e-Termos: Um ambiente colaborativo web de gestão
terminológica

Leandro Henrique Mendonça de Oliveira

e-Termos: Um ambiente colaborativo web de gestão terminológica

Leandro Henrique Mendonça de Oliveira

Orientadora: *Prof^a Dr^a Sandra Maria Aluísio*

Tese apresentada ao Instituto de Ciências Matemáticas e de Computação - ICMC-USP como parte dos requisitos para obtenção do título de Doutor em Ciências de Computação e Matemática Computacional.

USP - São Carlos
Agosto/2009

Resumo

Em uma de suas definições, a Terminologia representa o conjunto de princípios e métodos adotados no processo de gestão e criação de produtos terminológicos, tais como glossários e dicionários de termos. A sistematização desses métodos envolve a aplicação de ferramentas computacionais específicas e compatíveis com as tarefas terminológicas, contribuindo para o desenvolvimento desses produtos e a difusão de conhecimento especializado. Entretanto, principalmente no Brasil, a combinação da Terminologia e Informática é incipiente, e dentre as atividades do trabalho terminológico é comum a utilização de várias ferramentas não especializados para esse fim. Isso torna o trabalho dos terminólogos muito moroso, pois esse trabalho geralmente é feito por uma equipe multidisciplinar que deve ter acesso, a todo o momento, à versão mais atual das várias etapas da geração de um produto terminológico. Além disso, deixa o gerenciamento dos dados mais complicado, pois não existe um padrão de entrada e saída definido para os programas. Apoiado nos pressupostos da Teoria Comunicativa da Terminologia (TCT), este trabalho apresenta a proposta de desenvolvimento e avaliação do **e-Termos**, um Ambiente Colaborativo Web composto por seis módulos de trabalho bem definidos, cujo propósito é automatizar as tarefas de gestão e criação de produtos terminológicos. Cada módulo do **e-Termos** possui a responsabilidade de abrigar tarefas inerentes ao processo de criação das terminologias, sendo atreladas a eles diferentes ferramentas de apoio lingüístico, que possuem a função de dar suporte às atividades de Processamento de Língua Natural envolvidas nesse processo. Além delas, há também ferramentas colaborativas, designadas para dar apoio às necessidades comunicacionais e de interação da equipe de trabalho. Particularmente com relação ao processo de avaliação proposto, uma de suas características é a capacidade de ser executado em um tempo curto, viabilizando a avaliação controlada de vários grupos, mas executada no ambiente de trabalho do público alvo. As principais contribuições desta pesquisa são o aspecto colaborativo instanciado na prática terminológica, a criação flexível da Ficha Terminológica, a possibilidade didática de uso para o ensino de terminologia, lexicografia e tradução e o processo de avaliação para sistemas colaborativos desenvolvido para o **e-Termos**, que combina *Cenários de Uso* e um *Questionário de Pesquisa*. Utilizando tecnologias Web e da área de *Computer Supported Collaborative Work (CSCW)* para o desenvolvimento da sua arquitetura computacional colaborativa, o **e-Termos** apresenta-se como um ambiente inovador para a pesquisa terminológica assistida por computador, pois automatiza um método prático que expõe os postulados da terminologia de orientação descritiva e evidencia todas as etapas do processo de criação de produtos terminológicos com o inédito diferencial colaborativo. Para certificar este êxito, o **e-Termos** tem recebido um número crescente de novas propostas de projeto, tendo até Agosto de 2009 mais de 130 usuários cadastrados, alocados em 68 diferentes projetos terminológicos.

Abstract

In one of its definitions, Terminology represents the set of principles and methods adopted in the creation and management of terminological products as glossaries and dictionaries of terms. A systematization of these methods includes the application of specific computational tools, compatible with terminological tasks, which contribute to developing such products and disseminating expert knowledge. However, especially in Brazil, the combination of Terminology and Computer Science is still incipient, and to perform the tasks of a terminological work it is typical to employ several nonspecialized tools, which make terminologists' work very time-consuming, since it is usually carried out by a multidisciplinary team that should have access, all the time, to the latest versions of the various stages of the generation of a terminological product. Moreover, it makes data management more complex, because there is no input/output standard defined for programs. Based on the presuppositions of the Communicative Theory of Terminology (CTT), this thesis proposes the development and evaluation of **e-Termos**, a Web Collaborative Environment composed of six well-defined working modules, whose purpose is to automatize tasks for creating and managing terminological products. Each module in **e-Termos** is responsible for tasks inherent to the process of creating terminologies. Linked to these modules, there are different linguistic support tools that assist the Natural Language Processing activities included in the process. Besides them, there are also collaborative tools for supporting the communication and interaction needs of team members. As far as the proposed evaluation process is concerned, one of its features is that it can be run in a short time, making viable a controlled evaluation of several groups that is, however, run in the work environment of the target audience. The main contributions of this research are the collaborative aspect instantiated in terminological practice, the flexible creation of Terminological Records, the possibility of being used for teaching terminology, lexicography and translation, and the evaluation of collaborative systems developed for **e-Termos**, which combines *Scenario-based Evaluations* and *Surveys*. Using Web technologies and *Computer Supported Collaborative Work (CSCW)* to develop its collaborative computational architecture, **e-Termos** is an innovative environment for computer-assisted terminological research, since it automatizes a useful method that represents the postulates of descriptive terminology and highlights all stages of the process of creating terminological products with the unprecedented collaborative differential. Confirming its success, **e-Termos** has been receiving a growing number of new project proposals, and in August 2009 has more than 130 registered users in 68 different terminological projects.

Esta tese foi preparada com o formatador de textos L^AT_EX. Foi utilizado um estilo (*style*) desenvolvido por Ronaldo Cristiano Prati. O sistema de citações de referências bibliográficas utiliza o padrão *Apalike* do sistema BibT_EX.

Sumário

Resumo	v
Abstract	vi
Sumário	xi
Lista de Figuras	xiii
Lista de Tabelas	xv
1 Introdução	1
1.1 Contextualização	1
1.2 Objetivos	8
1.3 Contribuições	10
1.4 Organização do trabalho	11
2 Princípios teóricos da Terminologia	13
2.1 Definições Básicas	14
2.2 Breve histórico da Terminologia	15
2.2.1 Escolas e Teorias	17
2.2.2 Histórico da Terminologia no Brasil: grupos de estudo	20
2.3 Interfaces da Terminologia	22
2.3.1 Lexicologia	22
2.3.2 Lexicografia	23
2.3.3 Terminografia	23
2.3.3.1 As aplicações terminológicas informatizadas	24
2.3.4 Tradução	25
2.4 Objetos da Terminologia	26
2.4.1 O Termo	27
2.4.2 A Definição Terminológica	29
2.4.3 Os Textos Especializados	32
2.4.4 Glossários, dicionários e bases de dados terminológicos	32
2.5 Considerações finais	34
3 Etapas do trabalho terminológico: descrição e problemas	35
3.1 Etapa 01 - Compilação de corpus de especialidade	37
3.1.1 Visão geral de corpus	37
3.1.2 O uso de corpus de especialidade informatizados em terminologia	39
3.1.3 Compilação e análise de corpus de especialidade informatizado	42
3.1.4 Ferramentas computacionais para análise de corpus de especialidade	46
3.1.5 Compilação automática de corpus de especialidade	49
3.1.6 Questões relacionadas com os direitos autorais	52
3.1.7 Problemas referentes à compilação de corpus de especialidade	52
3.2 Etapa 02 - Extração de candidatos a termos	53
3.2.1 Unidades terminológicas e suas dimensões	54
3.2.1.1 A dimensão lingüística dos termos	54
3.2.1.2 A dimensão conceitual dos termos	55
3.2.1.3 A dimensão comunicativa dos termos	56

3.2.2	O uso de córpus de especialidade para extração automática de termos . . .	56
3.2.3	A extração automática de candidatos a termos	59
3.2.3.1	O método estatístico	61
3.2.3.2	O método lingüístico	62
3.2.3.3	O método híbrido	64
3.2.4	Padrões de Formação dos Termos	65
3.2.5	Problemas relacionados com a extração automática de termos	66
3.3	Etapa 03 - Criação da Estrutura Conceitual do Domínio (Ontologia)	67
3.3.1	Ontologias e a terminologia	68
3.3.2	Linguagens e formatos de representação de ontologias	70
3.3.2.1	Ontologias e modelos de banco de dados relacionais	71
3.3.3	Metodologia para construção de ontologias	73
3.3.4	Ferramentas para construção de ontologias	74
3.3.5	Problemas relacionados com criação da estrutura conceitual do domínio .	75
3.4	Etapa 04 - Inclusão dos candidatos a termos na ontologia e a validação do especialista	77
3.4.1	Problemas relacionados às tarefas de inclusão e validação dos termos can- didatos	78
3.5	Etapa 05 - Criação da ficha terminológica e elaboração da base definicional . . .	80
3.5.1	Criação da Ficha Terminológica	80
3.5.2	Elaboração da Base Definicional	81
3.5.2.1	Normas ISO para redação das definições	84
3.5.3	Problemas relacionados com a criação da Ficha Terminológica e a elabo- ração da Base Definicional	84
3.6	Etapa 06 - Edição dos verbetes e uso de formatos para facilitar o intercâmbio e difusão dos produtos terminológicos	86
3.6.1	Edição dos verbetes	86
3.6.2	Intercâmbio e difusão dos termos	87
3.6.2.1	Formatos e padrões de intercâmbio e modelo de dados	88
3.6.3	Problemas relacionados com a edição dos verbetes e o intercâmbio de pro- dutos terminológicos	91
4	O Ambiente colaborativo Web e-Termos	93
4.1	Suporte teórico	94
4.2	Aspectos gerais do e-Termos	95
4.2.1	Perfil dos usuários do e-Termos	97
4.2.2	O ambiente modular	98
4.3	Descrição dos módulos de trabalho do e-Termos	104
4.3.1	Módulo Principal - Gestão terminológica	105
4.3.1.1	Gestão terminológica	105
4.3.1.2	Suporte a colaboração	107
4.3.2	Módulo 1 - Compilação automática de córpus	108
4.3.2.1	O <i>Portal de Córpus</i> do Projeto PLN-Br	109
4.3.3	Módulo 2 - Compilação e Suporte para Análise de Córpus	112
4.3.3.1	A interface Web de compilação de córpus e aplicação de ferra- mentas lingüísticas	113
4.3.4	Módulo 3 - Extração Automática de Candidatos a Termos	116
4.3.4.1	Métodos de extração no trabalho <i>Bloc-Eco</i>	119
4.3.4.2	Métodos estatísticos de extração de candidatos a termos	120
4.3.5	Módulo 4 - Edição do Mapa Conceitual e Categorização dos Termos . . .	121
4.3.5.1	Ferramentas de edição e visualização de ontologias	124
4.3.5.2	A interface Web para edição de ontologias no e-Termos – O OntoEditor	126
4.3.6	Módulo 5 - Criação e Gerenciamento de uma base de dados terminológica	131
4.3.6.1	Banco de Dados terminológicos	131
4.3.6.2	Sistemas de Gerenciamento do Banco de Dados terminológicos .	133
4.3.6.3	A população da Base Definicional	136
4.3.6.4	A criação flexível da Ficha Terminológica	138

4.3.7	Módulo 6 - Edição dos Verbetes e Uso de Formatos de Intercâmbio dos Produtos Terminológicos	143
4.3.7.1	A criação dos Modelos de Verbetes e Verbetes	143
4.3.7.2	A pré-publicação e publicação dos produtos terminológicos	144
4.3.7.3	A exportação de dados em formato TBX	148
4.4	O e-Termos como Ambiente Colaborativo Web	148
4.4.1	Ferramentas que favorecem a colaboração no Ambiente e-Termos	149
4.5	Aspectos e discussões relativas ao desenvolvimento e implementação do e-Termos	151
4.5.1	A escolha da Web como ambiente de desenvolvimento do e-Termos	153
4.5.2	O modelo de desenvolvimento em camadas	153
4.5.2.1	O modelo <i>Model Control View (MVC)</i>	154
4.6	Considerações finais	157
5	Sistemas colaborativos: métodos de pesquisa e avaliação	159
5.1	Aspectos gerais dos métodos de pesquisa e avaliação	161
5.1.1	Planos de pesquisa científica	162
5.1.2	Experimentos de Laboratório e de Campo	163
5.1.3	Natureza dos dados	163
5.1.4	Abordagens de Coleta de Dados	164
5.1.4.1	Observação	165
5.1.4.2	Entrevistas e/ou Questionários	165
5.1.5	Mensuração e Escala	166
5.2	Aspectos Gerais dos Sistemas Colaborativos	169
5.2.1	Principais características	170
5.2.2	Aplicações e benefícios de sistemas colaborativos	171
5.3	Avaliações nos Sistemas Colaborativos	173
5.3.1	Princípios e Dificuldades da Avaliação Colaborativa	173
5.3.2	Classificação e Organização dos Métodos de Avaliação	175
5.4	Cenários de Uso na Avaliação Colaborativa	178
5.4.1	Principais características e aplicação dos Cenários de Uso	178
5.4.2	Criação dos Cenários de Uso	179
5.5	Considerações finais	180
6	Avaliação do e-Termos	181
6.1	Avaliação da qualidade e da colaboração no e-Termos	181
6.1.1	Método de avaliação para o e-Termos : seleção e justificativas	182
6.1.2	Componentes da avaliação da colaboração no e-Termos	184
6.1.2.1	O Cenário de Uso	184
6.1.2.2	Questionário de Pesquisa	187
6.1.2.3	Grupos de usuários avaliadores	188
6.1.3	O método do <i>Valor Percebido</i> como componente da avaliação	190
6.2	Aplicação do Cenário de Uso e execução da avaliação	190
6.3	Descrição, análise e interpretação dos resultados	191
6.3.1	Conjunto de dados da avaliação do e-Termos	191
6.3.2	Interpretação e análise dos resultados da avaliação	192
6.3.2.1	Análise pelas Distribuições de Frequência	192
6.3.2.2	Análise pelas Correlações	206
6.4	Considerações Finais	219
7	Conclusão e trabalhos futuros	221
7.1	Conclusão	221
7.2	Contribuições	225
7.3	Limitações	227
7.4	Trabalhos futuros	228
	Referências	238
	A Caderno de Atividades Perfil Gerente de Projeto	239

B	Caderno de Atividades Perfil Terminólogo/Linguista	275
C	Caderno de Atividades Perfil Especialista do Domínio	299
D	Questionário de Pesquisa	311
E	Distribuição de frequência das questões do Questionário de Pesquisa	316

