name="SecaoI" type="TipoSecaoI" maxOccurs="unbounded"/>
  </xsd:sequence>
</xsd:complexType>

<xsd:complexType name="TipoConclusao">
  <xsd:sequence>
    <xsd:element name="Titulo" type="xsd:string"/>
    <xsd:element name="SecaoC" type="TipoSecaoC" maxOccurs="unbounded"/>
  </xsd:sequence>
</xsd:complexType>

<xsd:complexType name="TipoCapitulo">
  <xsd:sequence>
    <xsd:element name="Titulo" type="xsd:string"/>
    <xsd:element name="Secao" type="TipoSecao" maxOccurs="unbounded"/>
  </xsd:sequence>
</xsd:complexType>

<xsd:complexType name="TipoSecaoI">
  <xsd:sequence>
    <xsd:element name="Titulo" type="xsd:string" minOccurs="0"/>
    <xsd:element name="Bloco" type="TipoBloco" minOccurs="0" maxOccurs="unbounded"/>
    <xsd:element name="SubsecaoI" type="TipoSecaoI" minOccurs="0" maxOccurs="unbounded"/>
  </xsd:sequence>
</xsd:complexType>

<xsd:complexType name="TipoSecaoC">
  <xsd:sequence>
    <xsd:element name="Titulo" type="xsd:string" minOccurs="0"/>
    <xsd:element name="Bloco" type="TipoBloco" minOccurs="0" maxOccurs="unbounded"/>
    <xsd:element name="SubsecaoC" type="TipoSecaoI" minOccurs="0" maxOccurs="unbounded"/>
  </xsd:sequence>
</xsd:complexType>

```

```
<xsd:complexType name="TipoSecao" mixed="true">
  <xsd:sequence>
    <xsd:element name="Titulo" type="xsd:string" minOccurs="0"/>
    <xsd:element name="Elemento" type="TipoElemento" minOccurs="0" maxOccurs="unbounded"/>
    <xsd:element name="Subsecao" type="TipoSecao" minOccurs="0" maxOccurs="unbounded"/>
  </xsd:sequence>
</xsd:complexType>

<xsd:complexType name="TipoElemento">
  <xsd:choice>
    <xsd:element name="Figura" type="xsd:string"/>
    <xsd:element name="Tabela" type="xsd:string"/>
    <xsd:element name="Grafico" type="xsd:string"/>
    <xsd:element name="Quadro" type="xsd:string"/>
    <xsd:element name="Equacao" type="xsd:string"/>
    <xsd:element name="Rodape" type="xsd:string"/>
  </xsd:choice>
</xsd:complexType>

<xsd:complexType name="TipoBloco">
  <xsd:choice>
    <xsd:element name="Subcomponente" type="TipoSubcomponente"/>
    <xsd:element name="Figura" type="xsd:string" minOccurs="0"/>
    <xsd:element name="Tabela" type="xsd:string" minOccurs="0"/>
    <xsd:element name="Grafico" type="xsd:string" minOccurs="0"/>
    <xsd:element name="Quadro" type="xsd:string" minOccurs="0"/>
    <xsd:element name="Equacao" type="xsd:string" minOccurs="0"/>
    <xsd:element name="Rodape" type="xsd:string" minOccurs="0"/>
  </xsd:choice>
</xsd:complexType>

<xsd:complexType name="TipoSubcomponente">
  <xsd:sequence>
    <xsd:element name="Nome" type="xsd:string"/>
    <xsd:element name="Estrategia" type="TipoEstrategia" minOccurs="0" maxOccurs="unbounded"/>
  </xsd:sequence>
</xsd:complexType>

<xsd:complexType name="TipoEstrategia" mixed="true">
  <xsd:sequence>
    <xsd:element name="Nome" type="xsd:string"/>
    <xsd:element name="Reutilizavel" type="TipoReutilizavel" minOccurs="0" maxOccurs="unbounded"/>
  </xsd:sequence>
</xsd:complexType>
```

```
<xsd:complexType name="TipoReutilizavel" mixed="true">
  <xsd:sequence>
    <xsd:element name="Marcador" type="xsd:string" minOccurs="0" maxOccurs="unbounded"/>
  </xsd:sequence>
</xsd:complexType>

<xsd:complexType name="TipoResumo">
  <xsd:sequence>
    <xsd:element name="Subcomponente" type="TipoSubcomponente" maxOccurs="unbounded"/>
  </xsd:sequence>
</xsd:complexType>