Lista de Figuras

3.1	Possíveis classificações entre língua geral e língua de especialidade Hernández (2002). . .	40
3.2	Tela inicial do <i>TACTWeb</i>	47
3.3	Tela inicial do <i>AneLL</i>	48
3.4	Tela inicial do <i>Word Smith Tools</i> (versão 4.0)	48
3.5	Localização dos erros de <i>silêncio</i> e <i>ruído</i> no processo de extração automática de termos.	62
3.6	Exemplos de verbetes do <i>Glossário de termos neológicos da economia</i>	87
4.1	Página principal do Portal do e-Termos	96
4.2	Exemplo de tela do Sistema e-Termos	96
4.3	Esquema geral modular da Parte Operacional do e-Termos	99
4.4	Subprojetos vinculados ao Projeto PLN-Br.	110
4.5	Relação de confiança entre os sistemas e-Termos e Portal de Córpus	111
4.6	Tela de pesquisa de texto do Portal de Córpus	111
4.7	Tela de resultado da montagem de subcórpus e de envio para o e-Termos	112
4.8	Tela de recebimento de subcórpus do e-Termos	112
4.9	Tela de gerência de textos.	114
4.10	Tela de compilação de córpus.	114
4.11	Tela de gerencia de córpus.	115
4.12	Tela de edição de uma Lista de Termos.	117
4.13	Tela de concatenação de Listas de Termos.	118
4.14	Menu de seleção dos métodos de extração automática de candidatos a termos.	120
4.15	Tela que apresenta as opções de extração de candidatos a termos do e-Termos	120
4.16	Tela de inclusão de nova Relação Específica.	124
4.17	Menu do e-Termos com as funcionalidades do <i>OntoEditor</i>	127
4.18	Estruturas da ontologia (EXEMPLO DE ONTOLOGIA)	128
4.19	Visualização hiperbólica da ontologia de Revestimentos Cerâmicos e o destaque para busca do termo “Mistura”.	129
4.20	Visualização em grafos da ontologia de Revestimentos Cerâmicos e o destaque para o termo “Mistura”.	129
4.21	Tela de categorização de termos na ontologia.	130
4.22	Tela de inclusão de um novo excerto definicional.	137
4.23	Tela do excerto do termo “Agitador”.	137
4.24	Menu das funcionalidades da Ficha Terminológica.	138
4.25	Tela de criação e edição da Ficha Terminológica.	139
4.26	Tela de adição de novos campos da Ficha Terminológica.	139
4.27	Tela de exclusão e adição de novos campos da Ficha Terminológica.	139
4.28	Tela de exclusão da Ficha Terminológica.	140
4.29	Tela do <i>Editor de Definição</i>	142
4.30	Exemplos de Modelos de Verbetes.	144
4.31	Verbetes formatados a partir dos Modelos de Verbetes.	144
4.32	Tela de criação do Modelo de Verbetes.	145
4.33	Tela de entrada da pré-publicação.	145
4.34	<i>Template</i> de publicação de produtos do e-Termos	146

4.35	Termo “eletrocardiograma” na visualização hipertexto.	146
4.36	Termo “eletrocardiograma” na visualização hiperbólica.	147
4.37	Termo “eletrocardiograma” na visualização em grafos.	147
4.38	Representação de uma aplicação monolítica - uma única camada.	153
4.39	Representação de uma aplicação em duas camadas, com a separação dos dados.	154
4.40	Representação de uma aplicação em três camadas, com a separação das camadas de apresentação, negócios e dados.	154
4.41	Arquitetura computacional modular em MVC do e-Termos	156
5.1	Fases de um Plano de Pesquisa.	161
5.2	Questão com respostas em escala nominal	167
5.3	Questão com respostas em escala ordinal	168
5.4	Questão com respostas em escala intervalar	169
6.1	Ícones utilizados nos cadernos de atividades.	185
6.2	Descrição de uma tarefa no caderno de atividades.	186
6.3	Descrição de tarefas que estimulam a comunicação da equipe de trabalho.	186
6.4	Distribuições de Frequências das Questões 8a e 8b	193
6.5	Distribuições de Frequências das Questões 8h e 8i	194
6.6	Distribuições de Frequências das Questões 8c e 8f	194
6.7	Distribuições de Frequências das Questões 9e e 8e	195
6.8	Distribuições de Frequências das Questões 7a e 7b	196
6.9	Distribuições de Frequências das Questões 7c e 7e	196
6.10	Distribuições de Frequências das Questões 7i e 7g	197
6.11	Distribuições de Frequências das Questões 7k e 7l	197
6.12	Distribuição de frequências das respostas sobre o nível de satisfação do participante na experiência de uso do e-Termos	198
6.13	Distribuições de Frequências das Questões 8d e 9a	199
6.14	Distribuições de Frequências das Questões 9b e 9c	199
6.15	Distribuição de frequências das respostas sobre a integração da equipe de trabalho no sistema e-Termos	200
6.16	Distribuições de Frequências das Questões 11 e 14	201
6.17	Distribuição de frequências da probabilidade de recomendação de uso do e-Termos para colegas de profissão.	201
6.18	Distribuições de Frequências das Questões 5 e 6	202
6.19	Distribuição de frequências da quantidade de aprendizado adquirido durante o uso do e-Termos	203
6.20	Percentual de respostas da notas da avaliação geral do e-Termos	203
6.21	Percentual de respostas da probabilidade de usar o e-Termos em programas de graduação e pós-graduação.	204
6.22	Distribuições de Frequências das Questões 19a e 19b	205
6.23	Distribuições de Frequências das Questões 19c e 19d	206

Lista de Tabelas

2.1	Grupos de Pesquisas e Estudos Terminológicos no Brasil no CNPq.	22
3.1	Linguagens para representação de ontologias.	72
3.2	Ferramentas de edição de ontologias.	76
3.3	Formatos de intercâmbios de dados terminológicos e lexicográficos.	89
4.1	Conjunto de ferramentas de PLN disponíveis no e-Termos	101
4.2	Conjunto de futuras ferramentas de PLN	101
4.3	Relações Conceituais Genéricas do e-Termos Lenci et al. (2000)	122
4.4	Principais sistemas de gerenciamento de base de dados terminológicos	134
4.5	Tipos de Dados disponíveis no e-Termos	140
4.6	Conjunto de ferramentas colaborativas disponíveis no e-Termos	149
6.1	Seleções da avaliação colaborativa do Ambiente e-Termos	182
6.2	Focos da Avaliação Colaborativa do e-Termos	182
6.3	Valores da relação entre perfil do usuário e gestão do tempo	209
6.4	Valores da relação entre perfil do usuário e gestão da informação	209
6.5	Valores da relação entre perfil do usuário e gestão da equipe de trabalho	210
6.6	Valores da relação entre perfil do usuário e gestão do trabalho terminológico	210
6.7	Valores da relação entre perfil do usuário e nível interesse	211
6.8	Valores da relação entre perfil do usuário e probabilidade de uso	212
6.9	Valores da relação entre perfil do usuário e o nível de satisfação	212
6.10	Valores da relação entre perfil do usuário e o nível de concordância da metodologia de trabalho	213
6.11	Valores da relação entre perfil do usuário e o tempo de experiência	214
6.12	Valores da relação entre perfil do usuário e as notas da avaliação geral	214
6.13	Correlação de Spearman entre a Aparência da Interface e Posicionamentos dos componentes na tela	215
6.14	Correlação de Spearman entre a Aparência da Interface e a Disposição dos menus	216
6.15	Correlação de Spearman entre o controle e a gestão do trabalho terminológico e a confiabilidade final do produto	216
6.16	Correlação de Spearman entre a interatividade da equipe e a tomada de decisões	217
6.17	Correlação de Spearman entre a comunicação dos membros da equipe e a promoção da integração	217
6.18	Correlação de Spearman entre a comunicação dos membros da equipe e o controle e gestão do trabalho terminológico	218
6.19	Correlação de Spearman entre o nível de interesse do usuário e a uniformização das atividades e tarefas da equipe de trabalho	218
6.20	Correlação de Spearman entre o nível de satisfação do usuário e a aparência da interface	218
6.21	Correlação de Spearman entre o nível de satisfação do usuário e a interatividade com o usuário	219
6.22	Correlação de Spearman entre o nível de satisfação do usuário e a utilização de recursos informatizados	219

Introdução

1.1 Contextualização

Não é de hoje que os homens criam e utilizam palavras para expressar e denominar conceitos, objetos e processos dos diversos campos do conhecimento. O atual desenvolvimento das ciências tem provocado uma situação nunca antes experimentada pela sociedade: a presença constante da tecnologia em nosso cotidiano, não se restringindo apenas a grandes laboratórios, instituições e empresas de pesquisa. Com a presença diária desta tecnologia, a língua deve ser capaz de nomear novos conceitos e novos objetos, a ponto de ser eficaz em todos os âmbitos da comunicação do homem.

Devido aos avanços tecnológicos e à especialização das ciências, estes novos conceitos e objetos ocorrem em domínios específicos do conhecimento humano, sendo necessário, para referenciá-los, o uso das línguas de especialidades. Diferentemente da língua geral, aquela em que usamos diariamente, a língua de especialidade é utilizada nas comunicações técnicas e especializadas e exige uma comunicação precisa baseada em um vocabulário específico de um determinado domínio de especialidade. Este vocabulário específico é formado pelos *Termos*, um grupo de palavras com significados especializados que são usados para expressar e referenciar o conjunto de objetos, conceitos, métodos e processos pertencentes a uma área específica do saber humano.

A Terminologia é a ciência que estuda a natureza dos *Termos* e seu emprego nas obras de referência. Entretanto, a Terminologia é naturalmente polissêmica¹ e de acordo com Cano (2001) tem pelo menos três significados. Em sua primeira acepção, como **disciplina**, Cabré (1993) define a Terminologia como ciência autônoma que estuda a comunicação e expressão do conhecimento especializado de uma área específica por meio da análise comportamental dos *Termos*. A segunda acepção, como **prática**, a Terminologia designa um campo teórico de estudo

¹Entidade ou objeto que possui vários significados.

aplicado aos Termos, utilizados nas línguas de especialidade Krieger e Finatto (2004), Hernández (2002), Dias (2000). E finalmente, o terceiro significado trata a Terminologia como **produto**, que é a obra de referência ou o conjunto de *Termos* de um determinado domínio do conhecimento.

Cada vez mais a Terminologia torna-se relevante para a sociedade atual. Isto porque, em seus diferentes aspectos, vincula-se muito diretamente aos paradigmas de desenvolvimento econômico, tecnológico e cultural de um país, no mundo contemporâneo. Esses paradigmas, por sua vez, estão intimamente relacionados à valorização da língua, ao processo de globalização, ao acelerado desenvolvimento da ciência e da tecnologia e às novas tecnologias informatizadas Krieger (1998).

De acordo com Cabré (2005), poucos campos de estudo têm avançado tanto como a Terminologia nos últimos 20 anos. Agrega-se a esse fato, o reconhecimento da necessidade de organização sistemática das obras de referência, como dicionários e glossários de termos, de modo a facilitar a comunicação interna entre especialistas além favorecer da comunicação em âmbito internacional, principalmente com o advento dos mercados comuns de comércio internacional, como por exemplo, o Mercosul.

Este avanço está vinculado a uma série de fatores que representaram influências positivas, oriundas da chamada *Escola de Viena*, formada por países cuja maturidade do desenvolvimento terminológico permitiu grandes avanços dos conhecimentos teóricos e práticos, além das experiências pioneiras no campo da Terminologia. Nesse panorama, o que se fazia no universo da *Escola de Viena*, orientadas pelas proposições teóricas e metodológicas apresentadas pelo engenheiro Eugen Wüster, constituiu a conhecida Teoria Geral da Terminologia (TGT) e forneceu as primeiras coordenadas de saber e de práticas em Terminologia, sendo acolhidas tanto na Europa, quanto na América.

A Teoria Geral da Terminologia é o pilar referencial dos estudos terminológicos. Embora seu objetivo tenha sido o de padronizar o uso dos termos nas comunicações especializadas, a TGT justifica seu papel de referência porque favoreceu a Terminologia a estabelecer-se como área de conhecimento e de pesquisa Krieger e Finatto (2004). Sua influência foi, sobretudo nos estudos dos mecanismos de funcionamento dos termos, em especial na menção de seu valor monovalente em um determinado discurso específico. Segundo Wüster *apud* Cabré (2005) e Krieger (2000), um termo monovalente é um termo que em relação ao contexto da área objeto possui apenas um significado atual. Em outras palavras, a idéia de monovalência prevê que um determinado termo possui somente um significado em uma determinada área do conhecimento, em um discurso específico. Dessa forma, a TGT, também conhecida como terminologia de orientação prescritiva ou normalizadora, apresenta concepções muito produtivas em relação ao valor terminológico de um determinado item lexical.

Outro fator responsável pelo atual desenvolvimento da Terminologia está relacionado à in-

clusão dos lingüistas no campo de estudos terminológicos, principalmente no estudo do léxico especializado para esclarecimento dos seus vínculos e avaliação dos seus resultados e aplicações na Terminologia. Em decorrência disso, o léxico especializado passou a ser compreendido como componente natural das línguas naturais com todos os efeitos que o funcionamento da língua acarreta. Esta nova visão recusou muitos postulados da TGT, cujos fundamentos foram revistos criticamente, dando origem a uma valorização dos aspectos lingüísticos e comunicacionais dos *Termos*. Estes novos posicionamentos estimularam o aparecimento e a consolidação da Teoria Comunicativa da Terminologia (TCT) formulada por Maria Teresa Cabré e o grupo IULATERM da Universidade Pompeu Fabra, de Barcelona, Espanha. Ao contrário da TGT, a TCT contempla a variação lingüística dos termos em toda a sua dimensionalidade e os considera a partir de um ponto de vista descritivo Krieger (2000), Hernández (2002). Dessa maneira, a principal característica da TCT é a capacidade de entender que os *Termos* comportam diversidades conceituais e variações de dimensionalidade, admitindo inclusive a polissemia. Assim, enquanto a TGT considera os *Termos* como representações conceituais que possuem somente um significado, a TCT considera os *Termos* no contexto da linguagem, admitindo variações conceituais e podendo assumir diferentes significados.

No Brasil, segundo Krieger e Bevilacqua (2005), o início das pesquisas terminológicas ocorreu na década de 90. Nessa época, a pesquisa e os trabalhos terminológicos no Brasil situavam em Brasília e São Paulo (capital). Em Brasília, os representantes destes trabalhos eram: 1) a Universidade de Brasília (UnB), pelas reflexões do Centro de Estudos Lexicais Terminológicos (Lexterm) sobre o léxico especializado e 2) o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), instituição ligada ao Ministério de Ciência e Tecnologia (MCT), que ofereceu subsídios para que as práticas e o conhecimento sobre as questões de documentação especializada e tratamento dos termos avançassem.

Em São Paulo, por sua vez, destacaram-se os trabalhos desenvolvidos na Universidade de São Paulo (USP) que visavam aos aspectos basilares da Terminologia, incluindo a relação existente entre a terminologia e a neologia e, conseqüentemente, a formação dos termos. Além dos trabalhos realizados em São Paulo e Brasília, paralelamente, a Terminologia também ganhava espaço no Rio de Janeiro, onde se destacavam os estudos de tradução na Pontifícia Universidade da Católica (PUC-Rio) e as primeiras atividades de padronização terminológica realizados pela Associação Brasileira de Normas Técnicas (ABNT).

Nesse cenário, embora a Terminologia no Brasil tenha iniciado no começo dos anos 90, paralelamente à mudança de paradigma em direção a TCT, os primeiros trabalhos brasileiros seguiram a visão normativa, predominantemente da TGT. No entanto, os estudos nas universidades e centros de pesquisas, principalmente nos programas de pós-graduação, sob a luz da lingüística e dos

pressupostos descritivos da TCT, contribuíram para superar os enfoques normativos da Terminologia brasileira. No contexto dessa mudança, é importante destacar também os trabalhos da Associação Nacional de Pesquisa e Pós-Graduação em Letras e Linguística (ANPOLL), fundada em 1986, por meio dos trabalhos do Grupo de Lexicologia e Lexicografia.

As universidades brasileiras junto com a ANPOLL tiveram um papel crucial na determinação da terminologia brasileira. Recentemente, outras universidades entraram nesse cenário de pesquisa e desenvolvimento, como a Universidade Federal do Ceará (UFCE), a Universidade Estadual de São Paulo (UNESP), a Universidade Federal de São Carlos (UFSCar) e a Universidade Federal de Uberlândia (UFU), de forma que nos últimos anos o mapa terminológico no Brasil cresceu de forma significativa.

Na esteira deste desenvolvimento, de 2000 em diante, a Terminologia têm se dedicado aos aspectos lingüísticos textuais sob diferentes abordagens, destacando-se neste meio: 1) a formação e identificação das especificidades dos termos, cuja principal referência são as pesquisas de Alves (1999) e 2) o seu tratamento, baseado no princípio de que a identificação dos termos, para fins de criação de produtos terminológicos, deve estar vinculada ao exame dos universos de discurso em que os mesmo aparecem Krieger e Finatto (2004).

Todavia, apesar dessa produtividade, a Terminologia ainda carece de soluções informatizadas que possam agilizar o trabalho, principalmente em sua face terminográfica. A Terminografia corresponde à face prática da Terminologia, direcionada para criação de produtos terminológicos, como dicionários, glossários e banco de dados terminológicos Meyer et al. (1997). Genericamente, a proposta da Terminografia é identificar e analisar as unidades lexicais usadas em domínios específicos do conhecimento, para os fins de criação e elaboração de obras de referência. Por outro lado, essa carência de recursos informatizados é muito salutar, pois revela um campo de investigação pleno de questionamentos e de busca de soluções metodológicas para a tarefa de criação de produtos terminológicos de referência temática que movimentam a ciência da computação.

A Terminologia e a Informática não são independentes uma da outra. Desde 1960, em países desenvolvidos e com grande tradição em pesquisa terminológica, a Terminologia e a Informática estão ligadas para principalmente facilitar o armazenamento e a difusão de dados terminológicos, principalmente na elaboração de grandes bases de dados especializadas, denominados bancos de terminologia. A integração entre as duas áreas é tal que se cunhou o termo **Terminótica**, advindo da generalização de **Terminologia+Informática**, dando origem a um novo conceito e inaugurando um novo paradigma metodológico nas pesquisas terminológicas. Assim, a Terminótica oferece a possibilidade de (semi-)automatizar as atividades do terminólogo, fazendo com que a tarefa exaustiva e prolongada da confecção de produtos terminológicos, que se esten-

dia por longos períodos, possa ser finalizada mais rapidamente.

O primeiro trabalho desenvolvido com o benefício da Terminótica foi o EURODICAUTOM (2004), que surgiu da conjugação de fatores políticos-sociais dos países da Europa, partir da necessidade de um fluxo de comunicação uniforme, eficiente e rápido entre estes países. Atualmente, os principais representantes da junção da Terminologia e a Informática são o IATE (Inter-Active Terminology for Europe)² uma base de dados terminológica interinstitucional de países da Europa, o Terminus³, sistema que atua na sistematização do trabalho terminológico e o *Corpógrafo*⁴ Sarmiento et al. (2004), um ambiente Web integrado que dá suporte para várias tarefas terminográficas.

No Brasil, um dos primeiros projetos a utilizar a informática foi o Grupo Termisul⁵, nos anos 90, como pode ser visto em Maciel (1993) e Maciel et al. (1994). Entretanto, atualmente no Brasil há uma carência grande de recursos computacionais compatíveis com o trabalho terminográfico, que são de fundamental importância para a tarefa de sistematizar (ou harmonizar) terminologias Krieger e Bevilacqua (2005). Os terminólogos brasileiros que coordenam grupos de pesquisa e desenvolvem produtos terminológicos trabalham na maioria das vezes com ferramentas adaptadas e não elaboradas especificamente para a finalidade das pesquisas terminológicas, evidenciando a grande necessidade de exploração da Terminótica em direção a sistematização deste trabalho.

A contribuição advinda da Informática começa de fato a aparecer nas pesquisas terminológicas no Brasil nos últimos dez anos. Isso porque o crescimento de estudos e pesquisas na área de Processamento de Língua Natural (PLN) e o conseqüente aprimoramento e desenvolvimento de ferramentas computacionais voltadas para o português passaram a interferir diretamente na prática terminográfica.

Adicionalmente, a Terminologia espera da Informática o desenvolvimento de sistemas integrados, que possam utilizar conhecimentos empíricos e estruturados nas atividades da terminografia Cabré (2005). Nesse sentido, as aplicações da Terminótica podem ser entendidas como adaptação das práticas terminológicas em benefício de uma necessidade, tanto sob a forma de uma metodologia de gestão automatizada, quanto na criação de um produto terminológico.

O trabalho de gestão e criação dos produtos terminológicos é longo, minucioso e na maioria exige muita pesquisa e estudo da área alvo⁶. Além disso, a equipe de trabalho responsável pela

²O IATE pode ser acessado pelo endereço: iate.europa.eu/

³O Terminus é um sistema desenvolvido pelo grupo IULATERM e representa uma positiva iniciativa em direção à automatização das tarefas de terminologia. Pode ser acessado pelo endereço: <http://igraine.upf.edu/Terminus2/index.html>

⁴O Corpógrafo pode ser acessado pelo endereço: www.linguateca.pt

⁵O Termisul é um grupo de pesquisa responsável pelo Projeto Terminológico Cone Sul, originado no Instituto de Letras, da Universidade Federal do Rio Grande do Sul, em 1991. Mais informações no endereço: <http://www6.ufrgs.br/termisul/index.html>

⁶Área-alvo consiste na área específica do conhecimento sobre a qual o produto terminológico será desenvolvido.

criação de tais produtos deve ser multidisciplinar e formada por variados perfis profissionais, como terminólogos ou lingüistas, e principalmente, o especialista do domínio, indispensável para produzir uma obra de qualidade que seja realmente útil para o consumidor que utiliza estas obras Pavel e Nolet (2002).

Em sintonia com os pressupostos teórico da TCT, uma seqüência de etapas deve ser seguida para a construção de produtos terminológicos: 1) compilação de *cópus*⁷ de especialidade; 2) coleta (ou extração) de candidatos a termos; 3) elaboração do mapa conceitual (também denominado de estrutura conceitual do domínio), doravante *ontologia*⁸; 4) categorização dos candidatos a termos na ontologia e sua validação pelos especialistas; 5) elaboração e incremento da Base Definicional e criação e preenchimento das Fichas Terminológica e; 6) edição dos verbetes e o intercâmbio⁹ e/ou difusão dos produtos terminológicos.

Dada essa seqüência de etapas, outro problema que merece destaque na relação entre Terminologia e Informática, e conseqüentemente na prática terminológica, é o uso de várias ferramentas/sistemas distintas para executar as etapas acima mencionadas. O uso de diferentes sistemas para as tarefas terminológicas torna o trabalho dos terminólogos muito moroso, além de deixar o gerenciamento dos dados mais complicado, pois não existe um padrão de entrada e saída definido entre eles.

Outra deficiência provocada pelo uso de diferentes sistemas é a impossibilidade de se estabelecer e atender as necessidades colaborativas e comunicacionais da equipe multidisciplinar que participa do trabalho terminológico. Isso porque tais necessidades são inerentes a este trabalho e naturais para a interação da equipe, pois a comunicação e o relacionamento colaborativo entre seus membros além de essencial são inevitáveis, e se caracterizam pela troca e o compartilhamento de idéias, opiniões e informações acerca das atividades e produtos intermediários gerados ao longo do trabalho.

Dessa forma, se a base de dados terminológica fosse acessível pela Internet e todas as atividades inerentes ao processo de desenvolvimento das terminologias estivessem num único sistema, uma equipe de trabalho maior, diversificada e situada em diferentes localidades poderia trabalhar unida, cada membro alocado numa atividade e sem prejuízo para o projeto, pois todos compartilhariam uma única visão do trabalho desenvolvido. Além disso, uma base de dados centralizada e acessível permite o gerenciamento mais eficiente e padroniza o fluxo de dados.

Segundo Eseryel et al. (2002), o *Computer-Supported Collaborative Work (CSCW)* é um campo de pesquisa multidisciplinar focado sobre ferramentas e técnicas computacionais utilizadas para prover e garantir o trabalho em grupo de várias pessoas que possuem tarefas rela-

⁷Nesta monografia, usamos o aportuguesamento da palavra *corpus/corpora* como *cópus/cópus*.

⁸O termo *ontologia* será utilizado neste trabalho com o sentido empregado em Ciência da Informação.

⁹Vale ressaltar que o intercâmbio dos produtos terminológicos é dependente do objetivo da equipe de trabalho, bem como de sua própria natureza, pois pode não propósito da equipe a realização de permuta de produtos.

cionadas. Ainda, de acordo com Fluckiger *apud* Santos (1998) e Vieira e Fortes (2004), o CSCW pode ser entendido como o projeto e desenvolvimento de sistemas computacionais para apoiar e aprimorar o trabalho em grupo de usuários envolvidos em tarefas ou objetivos comuns, bem como para a compreensão da utilização de tais sistemas. De acordo com Ehrlich (1999), as aplicações que implementam o trabalho colaborativo, também conhecidas como *Groupware*, são diferentes daquelas aplicações que não apóiam a colaboração, também chamadas de *single-user*. As aplicações *single-user* são desenvolvidas para apoiar as tarefas dos usuários, aquelas executadas no dia-a-dia, freqüentemente bem definidas, observáveis e concretas. Já as aplicações de *Groupware* são desenvolvidas para dar apoio ao trabalho dos usuários, em todo seu contexto, incluindo os aspectos comunicacionais e de relacionamentos com outros membros da equipe, freqüentemente tácito, invisível e amorfo.

Fica caracterizado então, como salienta Bannon (1992), que as aplicações de apoio colaborativo representam um grande esforço para se entender a natureza e as características deste trabalho “invisível”, resultado das atividades colaborativas, com o objetivo de desenvolver adequados sistemas computacionais que buscam compreender como as pessoas trabalham em grupo e o que pode ser desenvolvido para apoiar tais atividades. Isso porque o trabalho em grupo, por meio das aplicações CSCW, permite a interação entre diferentes perfis individuais, podendo eliminar as distâncias, aproximar as pessoas, reduzir o tempo gasto no trabalho, aumentar a produtividade e diminuir os custos. Dessa maneira, um sistema que efetivamente apóie o trabalho colaborativo deve considerar em seu projeto tanto o trabalho em si, objeto da tarefa colaborativa, quanto o aspecto da colaboração.

Devido a estas características, os desafios de avaliar os sistemas desta categoria têm sido tema de várias pesquisas em CSCW Wainer e Barsottini (2004) Pinelle e Gutwin (2000b). Tais pesquisas trabalham na tentativa de desenvolver uma abordagem de avaliação ampla e eficaz que considera os contextos do trabalho colaborativo e a redução de custos de sua aplicação. Entretanto, não há nenhuma definição sobre as abordagens e métodos de avaliação de sistemas colaborativos, principalmente quando são consideradas todas as variáveis contextuais envolvidas neste processo, frutos da alteração da dinâmica de trabalho em grupo que a implantação de um sistema colaborativo causa no ambiente em que é utilizado Prates e Raposo (2006), Neale et al. (2004).

Neste contexto, existem diversos motivos que justificam a carência de abordagens e métodos de avaliação, tais como: 1) os métodos de avaliação *single-user* não são aplicáveis aos sistemas *Groupware*, 2) alto custo de criação e aplicação, 3) demora na apuração dos resultados, já que os benefícios dos sistemas *Groupware* aparecem depois de períodos longos de adoção, e 4) a dificuldade dos criadores de definir o quê, como e quando avaliar. Assim, o processo de

avaliação dos sistemas colaborativos é difícil porque depende de vários fatores externos, ligados principalmente ao contexto real de uso destes sistemas, havendo a necessidade de consideração destes fatores no momento de avaliar.

Apesar desta dificuldade, vários métodos de avaliação têm sido propostos nos últimos anos. Para sistemas já desenvolvidos, um dos métodos formais de avaliação mais eficientes é o *Cenário de Uso*, que pode ser entendido como um conjunto organizado de tarefas distribuídas sistematicamente em um contexto que descreve a interação dos usuários em um sistema colaborativo. A principal vantagem de aplicação dos *Cenário de Uso* está na capacidade de simular contextos reais de uso, bem como a colaboração prevista nos sistemas *Groupware*, além da possibilidade de inserir os objetivos da avaliação de maneira que sejam coletados e aferidos durante sua aplicação.

1.2 Objetivos

Abaixo, definimos a hipótese ou questão central desta pesquisa:

Seria possível construir um ambiente colaborativo que apóia-se todas as etapas do trabalho terminológico de orientação descritiva e avaliar o impacto que tal sistema causa no trabalho terminológico?

Assim, com tal suposição e com base nessa contextualização, na qual destacamos a simbiose entre a Terminótica, a seqüência de etapas do trabalho terminológico e a carência de sistemas integrados e colaborativos para sua gestão, que propomos nesta pesquisa de doutoramento, o projeto, a implementação e a avaliação de um ambiente computacional que integra num único espaço ferramentas informatizadas para auxiliar a pesquisa terminológica e gerir o trabalho terminográfico. Chamado de **e-Termos** (www.etermos.ufscar.br), um acrônimo de “*Termos Eletrônicos*”, este ambiente computacional reúne um conjunto de ferramentas lingüísticas e colaborativas interligadas a uma estrutura modular dedicada às tarefas e pesquisas terminológicas.