</xsd:schema>
```


Apêndice C

***Questionário Aplicado aos Potenciais Usuários de
uma Ferramenta de Auxílio à Escrita Acadêmica***

1. Como você classificaria o seu conhecimento sobre a redação de textos acadêmicos (p.e. relatório técnico, artigo, monografias de qualificação, dissertação, tese, projeto científico) quanto à estrutura geral das seções/capítulos, seus componentes e suas funções?
() Excelente () Bom () Razoável () Pouco () Insuficiente
2. Quantos textos desse tipo você já produziu individualmente?
() Nenhum () Até 5 () Até 10 () Mais que 10
3. Considerando as seguintes seções de um texto acadêmico típico, assinale seu nível de dificuldade ao escrevê-las. Use uma escala de 1 a 5 para indicar o grau de dificuldade, sendo que 1 indica maior dificuldade.
() Resumo () Abstract () Introdução () Revisão Bibliográfica
() Metodologia () Resultados () Conclusão
4. Quantas revisões, além das suas, são feitas em média em seus textos acadêmicos até que você os considere prontos?
() Nenhuma () 1 () 2 () Outro. Neste caso, indique o número. _____
5. Qual o grau de interferência de alguém mais experiente (p.e. seu orientador) na produção de seus textos acadêmicos?
() Nenhuma () Pouca () Significativa () Muita
6. Você se baseia em algum modelo de texto para produzir o seu?
() Não
() Sim. Neste caso, qual?
() Texto anterior seu
() Texto de um colega
() Texto anterior de seu orientador ou de quem foi orientado por ele
() Texto acadêmico da literatura nacional
() Texto acadêmico da literatura em inglês

1. Indispensável	<i>sem essa característica não valeria a pena usar a ferramenta</i>
2. Desejável	<i>a presença da característica motivaria o uso da ferramenta</i>
3. Interessante	<i>a característica é interessante, mas não influenciaria sua motivação em usar a ferramenta</i>
4. Dispensável	<i>a ferramenta não perderia valor sem essa característica</i>

7. () Apoio à organização das idéias através de *brainstorming* livre (sem interferência do sistema), por exemplo, fornecendo um espaço onde o usuário possa escrever suas idéias e relacioná-las (p.e. um grafo com nós e arcos).
8. () Apoio à organização das idéias através de *brainstorming* dirigido (p.e., através de perguntas que estimulem a geração das idéias iniciais) com o foco nos principais componentes de um texto acadêmico.
9. () Disponibilidade de explicações sobre o gênero acadêmico, indicando o papel de cada componente de uma seção/capítulo.
10. () Disponibilidade de modelos de textos acadêmicos, indicando os possíveis componentes e a ordem em que aparecem.
11. () Disponibilidade de exemplos autênticos de bons textos acadêmicos que instanciem os modelos.
12. () Disponibilidade de *guidelines* sobre como escrever bem cada componente.
13. () Disponibilidade de padrões de escrita para cada componente, i.e., expressões comumente utilizadas em um determinado componente.
14. () Alerta do sistema sobre erros comumente encontrados na escrita de cada componente.
15. () Disponibilidade de uma *checklist* para que o usuário possa verificar se o seu texto apresenta as principais características de um bom texto acadêmico.
16. () Apoio à correção ortográfica.
17. () Apoio à correção gramatical e de estilo.
18. () Revisão do texto depois de pronto com sugestões de melhoria.
19. Existem características que não foram listadas aqui que você considera importantes? Em caso afirmativo, faça sua sugestão sem se preocupar com a viabilidade da implementação.
- _____
- _____
- _____
- _____
20. Assinale qual o **estilo de interação com o usuário** de sua preferência para a ferramenta para auxiliar a escrita de um texto acadêmico
- () Um ambiente tutor, sob o controle do sistema, que especifica uma sequência de passos para a escrita do texto.
- () Um ambiente livre, em que o usuário aciona as funções do sistema (solicitando ajuda) quando achar necessário.
- () Um ambiente misto, em que o usuário é livre para acionar funções do sistema e o sistema também interfira quando julgar necessário.
- () Nenhuma das anteriores. Neste caso, faça sua sugestão.
- _____
- _____
- _____
- _____

21. Indique a sua **motivação** em usar a ferramenta para auxiliar a escrita de um texto acadêmico acoplada ao seu processador de texto.

- () Não a usaria caso não fosse acoplada ao editor de texto
- () Usaria mesmo não sendo acoplada ao editor de texto, desde que o formato dos arquivos sejam compatíveis.
- () Usaria mesmo não sendo acoplada ao editor de texto, pois essa característica é irrelevante.
- () Não gostaria de uma ferramenta acoplada ao editor de texto.