Apoiado nos pressupostos teóricos da Teoria Comunicativa da Terminologia e sendo uma aplicação CSCW, o **e-Termos** é um Ambiente Colaborativo Web que contempla as atividades da gestão terminológica. Definido como um projeto modular, tal Ambiente, composto por seis Módulos de trabalho bem definidos, tem o objetivo de automatizar as tarefas de criação e gerenciamento de produtos terminológicos. Cada Módulo abriga tarefas específicas e inerentes ao processo de criação destes produtos, sendo atreladas a eles diferentes ferramentas de análise lingüística, que possuem a função de dar suporte às tarefas de PLN envolvidas neste processo e também ferramentas colaborativas, designadas para dar apoio às necessidades comunicacionais da equipe de trabalho. O Ambiente foi desenvolvido para atender a grande maioria das necessidades e tarefas dos usuários de maneira que o fluxo de dados entre os Módulos seja garantido

e aconteça de forma transparente, perfazendo desde a seleção e análise do corpus até a difusão do produto terminológico final.

Dado que se trata de um trabalho de equipe multidisciplinar, a colaboração no **e-Termos** é um processo em que dois ou mais usuários trabalham juntos para atingir metas comuns. A colaboração implica em uma atividade coordenada, resultado da tentativa constante de construir e manter uma concepção compartilhada do problema entre todos os usuários envolvidos Blandford (1994). No **e-Termos** estão definidos quatro perfis de usuários que possuem visões específicas do Ambiente, variando de acordo com a atividade que desenvolvem.

No contexto de uso e implantação do Ambiente, esta pesquisa também apresenta a aplicação do método de *Cenário de Uso* no processo de avaliação colaborativa do **e-Termos**, revelando os detalhes da criação e aplicação do *Cenário* utilizado, bem como da elaboração do seu enredo, seus objetivos, sua organização e principais características. Além disso, são mostrados todos os passos da preparação de sua execução, da configuração do ambiente, da formação das equipes participantes e a condução da avaliação. Adicionalmente, como parte do aparato e do processo de avaliação, o método de coleta de dados utilizando *Questionários de Pesquisa* é apresentado, juntamente com sua organização, aplicação e características.

Especificamente, os objetivos deste trabalho de doutoramento podem ser agrupados da seguinte maneira:

- sistematizar o conhecimento necessário para a aplicação ou implementação de ferramentas computacionais no contexto da pesquisa terminológica;
- reunir em uma única arquitetura computacional todas as etapas do trabalho da gestão terminológica em consonância com a TCT;
- adaptar ou implementar ferramentas de PLN que sejam relevantes e condizentes com as tarefas desenvolvidas em cada Módulo do **e-Termos**;
- disponibilizar, adaptar e/ou implementar ferramentas colaborativas que suportem o trabalho terminológico e as necessidades de comunicação entre a equipe de trabalho;
- modelar e criar um banco de dados terminológico que: 1) suporte todas as tarefas e atividades das ferramentas computacionais inerentes aos Módulos, 2) permita o intercâmbio de dados entre padrões e/ou formatos amplamente aceitos pela comunidade científica da terminologia; 3) armazene com segurança todo conhecimento contido em um produto terminológico e 4) suporte o aspecto da colaboração inerente ao método de trabalho e que é refletido nas ferramentas colaborativas;

- projetar o Ambiente Colaborativo Web para gerenciar os aspectos da criação de produtos terminológicos;
- disponibilizar ou implementar ferramentas que permitem a exportação de dados num formato de intercâmbio de dados terminológicos para possibilitar a difusão do conhecimento disponível;
- analisar e avaliar a percepção do efeito e dos impactos de um sistema colaborativo sobre a rotina do trabalho terminológico.

1.3 Contribuições

Na área da Terminologia, a principal contribuição deste trabalho é a automatização de um método prático que procura explicitar os postulados da terminologia de orientação descritiva, fundamentada em princípios da Lingüística, e que tem como novidade a ação colaborativa, inédita neste âmbito de aplicações da computação. Nesse contexto, esta pesquisa é de grande importância para cenário da Terminologia, pois trouxe para a prática todas as etapas do processo de criação de produtos terminológicos com o diferencial colaborativo.

Decorrente desta contribuição está o fato que o **e-Termos** insere diretamente a possibilidade de instanciação de aspectos pedagógicos e didáticos no ensino da Terminologia, da Lexicologia e da Tradução, visto que é possível utilizá-lo para tais práticas, principalmente nos âmbitos de ensino da graduação e pós-graduação.

Ainda, no conjunto das contribuições terminológicas deste doutorado, está a necessidade de integração e cooperação técnica e científica, principalmente quando nos baseamos na Terminologia de orientação descritiva para favorecer a difusão e o intercâmbio do conhecimento especializado, já que nesse contexto, a língua passa a funcionar como veículo de disseminação de conhecimento, facilitando a comunicação e ao mesmo tempo contribuindo para sua consolidação.

Uma das contribuições do **e-Termos** é o fato de estar sempre acessível, na Internet, permitindo a interação da equipe de trabalho em um processo incremental de construção do produto terminológico, permitindo que este esteja sempre sendo atualizado. Outra importante contribuição é o acoplamento de ferramentas de PLN com intuito de oferecer suporte lingüístico às tarefas envolvidas no trabalho terminológico. O uso dessas ferramentas é um claro diferencial resultante desse projeto, já que tais ferramentas são muito úteis para determinadas atividades, como por exemplo, a análise de corpus e a redação das definições nas Fichas Terminológicas, tão importantes para o contexto deste trabalho.

No entanto, o principal diferencial e a notável contribuição do **e-Termos** está na característica colaborativa que este ambiente computacional implementa. Perfazendo desde a criação

automática de corpúsculos especializados (correspondente ao Módulo 1) até a difusão e intercâmbio do conjunto de verbetes (correspondente ao Módulo 6), baseando-se nos processos de apoio e cooperação de um conjunto diferenciado de profissionais, o **e-Termos** possibilita o trabalho, a produção em conjunto e a troca de informações para melhorar as atividades de grupos de usuários com interesses e propósitos comuns. Além disso, o aspecto colaborativo permite a sistematização e o mapeamento do fluxo de atividades dos diversos profissionais envolvidos na criação de produtos terminológicos, produzindo resultados mais rápidos e fiáveis.

Além disso, para área de sistemas colaborativos esta pesquisa forneceu uma singular contribuição, pois adaptou e empregou um método formal de avaliação colaborativa, resultante da aplicação de *Cenários de Uso* em conjunção com *Questionários de Pesquisa* para coleta de dados, para avaliar as percepções de impactos resultantes da implantação e uso do **e-Termos** na esfera do trabalho terminológico e lexicográfico.

Em outras palavras, o **e-Termos** permite que os diferentes integrantes de uma mesma equipe de pesquisa possam trabalhar juntos, de maneira colaborativa, interagindo, editando, atualizando, inserindo e retirando informações de todos os Módulos, bastando conectar-se à Internet, buscar a URL e utilizar uma senha de acesso. O mesmo procedimento pode ser utilizado para a interação com os especialistas da área alvo, ou seja, especialistas previamente selecionados podem opinar, criticar, sugerir alterações, ratificar os dados (lista de termos, definições, etc.) por meio de acesso à Internet, num processo incremental de gestão do trabalho e da equipe para a construção do produto terminológico.

Portanto, entendemos que fazer Terminologia na era da Informática significa criar um conjunto de procedimentos automatizados ou semi-automatizados que dêem suporte às tarefas envolvidas na gestão terminológica, bem como suporte todas as necessidades de comunicação inerentes à natureza deste trabalho.

1.4 Organização do trabalho

Esta monografia está organizada da seguinte forma. O Capítulo 2 faz uma revisão bibliográfica da Terminologia para apontar as definições básicas, histórico, áreas de interface e objetos de estudo. No Capítulo 3 é realizada uma explanação detalhada de todas as etapas do processo de gestão e criação de produtos terminológicos, relatando as atividades, características e os problemas envolvidos em cada uma delas. Também neste capítulo é apresentada uma revisão dos principais programas ou ferramentas que implementam as tarefas de cada etapa do trabalho terminológico. O Capítulo 4 apresenta o **e-Termos**, destacando suas principais funcionalidades e características, o suporte teórico, a descrição detalhada de cada Módulo de trabalho, sua natureza colaborativa e os principais aspectos que permearam sua implementação. O Capítulo 5

apresenta uma revisão bibliográfica das abordagens e os aspectos gerais sobre a avaliação de sistemas colaborativos com destaque especial para os *Cenários de Uso*. O Capítulo 6 revela e explica a avaliação colaborativa empregada no **e-Termos**, enfatizando os detalhes, o planejamento, a configuração, a elaboração e a execução. Além disso, este capítulo também apresenta e discute os resultados obtidos no processo de avaliação. Finalmente, o Capítulo 7 apresenta a conclusão e as contribuições desta pesquisa e propõe os trabalhos futuros previstos.

Os Apêndices A, B e C desta monografia correspondem aos 3 *Cadernos de Atividades* elaborados para o processo de avaliação do **e-Termos**, sendo um para cada perfil de usuário (Gerente de Projeto, Terminólogo/Linguista e Especialista do Domínio), respectivamente. O Apêndice D apresenta o *Questionário de Pesquisa* aplicado durante o processo de avaliação e o Apêndice E as distribuições de frequência das questões dos questionário.

Princípios teóricos da Terminologia

O propósito deste capítulo é estabelecer os princípios teóricos da Terminologia, familiarizando o leitor com a Terminologia e suas bases teóricas. Começaremos por caracterizar a identidade da área, que tanto pode significar o conjunto de termos técnicos científicos de uma área específica quanto o campo de estudos que se ocupa do chamado léxico especializado.

Apesar do aspecto multidisciplinar da Terminologia, é indiscutível que este campo de conhecimento possui identidade própria e está formado por objetos, estudos e práticas bem delimitados. Para identificar as faces da Terminologia, um breve histórico será delineado com destaque para suas escolas teóricas, indo desde a sua origem até o estágio de desenvolvimento atual da Terminologia, em relação aos termos técnico-científicos e a sua consolidação como disciplina, enfatizando-se seus fundamentos lingüísticos. Como parte integrante destes princípios teóricos, a interdisciplinaridade da Terminologia também é apresentada, com ênfase em suas inter-relações com a Lexicologia, Lexicografia, Terminografia e a Tradução, assinalando em cada uma os pontos de aproximação e distinção.

Por fim, especificamos os objetos da Terminologia, tratando principalmente das unidades lexicais com significado especializado, os *termos*, e os relacionamentos com a definição terminológica, um dos objetos constituintes da Terminologia, o uso de textos especializados e os aspectos teóricos relacionados aos glossários e dicionários especializados.

2.1 Definições Básicas

O nascimento da feição terminológica de uma determinada área do conhecimento ocorre quando os especialistas desta área criam e utilizam um conjunto específico de palavras para expressar seus conceitos, processos, métodos e objetos específicos. Genericamente, este conjunto limitado de palavras pode ser denominado de *termos*. Decorre que o uso dos termos em áreas específicas do conhecimento é antigo, já que os mesmos ocorrem no universo das ciências, das técnicas e do trabalho humano Krieger e Finatto (2004). Por outro lado, o estudo dos termos, em todas as suas dimensões (lingüística, conceitual e comunicativa, conforme será visto na Seção 2.4.1) é recente e data da segunda metade do século passado.

Diante destes fatos, é notório que a Terminologia é um termo naturalmente polissêmico. Em sua primeira acepção, a Terminologia se ocupa de um conjunto de palavras técnicas pertencentes a um domínio específico do conhecimento, como, por exemplo, a medicina ou a ciência da computação. Em uma segunda acepção, tomada em um sentido mais restrito, a Terminologia designa uma disciplina ou campo teórico de estudo aplicado aos termos técnico-científicos utilizados nas línguas de especialidade Krieger e Finatto (2004), Hernández (2002), Dias (2000). Nesse contexto, língua comum é aquela que usamos no cotidiano, ao passo que a língua de especialidade é utilizada nas comunicações técnicas, exigindo uma comunicação precisa baseada em um vocabulário específico de um determinado domínio de especialidade.

Em relação ao conjunto de termos propriamente dito, a Terminologia é entendida como um rol de atividades práticas que envolvem a compilação, descrição, tratamento e apresentação dos termos de uma determinada área de conhecimento. Entretanto, segundo Cabré (1993) não se pode considerar a Terminologia como “*uma atividade prática que se justifique por si só*”. Por sua vez, em relação à disciplina terminológica, Cabré (1993) também define a Terminologia como ciência autônoma que estuda a comunicação e expressão do conhecimento especializado de uma área específica por meio da análise comportamental dos termos técnico-científicos. Resume então que o objeto primordial de estudo da Terminologia é o termo técnico-científico Krieger e Finatto (2004). É este objeto que marca a principal característica da Terminologia, já que os mesmos realizam duas funções essenciais: a representação e a transmissão do conhecimento especializado. Os termos, por sua vez, podem ser considerados como signos lingüísticos de valor polissêmico que veiculam significados específicos de cada área, estabelecendo referências e ligações com o universo exterior, sempre no âmbito do domínio específico no qual estão inseridos. Ainda, dado este “valor polissêmico” dos termos, a precisão conceitual torna-se uma condição necessária para o intercâmbio comunicativo eficiente, seja ele tecnológico, científico ou cultural.

Sobre os termos é que se destacam as aplicações práticas da terminologia. Tais aplicações

compreendem uma variedade de produtos e ferramentas que vão desde a análise de textos especializados e o reconhecimento automático de termos até a criação e edição de glossários, dicionários técnico-científicos e banco de dados terminológicos.

Neste campo, segundo Wüster *apud* Krieger e Finatto (2004), a Terminologia foi concebida como um ramo da Lingüística Aplicada no sentido de que ela relaciona-se ao aspecto prático da Terminologia, particularmente na produção de obras de referência especializadas que promovem a comunicação. Entretanto, em Krieger e Finatto (2004) sinalizam as diferenças básicas entre a Lingüística Aplicada e a Terminologia, em que a primeira se ocupa dos objetos relacionados ao léxico geral enquanto que a segunda somente ao léxico especializado. No entanto, definir se a Terminologia pertence ou não à Lingüística Aplicada depende do enfoque sobre o qual os termos estão sendo empregados e analisados, já que um determinado termo faz parte do léxico geral e especializado, e seu valor terminológico depende do local onde está inserido.

Independente dos enfoques particulares, a Terminologia é uma área de conhecimento que se ocupa principalmente da constituição e comportamento dos termos, compreendendo desde sua formação lexical até o exame de suas relações nas mais diferentes áreas do conhecimento científico e técnico. Adicionalmente, a Terminologia, aproveitando-se da relação com tais áreas, se capacita para estabelecer princípios e métodos para a criação de ferramentas e produtos terminológicos.

A especialização dos estudos terminológicos tem acendido as pesquisas sobre os textos especializados (como será mostrado na Seção 2.4.3), visto que estes são componentes indispensáveis para a compreensão da natureza dos principais objetos terminológicos. Finalmente, a dupla face da Terminologia reúne tanto a descrição e explicação dos termos quanto o conjunto de diretrizes metodológicas para o tratamento destes, estabelecendo sua identidade própria e firmando-a cada vez mais no campo dos estudos lingüísticos, independentemente do seu caráter multidisciplinar Krieger e Finatto (2004).

2.2 Breve histórico da Terminologia

Iniciamos esta seção utilizando as palavras de Cabré (2005) quando afirma que falar do passado da Terminologia não pressupõe necessariamente um rompimento absoluto com o presente. Isso porque abordar o passado não significa somente rever a história de uma única etapa, e sim, relembrar o panorama terminológico que determina e caracteriza a Terminologia atual.

Tendo em vista o exposto na seção anterior em que a Terminologia é uma área teórica e aplicada, delinear seu panorama histórico implica mencionar vários aspectos que cobrem sua trajetória de desenvolvimento. Assim, considerando que a Terminologia está situada no universo dos saberes técnicos e científicos, suas atividades práticas datam de um tempo remoto, visto

que o conhecimento humano especializado não ocorre somente nos tempos atuais. Todavia, em contrapartida à antiguidade da prática terminológica, os estudos teóricos sobre os termos técnico-científicos nas comunicações especializadas são mais recentes e datam da primeira metade do século XX.

Embora o interesse pelos fundamentos teóricos da Terminologia seja recente, é no século XVII que surgem os primeiros vocabulários específicos, mais precisamente relacionados aos conceitos e denominações das ciências e das artes, em decorrência do aparecimento dos dicionários clássicos da cultura européia.

Com o advento da Revolução Industrial no século XVIII e o consequente crescimento da indústria, da ciência, das técnicas, do comércio e da administração, o interesse pelas línguas e vocabulários especializados progride de vários modos e por diferentes razões. Tal interesse vincula-se também ao fato de que neste século ocorre o estabelecimento das nomenclaturas técnico-científicas, derivadas principalmente do latim e grego.

Impulsionada por este desenvolvimento tecnológico, a discussão das propriedades e problemas que envolvem o uso das línguas de especialidade recebe especial atenção, principalmente pelos enciclopedistas. Esta discussão adentra o século XIX, que por sua vez, estimula, sobretudo, a internacionalização das ciências, fazendo com que os cientistas ocupem-se com estratégias para assegurar a univocidade da comunicação científica internacional Krieger e Finatto (2004).

Motivados por este objetivo e preocupados com as regras de formação dos termos técnico-científicos, os cientistas começam a estabelecer padrões terminológicos em seus domínios de especialidade de modo a instituir determinadas especificidades às linguagens. É nesse contexto que surge a demarcação nítida entre o chamado léxico das ciências e o léxico geral.

Este comportamento dos cientistas e a consequente consolidação dos léxicos especializados passam a ser marcantes na primeira metade do século XX, e corresponde à época dourada da Terminologia, segundo Cabré (2005) e Krieger e Finatto (2004). Sendo presidida na Europa pelos estudos de Eugen Wüster, este período torna-se de grande desenvolvimento e consolidação da Terminologia, a qual ganha reconhecimento disciplinar, social e político.

A consolidação da Terminologia ocorrida durante o século XX mostra a relação direta existente entre a Terminologia e o avanço das ciências. Isso porque o desenvolvimento da tecnologia e a ampliação do conhecimento científico requerem novas denominações para as novas descobertas e invenções do homem. Da mesma maneira, agora já situado na segunda metade do século XX, está o advento da globalização, que dado o aumento das transações comerciais internacionais e o surgimento dos atuais blocos econômicos, motivaram o interesse da comunidade global pelo componente lexical especializado, e consequentemente, a sua estruturação Krieger (1998). Outro fenômeno que intensificou a necessidade de uso das terminologias é o aparecimento acentuado

de intercâmbios culturais, científicos e tecnológicos, também frutos do processo de globalização. Esta necessidade fez crescer a preocupação com a tradução, e, conseqüentemente, com o uso correto dos termos técnico-científicos.

De modo semelhante, os avanços da informática também contribuíram para os avanços da Terminologia, principalmente no que diz respeito aos aspectos práticos. O crescimento da informática, tais como o desenvolvimento de software e os sistemas de banco de dados, favoreceram a criação bases de dados terminológicos, como, por exemplo, o EURODICAUTOM (2004), o Termium Pavel e Nolet (2002) e o TERMDIC no contexto de trabalho do TERMISUL Krieger (1998), Maciel (1993), Morales (2001), este último representando o Brasil e o Mercosul. O desenvolvimento de software, por sua vez, permitiu a produção de pequenos programas e aplicações, como os editores de fichas terminológicas e dicionários bilíngües em formato eletrônico. Um conjunto amplo desses programas pode ser visto no Volume 16 da Revista Terminómetro de 1994 Terminómetro (1994).

Percebe-se, portanto, que o interesse pelas terminologias não se restringe aos especialistas de uma área específica, objeto de trabalho da Terminologia, definido como “usuários diretos” em Krieger e Finatto (2004). No mesmo empenho estão os chamados “usuários indiretos” que correspondem aos profissionais ligados diretamente aos estudos da linguagem, como, por exemplo, os tradutores, lexicógrafos, documentalistas, redatores técnicos e terminógrafos. Além do mais, participa do mesmo interesse o próprio cidadão comum, devido à acelerada produção do conhecimento e os avanços tecnológicos que atualmente afeta a sociedade.

É neste cenário, na segunda metade do século passado, que identificamos a real expansão da Terminologia e o uso de léxicos e vocabulários especializados, principalmente como evolução da consciência da sociedade, política, científica e civil, no reconhecimento do valor comunicacional e social dos termos técnicos, bem como no seu papel de ampliação do conhecimento específico.

A transmissão desse conhecimento baseia-se fundamentalmente na linguagem, e no caso do conhecimento especializado, nos termos. Como veículo transmissor do conhecimento especializado, é preciso considerar a terminologia como uma peça fundamental na sociedade atual, ainda que seja necessário desenvolvê-la para poder seguir cumprindo sua missão.

2.2.1 Escolas e Teorias

A necessidade dos cientistas e especialistas de empreender uma comunicação especializada eficiente, principalmente em nível internacional, os levou a se preocuparem com os aspectos pragmáticos da Terminologia. Embora isolados, vários estudos sobre os propósitos práticos da Terminologia iniciaram na Europa e na América do Norte. Esses estudos deram origem ao que chamamos de Escolas da Terminologia, tais como a de Viena, a da Rússia, a de Praga e,

posteriormente, a do Canadá.

As Escolas clássicas, representadas pelas escolas de Viena, de Praga e da Rússia, são responsáveis pelo estabelecimento das bases da disciplina terminológica e pelo delineamento das diretrizes para a sistematização dos métodos de trabalho em terminologia, visando principalmente à padronização dos termos técnico-científicos para uma comunicação profissional eficiente. Nesse contexto, tais escolas procuraram estabelecer orientações metodológicas para o tratamento dos termos, baseados no pressuposto de que os termos são denominações de conceitos.

Sem dúvida, o representante mais importante dessas escolas é o engenheiro Eugen Wüster. Pertencente à *Escola de Viena* e fundador da Teoria Geral da Terminologia (TGT), seu nome está fortemente vinculado ao objetivo de traçar diretrizes para a normalização das terminologias, com o propósito de facilitar seu uso unívoco universal. Essa teoria se tornou um marco muito importante na área de Terminologia e esta baseada nos trabalhos de Wüster, principalmente sobre sua tese de doutorado, apresentada em 1931.

A Teoria Geral da Terminologia é o pilar referencial dos estudos terminológicos. Apesar do seu objetivo ter sido o de padronizar uso dos termos nas comunicações especializadas, a TGT justifica seu papel de referência porque favoreceu a Terminologia a estabelecer-se como área de conhecimento e de pesquisa Krieger e Finatto (2004).

O pensamento wüsteriano e consequentemente o enfoque da TGT estão fundamentados na atividade onomasiológica da terminologia, na qual o conhecimento sobre os conceitos de uma determinada área específica está precedido da seleção de denominações mais convenientes. Ou seja, o trabalho terminológico deve partir de uma estruturação dos conceitos e atribuir à cada um deles uma denominação de referência Cabré (2005).

A TGT atua, sobretudo, nos mecanismos de funcionamento dos termos, em especial na menção de seu valor monovalente em um determinado discurso específico. Segundo Wüster *apud* Cabré (2005) e Krieger (2000), um termo monovalente é um termo que, em relação ao contexto da área objeto, possui apenas um significado atual, embora possa ser polissêmico. Em outras palavras, a idéia de monovalência prevê que um determinado termo possua somente um significado em uma determinada área do conhecimento em um discurso específico. Dessa forma, a TGT, também conhecida como terminologia de orientação prescritiva ou normalizadora, apresenta concepções muito produtivas em relação ao valor terminológico de um determinado item lexical.

Entretanto, esta visão normalizadora reduz a abrangência dos significados dos termos, fazendo com que não sejam vistos como elementos das línguas naturais. Isso produz uma valorização extrapolada da dimensão conceitual dos termos, visto que aparecem como “unidades de conhecimento” que transmitem conceitos, em detrimento da hipótese que os considera como participantes

da língua geral. Portanto, fica claro que na teoria de Wüster, os termos expressam conceitos e não significados, ocorrendo uma sobreposição do componente lingüístico do termo em relação ao conceitual.

É esta característica da TGT que foi, e tem sido, segundo Krieger e Finatto (2004), alvo de críticas desde o início da década de 90, principalmente em relação à insuficiência em representar o funcionamento dos léxicos especializados direcionados aos aspectos lingüísticos e pragmáticos dos termos. Fato que tem estimulado os especialistas em Terminologia a dedicarem uma profunda revisão crítica sobre os fundamentos da terminologia normalizadora Krieger (2000), Hernández (2002).

No cenário de revisão da TGT, evidenciam-se os trabalhos de Maria Teresa Cabré, pesquisadora na Universidade Pompeu Fabra, em Barcelona, Espanha, pioneira na análise crítica dos fundamentos da TGT e responsável pela elaboração de uma nova teoria da Terminologia que privilegia os aspectos comunicativos das línguas de especialidade em detrimento dos propósitos normalizadores.

Esta “nova” teoria, chamada de Teoria Comunicativa da Terminologia (TCT), contempla a variação lingüística dos termos em toda a sua dimensionalidade e os considera a partir de um ponto de vista descritivo Krieger (2000), Hernández (2002). A característica inovadora da TCT está na capacidade de entender que as unidades terminológicas comportam diversidades conceituais e variações de dimensionalidade, admitindo inclusive a polissemia.

Dessa maneira, enquanto a TGT considera os termos como representações conceituais que ocupam uma determinada posição na hierarquia conceitual de uma área do conhecimento, a TCT considera o real funcionamento dos termos, no contexto da linguagem, registrando as variações denominativas e conceituais que os termos podem assumir.

Assim, a TCT se baseia na valorização dos aspectos comunicativos das linguagens especializadas, compreendendo que os termos fazem parte da língua natural e que uma determinada unidade lexical pode assumir caráter terminológico em situações e contextos particulares Cabré (2005). Dessa forma, como lembrado em Krieger e Finatto (2004), o conteúdo de um termo não é fixo, e sim relativo, variando conforme o cenário comunicativo especializado em que se inscreve. Decorre disso, então, a compreensão de que um determinado termo pertence à língua natural e sofre todas as implicações sistêmicas e contextuais como qualquer outra palavra da língua.

Resumidamente, podemos dizer que a principal ruptura da TCT em relação a TGT é que a primeira reconhece a polissemia dos termos no espaço das comunicações especializadas, pois acolhe o dinamismo das ciências e das línguas, reconhecendo que as unidades terminológicas estão em constante evolução. Entretanto, esta visão polissêmica dos termos não induz o desinteresse pela relação termo-conceito, tão evidenciada na TGT. Pelo contrário, o componente conceitual

no âmbito da TCT é importante na medida em que representa uma determinada identificação dos termos, tendo em vista sua íntima ligação com a definição terminológica.

2.2.2 Histórico da Terminologia no Brasil: grupos de estudo

Segundo Krieger e Bevilacqua (2005), o início das pesquisas terminológicas no Brasil ocorreu em 1990, na ocasião do Simpósio da RITerm ocorrido no Brasil neste mesmo ano. Os principais enfoques deste simpósio foram, entre outros, a necessidade de organização sistêmica dos repertórios terminológicos e o empreendimento da comunicação especializada, não só no Brasil, mas também no âmbito ibero-americano.

No início dos anos 90 a pesquisa e os trabalhos terminológicos no Brasil se situavam em Brasília e São Paulo (capital). Em Brasília, os representantes destes trabalhos eram: 1) a Universidade de Brasília (UnB), pelas reflexões do Centro de Estudos Lexicais Terminológicos (*Lexterm*) sobre o léxico especializado e 2) o Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), instituição ligada ao Ministério de Ciência e Tecnologia (MCT), que desempenhou papel muito importante no desenvolvimento da Terminologia no Brasil, tendo oferecido subsídios para que as práticas e o conhecimento sobre as questões de documentação especializada e tratamento dos termos avançassem.

Entre as realizações do IBICT, destaca-se o Projeto Integrado para Implantação e Difusão de Terminologia Científica e Técnica no Brasil, de 1994, que deu origem ao *BrasilTerm*, um banco nacional de dados terminológicos; além da realização de cursos, a consolidação de convênios e concessão de bolsas de estudo para formação e treinamento de capital humano em Terminologia.

Em São Paulo, por sua vez, destacam-se os trabalhos desenvolvidos na Universidade de São Paulo (USP) que visavam aos aspectos basilares da terminologia, incluindo a relação existente entre a terminologia e a neologia e à formação dos termos técnico-científicos. Como resultado destes trabalhos, foi criado em 1992 o Centro Interdepartamental de Tradução e Terminologia (CITRAT) cuja responsabilidade era representar as diversas interações da Terminologia com outras especialidades da área de Letras, evidenciando principalmente o caráter interdisciplinar da Terminologia.

Além dos trabalhos realizados em São Paulo e Brasília, paralelamente, a Terminologia também ganhava espaço no Rio de Janeiro, onde se destacavam os estudos de tradução na Pontifícia Universidade da Católica (PUC-Rio) e as primeiras atividades de padronização terminológica realizados pela Associação Brasileira de Normas Técnicas (ABNT).

Embora o berço da Terminologia no Brasil tenha sido na região sudeste e no Distrito Federal, também no início da década de 90, foi criado o Projeto TERMISUL, na Universidade Federal do Rio Grande do Sul (UFRGS), cuja origem foi motivada pela necessidade de produzir glossários

e dicionários especializados bilíngües (português-espanhol) em vista da intercomunicação dos países integrantes do Mercosul Krieger e Bevilacqua (2005), Maciel (1993). De forma isolada, mas igualmente importante, a Terminologia já era objeto de estudos na Universidade Federal de Pernambuco (UFPE) cujos trabalhos se dedicaram aos estudos das formações neológicas do português contemporâneo.

Apesar da Terminologia no Brasil ter iniciado no começo dos anos 90, início também da mudança de paradigma em direção à TCT, os primeiros trabalhos brasileiros seguiram a visão normativa, predominantemente da TGT. Todavia, os estudos nas universidades e centros de pesquisas, principalmente nos programas de pós-graduação, sob a luz da Lingüística e dos pressupostos descritivos da TCT vindos da Europa, contribuíram para superar os enfoques normativos da terminologia. No contexto dessa mudança, é importante destacar também a Associação Nacional de Pesquisa e Pós-graduação em Letras e Lingüística (ANPOLL), fundada em 1986, por meio dos trabalhos do Grupo de Lexicologia e Lexicografia.

Como se percebe, as universidades brasileiras, junto com a ANPOLL, tiveram um papel crucial na determinação da terminologia brasileira. Recentemente, outras universidades entraram nesse cenário de pesquisa e desenvolvimento, como a Universidade do Vale do Rio dos Sinos (UNISINOS) no Rio Grande do Sul, a Universidade Federal do Ceará (UFCE), a Universidade Estadual de São Paulo (UNESP) em seus campi de Araraquara e São José do Rio Preto, a Universidade Federal de São Carlos (UFSCar) e a Universidade Federal de Uberlândia (UFU). Dessa forma, os avanços qualitativos e quantitativos da Terminologia se devem, portanto, às pesquisas realizadas nessas instituições, principalmente sobre os objetos terminológicos, especialmente os termos, definições terminológicas, neologismos, e textos especializados.

Portanto, nos últimos anos o mapa terminológico no Brasil cresceu de forma significativa, permitindo-nos traçar um panorama dos grupos de pesquisas e estudos terminológicos nacionais. De acordo com Krieger e Bevilacqua (2005), há diversos grupos que foram consolidados e que integram o Diretório de Grupos de Pesquisa no Brasil no Conselho Nacional de Pesquisa e Desenvolvimento Tecnológico (CNPq), conforme mostra a Tabela 2.1.

Embora o traço das pesquisas em terminologia no Brasil tenham seguido as hipóteses “inovadoras” da TCT, é necessário salientar que apesar da terminologia brasileira ter acompanhado a visão lingüístico-comunicacional, ela é dotada de identidade própria, por meio de inúmeros trabalhos produzidos na esteira nacional que expressam postulados e comportamentos unicamente nacionais, principalmente aqueles que tratam de neologismos em textos especializados Krieger e Bevilacqua (2005).

No início da década de 2000 até os dias de hoje, como bem atesta Krieger e Finatto (2004) os estudos em terminologia têm se dedicado aos aspectos lingüísticos textuais nas línguas de

Tabela 2.1: Grupos de Pesquisas e Estudos Terminológicos no Brasil no CNPq.

UNIVERSIDADE	GRUPO DE PESQUISA
Universidade Estadual do Ceará (UFCE)	Tradução, Lexicologia e Processamento da Linguagem
Universidade Estadual de Londrina (UEL)	Grupo de Estudos Terminológicos
Universidade Estadual de São Paulo (UNESP)	Tradução, Lingüística Computacional, Estudos do léxico e dos dicionários
Universidade de Brasília (UnB)	Léxico e Terminologia
Universidade Federal de Goiás (UFG)	Léxico-Gramática e terminologia
Universidade Federal do Rio Grande do Sul (UFRGS)	TERMISUL
Universidade do Vale do Rio dos Sinos (UNISINOS)	TermiLex
Universidade Federal de São Carlos (UFSCAR)	Grupo de Estudos e Pesquisas em Terminologia (GETerm)
Universidade de São Paulo(USP)	Lexicologia, Lexicografia, Terminologia, Terminografia
	Observatório de Neologismos Científicos e Técnicos do Português Contemporâneo
	Núcleo Interinstitucional de Lingüística Computacional (NILC)

especialidade sob diferentes abordagens, destacando-se neste meio a formação e identificação das especificidades dos termos, como pode ser visto no trabalho de Ieda Maria Alves (Alves (1999)), bem como seu tratamento, baseado no princípio de que a identificação dos termos, para fins de criação e organização de repertórios terminológicos, deve estar vinculada ao exame dos universos de discurso em que os mesmo aparecem. Nesse campo, mais do que analisar os contextos discursivos onde os candidatos a termos se inserem, ocorre a adoção de princípios comunicacionais para as investigações terminológicas descritivas e aplicadas.

Observa-se, então, que os estudos de natureza terminológica e a produtividade da pesquisa brasileira têm contribuído para a fixação e consolidação da Terminologia não só no Brasil, mas também no mundo ibero-americano, dado o reconhecimento da importância destes estudos nos aspectos terminológicos contemporâneos.

2.3 Interfaces da Terminologia

O caráter interdisciplinar da Terminologia faz com que existam várias interfaces dela com outros saberes, tanto aqueles que representam a teoria e prática da Lingüística, quanto aqueles que representam o conhecimento específico das áreas fins, que servem como objeto de trabalho das práticas terminológicas. Nesse conjunto de saberes com que a Terminologia se relaciona, destacamos aqui somente a Lexicologia e a Lexicografia a Terminografia e a Tradução, apesar de existirem outros campos de conhecimento como a Semântica e a Documentação.

2.3.1 Lexicologia

De modo genérico, a Terminologia é uma disciplina cujo objeto primordial de estudo é o termo técnico-científico. Entretanto, ocupar-se de uma unidade lexical também é trabalho da Lexicologia, um ramo da Lingüística que trata do estudo do léxico, mais especificamente, do

conjunto de palavras de uma língua.

Dessa maneira, nos estudos da Lingüística, a Terminologia e Lexicologia são constituídas como ciências do léxico, porém distinguem-se pelas especificidades dos seus objetos. Assim, como salientam Krieger e Finatto (2004), a diferença entre elas depende da especificidade das unidades lexicais. A Lexicologia ocupa-se do componente lexical geral, não especializado das línguas. Por sua vez, a Terminologia ocupa-se dos termos, que denominam ou nomeiam processos e conceituações pertinentes a uma determinada área do conhecimento.

2.3.2 Lexicografia

Se a Lexicologia ocupa-se das palavras do léxico geral de uma língua, a Lexicografia é definida como a arte ou técnica de compor dicionários. Em outras palavras, podemos entender a Lexicografia como a parte prática da Lexicologia, tratando-se da criação e do fazer metodológico dos dicionários das línguas gerais. Tais dicionários registram o léxico geral de uma língua, reunindo seu conjunto de palavras e locuções de maneira mais abrangente possível, possuindo informações etimológicas, gramaticais, ortográficas e de gênero Hernández (2002).

A primeira visão leva-nos a identificar que a Lexicografia possui um caráter evidentemente prático, passando a entendê-la como atividade unicamente aplicada. Entretanto, o fazer lexicográfico é orientado por um paradigma teórico metodológico sintonizado com os propósitos de estudar o funcionamento morfossintático do léxico, categorizando a Lexicografia como um ramo da Lingüística Aplicada, conferindo a ela caráter descritivo, que define seu objeto de trabalho, postula princípios, e descreve os problemas e métodos envolvidos nas tarefas lexicográficas.

Nesse contexto, a dualidade teórica e prática da Lexicografia consiste em um dos pontos de interface com a Terminologia, já que esta última também comporta uma dimensão dupla, sendo que a parte prática é representada pela Terminografia (como será visto na Seção 2.3.3). Além disso, as duas áreas se envolvem com o componente léxico das línguas, seja de especialidade ou geral, buscando organizá-los como instrumentos de consulta e referência.

2.3.3 Terminografia

A Lexicografia Especializada ou Terminografia corresponde à face prática da Terminologia, direcionada para a produção de dicionários técnicos ou terminológicos, glossários e banco de dados de termos Meyer et al. (1997). Embora a Lexicografia e a Terminografia se pareçam em seus propósitos gerais, considerando seu aspecto aplicado da teoria, as duas áreas possuem objetos e metodologias de trabalho específicos Krieger e Finatto (2004).

A proposta da Terminografia é identificar e analisar as unidades lexicais usadas em domínios específicos do conhecimento, como a medicina, o direito, as engenharias ou a ciência da com-

putação, para os fins de criação e elaboração de obras de referência. Nesse sentido, a Terminografia toma o termo técnico-científico como seu principal objeto de definição, diferentemente da Lexicologia que usa as palavras do léxico geral.

Entretanto, como salienta Hernández (2002) definir a Terminografia como vertente da Terminologia encarregada da elaboração de dicionários especializados é uma definição muito genérica, pois é preciso ressaltar que esta área não se restringe a uma visão pragmática de criação destes produtos, mas é também um estudo sobre os termos. Assim, o propósito aplicado da Terminografia exige que os fundamentos teóricos necessários à identificação e tratamento dos termos sejam utilizados, tanto para o reconhecimento da variedade de suas formas quanto para a análise do seu funcionamento, principalmente para o adequado registro nas obras de referência especializada.

Dessa forma, os estudos terminográficos fornecem subsídios fundamentais para o estabelecimento de princípios metodológicos e diretrizes para o fazer terminológico, cuja estrutura de trabalho varia de acordo com o tipo e o conteúdo do produto final, seja ele um glossário, um dicionário monolíngüe ou multilíngüe, ou ainda um banco de dados terminológicos (tais produtos serão tratados com mais detalhes na Seção 2.4.4).

Independente do produto final, uma obra terminográfica é, com primazia, elaborada para oferecer informação técnica e específica de um determinado domínio do conhecimento, sobretudo no que diz respeito ao léxico com valor especializado, cujos conceitos são articulados pelas definições.

2.3.3.1 As aplicações terminológicas informatizadas

Segundo Meyer et al. (1997), os lexicógrafos alcançaram muito sucesso em suas tarefas com a adição do computador no seu dia a dia de trabalho. Da mesma forma, como sugere Maciel (1993) a tradução assistida por computador tem incrementado a rapidez e a facilidade do trabalho dos tradutores.

Da mesma forma, porém com menos intensidade, a eficiência e rapidez dos métodos práticos utilizados na Terminografia se beneficiaram dos progressos da informática. Os terminógrafos por sua vez, no início da informatização, usavam os computadores apenas para armazenar os conjuntos de dados lexicais, servindo apenas como um repertório de unidades lexicais com valor terminológico, enquanto que a maioria do trabalho terminográfico era realizada sem a utilização da informática e de forma manual Meyer et al. (1997).

O primeiro projeto que utilizou os recursos da informática para seu desenvolvimento foi o EURODICAUTOM (2004), que surgiu da conjugação de fatores políticos e sociais dos países da Europa, a partir da necessidade de um fluxo de comunicação uniforme, eficiente e rápida entre

estes países. No Brasil, um dos primeiros projetos a utilizar a informática foi o TERMISUL, nos anos 90, como pode ser visto em Maciel (1993) e Maciel et al. (1994).

O EUROIDICAUTOM marcou o nascimento de uma nova atividade no trabalho terminográfico: a junção da Terminologia e a Informática. A Terminologia e a Informática não são independentes uma da outra. Desde 1960, em países desenvolvidos e com grande tradição em pesquisa terminológica, a Terminologia e a Informática estão ligadas de forma a facilitar o armazenamento e a difusão de dados terminológicos na elaboração de grandes bases de dados especializados. A integração entre as duas áreas cunhou o termo Terminótica, dando origem a um novo conceito e inaugurando um novo paradigma metodológico nas pesquisas terminológicas.

De fato, entendendo a Terminologia como a ciência que estuda a formação dos termos, a Terminografia como atividade aplicada na constituição, gestão e difusão destes, pode-se compreender a Terminótica como o conjunto de operações e tarefas automáticas no tratamento e processamento dos termos. Dessa forma, a Terminótica oferece a possibilidade de semi-automatizar as atividades do terminólogo, fazendo com que a tarefa paciente, minuciosa, exaustiva e prolongada da confecção de dicionários, que se estendia por longos períodos, possa ser finalizada mais rapidamente.

Atualmente a Terminologia espera dos computadores o desenvolvimento de sistemas integrados, que possam utilizar conhecimentos empíricos e estruturados nas atividades do “fazer” terminológico Cabré (2005). Nesse sentido, as aplicações da Terminótica podem ser entendidas como sistematização das práticas terminológicas em contrapartida de uma carência, tanto sob a forma de uma metodologia de análise automatizada, quanto na criação de produtos terminológicos.

No Brasil, a contribuição advinda da Informática favoreceu o crescimento de estudos e pesquisas terminológicas na área de Linguística Computacional, e, conseqüentemente, o aprimoramento e desenvolvimento de ferramentas computacionais voltadas para o processamento de língua natural (PLN), que por sua vez, passam a interferir diretamente na prática terminográfica.

2.3.4 Tradução

Diferentemente da Terminologia, que começou a ser estabelecida como campo de conhecimento na segunda metade do século XX, a teoria e a prática da Tradução são de longa data e possuem tradição consolidada. Este avanço está ligado ao surgimento da Tradutologia, uma disciplina teórica que desenvolve reflexões sobre processo tradutório e seus diferentes aspectos.

Como parte da Tradutologia, a tradução técnica ou especializada esbarra na necessidade de utilização dos termos técnico-científicos, visto que estes também passam a ser objeto de trabalho dos tradutores, pois são elementos pertencentes aos textos especializados. Assim, este

tipo de tradução, que incide sobre os textos especializados, leva efetivamente os tradutores a se defrontarem com léxicos temáticos ou específicos.

Dessa forma, ao tradutor interessa um tratamento terminológico consistente, expressado pela adequada seleção, na língua de trabalho, dos termos equivalentes àqueles utilizados pelos especialistas na língua original. Isso porque, para o profissional da tradução é indispensável o conhecimento e o acesso aos repertórios terminológicos utilizado nas comunicações especializadas. Sob este enfoque, segundo Krieger e Finatto (2004), a utilização adequada da terminologia nos trabalhos de tradução contribui para a precisão semântico-conceitual da obra bilíngüe, requisito que todo trabalho tradutório obrigatoriamente requer.

Percebe-se então que os tradutores são usuários indiretos da Terminologia, visto que em virtude da natureza do seu trabalho, têm a preocupação constante com a eficiência e clareza da transferência conceitual e denominativa dos termos especializados. Entretanto, o contrário também é verdadeiro, pois os terminólogos, em benefício do seu trabalho, se relacionam com as atividades da Tradução na medida que se ocupam da elaboração de glossários, dicionários técnicos, e bancos de dados terminológicos multilíngües.

No conjunto das diferenças entre a Terminologia e a Tradução está a finalidade dos produtos de cada uma das áreas. Enquanto a Tradução tem o objetivo de produzir um texto em outra língua, mantendo a qualidade e a eficiência da comunicação, a Terminologia, ao gerar obras de referência específicas, realiza um trabalho de suporte, que serve para facilitar o trabalho dos tradutores.

Outra distinção entre a Tradução e a Terminologia é a metodologia de trabalho de cada uma delas. Cada área, em sua particularidade, possui suas próprias metodologias de trabalho, mediante ao produto final que oferecem.

Portanto, fica estabelecida a necessidade mútua existente entre a Tradução e a Terminologia, já que, sob este ponto de vista, o fazer tradutório é parceiro do fazer terminológico. Apesar disso, deve ficar claro que a Terminologia e a Tradução tratam-se de dois campos diferentes de pesquisa e de atuação, e não devem ser confundidos, embora a relação de parceria seja bastante produtiva.

2.4 Objetos da Terminologia

Embora a Terminologia seja uma disciplina de identidade própria e tenha os termos como objeto primordial de trabalho, outros objetos são essenciais para o desenvolvimento sistematizado de terminologias, tais como a Definição Terminológica, os Textos Especializados e os produtos do seu trabalho prático: os Glossários, Dicionários e Bases de Dados Terminológicos. Estes objetos merecem uma reflexão destacada, pois representam de diferentes maneiras as nuances e conceitos envolvidos no processo de criação de produtos terminológicos que transmitem o conhecimento

especializado.

2.4.1 O Termo

Começamos esta seção usando as palavras de Benveniste *apud* Krieger (2000) sobre a relação implícita existente entre os *termos* e as *ciências*:

“Poder-se-ia mesmo dizer que a história particular de uma ciência se resume na de seus termos específicos. Uma ciência só começa a existir ou consegue se impor na medida em que faz existir e em que impõe seus conceitos, através da sua denominação. Ela não tem outro meio de estabelecer sua legitimidade senão por especificar seu objeto denominando-o, podendo este constituir uma ordem de fenômenos, um domínio novo ou um modo novo de relação entre certos dados (Benveniste, 252).”

Parafraseando Benveniste, o conjunto de termos de uma área especializada representa o conhecimento dessa área ao mesmo tempo em que denomina seus conceitos. Os termos começam a existir quando se unem a conceitos determinados dentro do conjunto conceitual de uma área de especialidade.

Decorre do exposto que são os termos os primeiros traços que saltam aos olhos do leitor que lê um texto especializado ou científico, revelando à primeira vista as peculiaridades da área de especialidade Maciel (1998). Além disso, é unânime considerar que expressar e transferir o conhecimento de caráter especializado é o papel primordial dos termos e, consequentemente, da Terminologia Krieger (2001).

Sob uma perspectiva histórica, desde o surgimento e consolidação da Terminologia como área de conhecimento e pesquisa, os termos tiveram várias definições, principalmente devido às diferentes concepções teóricas aplicadas sobre a Terminologia. Sob os pressupostos da TGT, embora os estudos de Wüster concebessem o caráter interdisciplinar da Terminologia, os termos são considerados como unidades de conhecimento, caracterizando-se no sentido de que, para uma dada noção particular do conhecimento, há uma única denominação. Ou seja, entre o termo e o conceito existe uma relação unívoca e monoreferencial. Isso pode ser lido nas palavras do próprio Wüster *apud* Krieger (2001) que diz:

“Uma unidade terminológica consiste em uma palavra à qual se atribui um conceito como seu significado (Wüster, 21).”

Dessa forma, é possível estabelecer a diferença básica entre termos, no contexto da TGT, e as palavras da língua geral, visto que a compreensão de um termo está fundamentada no papel da dimensão conceitual a que se relaciona, neste caso, denominado de conteúdo especializado.

Outra propriedade vinculada à dimensão conceitual dos termos é a invariabilidade semântica. Isso porque o significado de uma determinada palavra do léxico geral está associada, na maioria das vezes, ao contexto discursivo em que está inserida, diferentemente do que ocorre com os termos, que se limitam a expressar conceitos específicos. É baseado nestas concepções sobre os termos que se destaca o caráter da monossemeia e univocidade conceitual da TGT Krieger e Finatto (2004).

Finalmente, percebe-se como denominador comum que, para a TGT, o estatuto terminológico de uma unidade lexical é definido pela sua dimensão conceitual, propriedade que a integra a um determinado domínio de especialidade.

Já na corrente histórica do pensamento da TCT, o termo é tido como unidade lingüística não sendo distinguido da palavra no ponto de vista do seu funcionamento. Consequentemente, os contextos lingüísticos e pragmáticos dos textos são elementos que configuram o status terminológico de uma unidade lexical. Isso pode ser visto pelas palavras de Cabré *apud* Krieger e Finatto (2004):

“... os termos não formam parte de um sistema independente das palavras, mas que conformam com elas o léxico do falante, mas ao mesmo tempo, pelo fato de serem multidimensionais, podem ser analisados de outras perspectivas e compartilham com os outros signos o espaço da comunicação especializada (Cabré, 119).”

Dessa forma, percebe-se que os termos fazem parte do conjunto de itens lexicais de uma língua e o valor terminológico é assumido a partir de um processo de “terminologização” sofrido por uma dada palavra em certo momento. Entende-se por terminologização o processo pelo qual as palavras da língua geral adquirem significados específicos, pertinentes a uma determinada área do saber científico, tornando-se então, elementos integrantes dos repertórios terminológicos Krieger (2001). Desse ponto de vista, uma unidade lexical pode assumir o valor de termo fazendo parte da representação dos princípios e propósitos de uma área.

A principal mudança de paradigma da TCT em relação aos termos é o entendimento de que os mesmos, mais do que elementos naturais dos sistemas lingüísticos são também componentes formais das linguagens em funcionamento, dada a sua presença nos textos especializados.

Atualmente, considera-se que o *habitat* natural das unidades terminológicas é o texto especializado, já que estes integram os processos discursivos da língua natural e sofrem os efeitos dos mecanismos dos sistemas lingüísticos. Decorre então que a partir da TCT surgiu uma nova visão sobre a constituição e funcionamento dos termos, cujo tratamento prescritivo das terminologias perde valor em detrimento dos enfoques descritivos, capazes de entender o léxico especializado como um elemento natural das línguas naturais, integrando as comunicações especializadas.

Nessa visão mais abrangente, apresenta-se uma estrutura tridimensional do termo, que compreende: 1) a *dimensão conceitual*, que expressa o conhecimento e os fundamentos dos saberes, 2) a *dimensão lingüística*, que determina sua naturalidade e integração aos sistemas lingüísticos, e 3) a *dimensão comunicacional* que agrega suas funcionalidades comunicacionais básicas para fixar e favorecer a transferência do conhecimento Krieger e Bevilacqua (2005).

É sobre este espectro dos termos que este trabalho de doutoramento se apóia, sendo mais restrito à valorização de seu componente conceitual, mas entendendo-os em toda sua complexidade constitutiva.

2.4.2 A Definição Terminológica

Em Terminologia, o enunciado definatório é uma componente chave na constituição e disseminação do conhecimento especializado, pois expressa o significado de um conjunto de coisas, objetos e processos pertencentes a uma área do saber humano. Sob esta perspectiva as definições em Terminologia identificam facetas específicas de compreensão dos fenômenos e valores das diferentes áreas do conhecimento. Assim, pelo exame cuidadoso de um conjunto de definições de uma área, torna-se possível uma percepção sobre o quanto de um conhecimento está sendo mobilizado e sobre como esse conhecimento pode ser multidimensional Finatto (1998).

No Dicionário Aurélio da Língua Portuguesa (edição eletrônica 2005) o significado da palavra “definição” remete ao verbo “definir”, que cujo enunciado é:

“definir
{Do lat. definire .}
Verbo transitivo direto
1. Determinar a extensão ou os limites de; limitar, demarcar:
2. Enunciar os atributos essenciais e específicos de (uma coisa), de modo que a torne inconfundível com outra:
3. Explicar o significado de; indicar o verdadeiro sentido de.”

Como pode ser visto no Aurélio, de maneira genérica, o ato de definir é caracterizado pela delimitação de fronteiras do sentido de uma determinada palavra, enfatizando as particularidades intrínsecas de maneira que transmita o sentido real e a caracterize unicamente. É a partir destes três significados que desejamos definir a tarefa de “definir” no âmbito terminológico.

Definir, no âmbito da terminologia é estabelecer um vínculo entre um termo, um conceito e um significado. A Definição Terminológica (DT) é responsável pela elaboração dos significados dos termos de um domínio específico em uma situação comunicativa específica, com o propósito de expressar os conceitos e transmitir um determinado saber ou uma porção do conhecimento especializado de uma área Finatto (1998).

Em contrapartida, como mencionado em Krieger e Finatto (2004), há diversos tipos de enunciados definatórios. A *definição terminológica* é reconhecida como aquela que mais se ocupa

de termos técnico-científicos. A *definição lexicográfica* é compreendida como aquela que mais se ocupa de palavras. A *definição enciclopédica*, por sua vez, contém informações variadas sobre um dado objeto da realidade.

De modo geral, a formulação do enunciado no conjunto dessas definições segue a referência delimitadora do **gênero próximo** e **diferença específica**. **Gênero próximo** é a porção do enunciado definitório que expressa a categoria ou classe genérica a que pertence o objeto está sendo definido. **Diferença específica** representa a indicação das particularidades ou características que distinguem este objeto em relação aos outros da mesma classe. Em outras palavras, o enunciado da definição deve sempre instanciar unicamente um objeto, salientando suas peculiaridades, em uma determinada classe. Por exemplo, uma definição de *PHP* seria: “*linguagem de programação para Web*”, cujo trecho “*linguagem de programação*” representa o gênero próximo, enquanto o trecho “*para Web*” a diferença específica.

Dessa maneira, uma boa definição deve salientar tanto o gênero próximo quanto a diferença específica, para que consigam, juntos, uma delimitação única, sob pena de impedir o eficaz entendimento da definição, principalmente se o objeto, representado pelo verbete de entrada, for novo para o consulente.

Por outro lado, a especificação das categorias do gênero próximo e da diferença específica nem sempre são suficientes para redigir uma boa definição. Isso ocorre, por exemplo, quando definimos a entrada *PHP*, pois em suas fronteiras definitórias cabem ainda outras linguagens como o *ASP* e o *JavaScript*. Nesse sentido a Definição Terminológica também configura uma situação que um objeto qualquer da definição precisa ser particularizado, classificado e diferenciado de outros objetos que lhe sejam aproximados.

Portanto, a separação de onde começa uma categoria e termina a outra num enunciado definitório pode ser incerta, de modo que não há margens seguras para garantir a eficiência da definição apenas por tais categorias. Além disso, fica nebulosa a distinção entre o que seria essencial, e estritamente definicional, e o que se pode chamar de acessório da definição. Tais acessórios podem ser uma gama de informações adicionais capazes de se tornar fundamentais tanto para Definição Terminológica quanto a compreensão do enunciado Finatto (1998), Krieger e Finatto (2004).

Por isso, reduzir a definição, como no exemplo de *PHP*, ao gênero próximo “*linguagem de programação*” e a diferença específica “*para Web*” equivale a desconsiderar outros elementos relevantes para o entendimento.

Conforme explicado em Krieger e Finatto (2004), encontramos o modelo de representação proposicional do enunciado definitório como alternativa da definição terminológica, além do gênero próximo e diferença específica. O enunciado, nessa sistemática de análise passa a ser

rescrito com auxílio de proposições explícitas que suportam o enunciado concreto. Assim, por exemplo, na definição “*PHP é uma linguagem de programação*” seria vista como uma representação lingüística na forma: {PHP (ser) {linguagem (ser) programação}}. Entretanto, para a definição “*PHP é uma linguagem de programação para Web, baseada em scripts e de código livre*” a representação seria mais complexa:

{PHP (ser) {linguagem (ser) {programação (desenvolvida para) Web (estar baseada em) {script (ser) de código livre}}}}}

A utilidade desse tipo de representação está na possibilidade de empreender um mapeamento de argumentos, entre sujeitos e predicados, do enunciado da definição de determinada especialidade. Esta proposição abre uma ótima questão de pesquisa em Terminologia, pois seria possível analisar quais nomes e verbos aparecem com mais frequência num conjunto de definições sob estudo, além de se poder perceber que tipos de relações seriam mais presentes entre os argumentos e conseqüentemente descobrir o perfil das definições.

Contudo a idéia de transformação de uma definição em conjunto proposicional não oferece um padrão suficiente de análise definitória, principalmente quando há definições com estruturas sintáticas mais complexas. Por exemplo, esta definição de PHP:

PHP é uma linguagem de programação de computadores interpretada, livre e muito utilizada para gerar conteúdo dinâmico na Web. Trata-se de uma linguagem modularizada, o que a torna ideal para instalação e uso em servidores Web. É muito parecida, em tipos de dados, sintaxe e mesmo funções, com a linguagem C e com a C++. Pode ser, dependendo da configuração do servidor, embutida no código HTML. Além disso, destaca-se a extrema facilidade com que PHP lida com servidores de base de dados, como MySQL, Microsoft SQL Server e Oracle.

No trabalho de doutorado de Finatto, finalizado em 2001 no Instituto de Letras da Universidade Federal do Rio Grande do Sul (UFRGS), há uma extensa discussão sobre pressupostos teórico-metodológicos da Definição Terminológica. Nele, a autora expõe outros métodos circunscritos na tarefa de definir em terminologia, como por exemplo, as concepções de Lakoff (1987) e Temmermann (1998), além de identificar a ausência de uma compreensão realmente lingüística da Definição Terminológica, principalmente no mapeamento do que é interno e externo no enunciado definitório da teoria terminológica de cunho lingüístico.

Portanto, dado que o objetivo deste trabalho não é uma discussão das vertentes teóricas da Definição Terminológica, a melhor opção é assumir como inerente a relação entre linguagem e conhecimento privilegiando nos enunciados definitórios e os significados a eles atrelados. Conforme mencionado Krieger e Finatto (2004), esse recorte é algo muito importante e tende a nos afastar das contradições.

Ainda neste contexto vale destacar os avanços em direção à sistematização da tarefa de

Definição Terminológica, principalmente na direção do mapeamento dos aspectos conceituais e lingüísticos que rodeiam este objeto Souza (2009).

2.4.3 Os Textos Especializados

A mudança de paradigma da Terminologia, de TGT para a TCT, ocorrida na segunda metade do século XX fez com que a Terminologia assumisse uma face lingüística que tomou como referência o exame do comportamento dos termos nos contextos onde ocorrem. Consequentemente, muitos dos trabalhos realizados sobre esta visão conceberam os termos como elementos da língua natural, valorizando principalmente os contextos discursivos das comunicações especializadas, com todas as implicações daí decorrentes.

Norteados por esse novo paradigma, as pesquisas sobre as linguagens especializadas também ganharam impulso, fazendo com que a Terminologia atuasse mais sistematicamente sobre as estruturas e tipologias de textos utilizadas nas comunicações especializadas. Dessa forma, a TCT considera o texto especializado como *habitat* natural dos termos, e objeto essencial na comunicação entre o autor e o destinatário Krieger e Finatto (2004).

Da mesma forma, tais pressupostos promovem a análise detalhada do comportamento dos termos em seu real contexto de ocorrência, nos textos especializados, compreendendo que essas unidades aparecem de maneira natural no discurso. Logo, decorre que os termos sofrem os efeitos de todos os fatores pragmáticos do discurso que suporta a comunicação especializada, de maneira que, a valorização do contexto discursivo, portanto, é de suma importância, pois os estudos das várias modalidades e temáticas apresentam especificidades e propósitos que contribuem para que uma unidade lexical seja reconhecida e alcance valor terminológico.

Dessa maneira, o conhecimento e a exploração dos textos especializados passam a ser requisitos metodológicos essenciais no contexto das pesquisas e aplicações terminológicas, pois são fundamentais para entender o funcionamento das terminologias.

Na Seção 3.1.3 do Capítulo 3, os textos especializados voltam a ter foco importante neste trabalho, dessa vez sob os aspectos dos Córpus de Especialidade como fonte terminológica.

2.4.4 Glossários, dicionários e bases de dados terminológicos

Genericamente, os glossários, dicionários e banco de dados terminológicos contêm o conjunto de termos de uma área especializada, isto é, a terminologia específica de um domínio do conhecimento. Estes três tipos de obras de referência, embora cumpram finalidades semelhantes de organização e apresentação dos termos, são diferentes entre si, e possuem características próprias.

Um Glossário de termos costuma ser definido como um repertório de unidades lexicais de

uma determinada especialidade, com suas respectivas definições e outras especificações sobre seus sentidos. Um Glossário terminológico não tem a pretensão de conter a totalidade de entradas que perfazem a terminologia de um domínio Krieger e Finatto (2004). Em outras palavras, pode ser entendido como um conjunto limitado de termos com seus respectivos significados e umas poucas informações adicionais, sem a necessidade de ser exaustivo.

Um Dicionário terminológico, por sua vez, é uma obra que registra o conjunto de termos de um domínio, oferecendo primordialmente informações conceituais e lingüísticas. A principal diferença em relação ao Glossário é a intenção de exaustividade, ou seja, o dicionário pretende alcançar a maior cobertura possível dos itens lexicais Maciel (1998). Entretanto, tal cobertura, depende do trabalho de revisão constante do dicionário, já que a vitalidade da língua sempre insere novos termos, ou variações de sentidos, em uma dada área. Portanto, ser um repertório exaustivo de termos e conter o significado destes em forma de definição são as duas principais características que identificam um Dicionário.

Dentre os tipos de Dicionários, vale destacar duas grandes categorias: o Dicionário de Língua Geral e o Dicionário Especializado. O Dicionário de Língua Geral apresenta um inventário, que tende a ser completo, do léxico de uma língua. Cada verbete de entrada do dicionário possui informações que representam todos os seus significados e usos possíveis, dados morfológicos, campos semânticos, quadro derivacional, níveis de uso, sinônimos, antônimos entre outras Maciel (1998). Todas essas informações são úteis para capacitar o consulente a reintegrar o verbete no contexto que foi encontrado e utilizá-lo de maneira eficiente para exprimir seu conteúdo semântico no processo de comunicação.

O Dicionário Especializado concentra-se em um recorte na língua geral e enfoca unidades lexicais especiais usadas em contextos particulares. Cada verbete de entrada neste dicionário representa um termo, e remete a um conjunto de informações que representa o significado específico do termo, e constitui a explicitação das características de um conceito dentro de um sistema conceitual Krieger e Finatto (2004). Dessa forma, o aspecto morfológico e a feição conceitual do termo, juntamente com os dados lingüísticos e enciclopédicos são apresentados para validar a pertinência temática e ou pragmática do termo no domínio de especialidade a que pertence.

Por último, um Banco de Dados terminológico é uma estrutura lógica informatizada que representa o relacionamento conceitual de dados terminológicos. Desenvolvido para armazenar e gerenciar um repertório de termos, além de uma série de outras informações, os bancos de dados terminológicos são peças fundamentais para o uso e funcionamento organizado das terminologias, principalmente quando inserido em um ambiente informatizado.

A principal característica dos Bancos de Dados terminológicos é a capacidade de *integração* e *estruturação* de um grande volume de dados em uma única estrutura lógica. A *integração*

consiste na incorporação em local único de informações terminológicas variadas; a *estruturação* compreende a organização sistemática e lógica destas informações.

Independentemente da diversidade de conteúdos que cada tipo de “repertório” de termos pode abarcar, e, principalmente, das diferenças estruturais que apresentam, há alguns elementos e princípios que são comuns e refletem como a Terminografia trata seus objetos de estudo. Dentre estes princípios, os que mais se destacam são as normas da *International Standardization Organization* (ISO) que, representada pelo Comitê Técnico 37, estabelece diretrizes úteis para elaboração de glossários e dicionários terminológicos, de modo a assegurar padrões mínimos de qualidade. As normas da ISO servem como orientações básicas para a implementação de bases de dados terminológicas. Conforme citado em Krieger e Finatto (2004), algumas destas normas são: ISO 860 (*Terminology Work, harmonization of concepts and terms*), ISO 704 (*Principles and methods of terminology*) ISO 704 (2000) e ISO 10241 (*International terminology stands - preparation and layout*).

Todavia, é importante estabelecer que embora as recomendações e diretrizes ditas pelas normas tornem o trabalho terminográfico mais organizado, é fundamental perceber as particularidades de cada empreendimento, principalmente observando os aspectos comunicacionais envolvidos e os objetivos almejados, já que somente obedecer as normas não garante o sucesso do trabalho terminológico.

2.5 Considerações finais

Este capítulo apresentou os princípios teóricos referentes à teoria terminológica e serviu de base para uniformizar e familiarizar os conceitos da Terminologia. Além da apresentação das escolas e teorias e do histórico da terminologia brasileira, foram delineados as interfaces com as áreas afins com destaque especial para os aspectos da informatização das práticas terminológicas. Adicionalmente, os principais objetos que concernem a Terminologia foram mostrados, evidenciando o termo, a definição terminológica e os textos especializados.

Estes elementos conceituais e teóricos são fundamentais para leitura do Capítulo 3 que destaca a face terminográfica e apresenta todas as etapas que permeiam o processo de gestão e criação de produtos terminológicos, mostrando os problemas e dificuldades encontradas e cada uma delas.

Etapas do trabalho terminológico: descrição e problemas

Conforme visto no Capítulo 2, a Terminografia corresponde à face prática da Terminologia, sendo responsável pela produção de produtos terminológicos que devem atender os seguintes requisitos Krieger e Finatto (2004):

1. As necessidades de um público-alvo ou geral devem ser atendidas para, de preferência, preencher uma lacuna de informação;
2. Todos os dados registrados ou utilizados devem ser confiáveis para futuras versões do produto;
3. A organização dos dados deve ser sistemática e seguir padrões de normas nacionais ou internacionais;
4. A estrutura dos dados sobre os termos na ficha de registro ou de uma base de dados deve atender aos objetivos do trabalho, principalmente em relação à organização de entradas do dicionário/glossário.

O processo de elaboração dos produtos terminológicos é longo, geralmente lento, demanda um grande volume de pesquisa na área alvo e necessita de uma equipe de trabalho multidisciplinar, formada de vários perfis profissionais. Nesse sentido, importa saber também que a colaboração dos especialistas da área alvo é indispensável para se produzir uma obra de qualidade que seja realmente útil para o consultante. Além disso, é preciso refletir sobre a natureza

e as peculiaridades do trabalho diante das especificações e objetivos que se quer atingir. Do mesmo modo, é necessária uma reflexão de cunho lingüístico inerente a este trabalho e uma contínua revisão da *nomenclatura*¹ durante sua composição Pavel e Nolet (2002).

Previamente, antes de iniciar o trabalho, é preciso caracterizar e identificar a área alvo que será focada. Esse recorte é muito importante para evitar que algumas tarefas (por exemplo: a pesquisa textual, a extração de termos, a definição terminológica, entre outras) não extrapolem os limites da área alvo apontada. Após este reconhecimento inicial, a publicação ou disponibilização do produto terminológico será a fase final do processo integrado por uma sequência de etapas de pesquisa e de trabalho.

Para este trabalho de doutoramento, propomos sistematizar um método de trabalho que, além de ser eficiente para a geração de produtos terminológicos, está em consonância com os postulados da Teoria Comunicativa da Terminologia (TCT). Assim, estabelecemos como método de trabalho uma sequência de etapas que entendemos serem constitutivas de qualquer projeto terminológico com fins terminográficos e que tenham tal filiação teórica. Essas etapas são as seguintes: 1) compilação de *cópus* de especialidade; 2) coleta (ou extração) de candidatos a termos; 3) elaboração do mapa conceitual ou *ontologia*; 4) inserção dos candidatos a termos na ontologia e sua validação (dos candidatos a termos) pelos especialistas; 5) elaboração e incremento da Base Definicional e criação e preenchimento das fichas terminológicas; 6) edição dos verbetes e o uso de formatos padrões de alcance internacional para facilitar o intercâmbio e/ou difusão dos produtos terminológicos. Esta metodologia, conforme foi descrito no trabalho de Almeida (2000), pode oferecer dificuldade para o especialista no momento de validar os candidatos a termos, pois estes não estão alocados numa ontologia, impedindo que seja analisado o contexto de uso dos termos. Outras metodologias, como a descrita na pesquisa de Krieger e Finatto (2004) não consideram a compilação do *cópus* como a primeira etapa, apesar do reconhecimento e extração dos candidatos a termos ocorrer predominantemente em *corpus* textuais. Ainda, como acontece nos processos terminográficos do Departamento de Tradução do governo Canadense, a ordem de algumas dessas etapas são trocadas, como no caso em que a etapa de criação da ontologia acontece antes da extração de candidatos a termos Pavel e Nolet (2002).

No nosso caso particular, resolvemos estabelecer a compilação de *cópus* de especialidade como etapa da metodologia, por entender que este recurso é de fundamental importância para a qualidade de produto terminológico, já que a grande maioria dos termos constantes no produto final vem destes *cópus*. Além disso, o uso de *cópus* especializados é essencial para as etapas de: extração de candidatos a termos, elaboração da Base Definicional, preenchimento da Ficha Terminológica, entre outras.

¹Nomenclatura consiste na lista de verbetes ou entradas que perfazem todo o dicionário ou glossário.

Nas próximas seções, faremos uma explanação das seis etapas de um projeto terminológico, relatando as atividades, as características e os problemas envolvidos em cada uma delas.

3.1 Etapa 01 - Compilação de corpus de especialidade

Atualmente, nos estudos da lingüística, quando se investiga os diferentes aspectos de uma língua, seja para fins de descrição lingüística, lexicológicos ou terminológicos, o uso de corpus é fundamental, porque facilita a comparação de diferentes aspectos léxicos, gramaticais ou textuais de diversos tipos de textos ou gêneros Hernández (2002). Além disso, a abordagem baseada em corpus oferece uma visão mais objetiva da língua do que o uso de introspecção ou intuição porque os falantes da língua não têm acesso aos padrões subliminares que ocorrem numa língua Sinclair (1997). Com o uso de ferramentas de análise de corpus pode-se não somente se investigar os padrões de uso, mas a sua frequência e os fatores contextuais que afetam a variabilidade deles.

Nos trabalhos de terminografia esse uso ainda é mais marcante, pois é praticamente impossível entender o funcionamento e a natureza dos termos sem ter em conta os contextos comunicativos e lingüísticos em que os mesmos aparecem. Assim, é importante destacar que os corpus representam um recurso lingüístico fundamental no processo de criação de produtos terminológicos.

3.1.1 Visão geral de corpus

Definir o conceito de corpus tal como é utilizado atualmente no âmbito da lingüística ou da lingüística computacional não é tão simples como pode parecer em uma primeira impressão. Genericamente, pode-se chamar de corpus qualquer coleção de textos selecionados para caracterizar uma língua ou variedade de língua Sardinha (2004). No entanto, a definição de corpus possui uma série de envolvimento que vão além de um simples “conjunto textual”.

Um das principais características é que os textos que compõem o corpus devem ser necessariamente textos naturais, produzidos por humanos, retirando desse conjunto aqueles textos criados por geração automática. No que compete à criação do corpus, é essencial ter bem definido o propósito de sua criação. Assim, temos a definição mais completa, ainda tomada de Sardinha (2004):

“A Lingüística de Corpus ocupa-se da coleta e exploração de corpora, ou conjuntos de dados lingüísticos textuais que foram coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade lingüística. Como tal, dedica-se à exploração da linguagem através de evidências empíricas, extraídas por meio de computador. (Sardinha, 2004: 3)”

Com o advento da lingüística computacional, tornou-se comum considerar qualquer coletânea de textos em formato eletrônico como um *corpus*. Nesse contexto, a definição de *corpus* expressa como um conjunto de textos em formato eletrônico que podem ser processados por computador é inválida, pois não segue os preceitos delineados no parágrafo anterior. Dessa forma, para dar conta da possibilidade do armazenamento eletrônico dos *corpus* Sardinha (2004) baseado em Sanchez (1996) define *corpus* como:

“Um conjunto de dados lingüísticos (pertencentes ao uso geral ou escrito da língua, ou ambos) sistematizados segundo determinados critérios, suficientemente extensos em amplitude e profundidade, de maneira que sejam representativos da totalidade do uso lingüístico ou de algum de seus âmbitos, dispostos de tal modo que possam ser processados por computador, com a finalidade de propiciar resultados vários e úteis para a descrição e análise.”

Decorrem dessa definição seis aspectos fundamentais que devem ser considerados: 1) o *corpus* deve ser autêntico, composto por textos produzidos por humanos em situações reais; 2) a inclusão dos textos que compõem o *corpus* deve ser guiada por uma série de critérios lingüísticos explícitos para assegurar seu uso como mostra representativa da língua 3) deve haver um propósito específico de pesquisa ou estudo; 4) os textos devem ser processados por ferramentas computacionais; 5) o *corpus* deve ser representativo em relação ao seu propósito e 6) o *corpus* deve ser extenso para ser representativo.

Dentre esses aspectos destaca-se uma importante característica dos *corpus*: a *representatividade*. Sob essa característica, o que mais se associa à idéia de representatividade de um *corpus* é o seu tamanho ou extensão, baseado no pressuposto de quanto maior o *corpus* mais representativo ele é Sardinha (2000). Embora existam tentativas relacionadas à frequência lexical (palavras de maior e menor frequência), bem como no cálculo da relação entre população e amostra, os critérios objetivos (estatísticos) para determinação da representatividade do *corpus*, como os citados em Biber (1993), são raramente usados por linguistas de *corpus* e lingüistas computacionais.

Desse modo, já que não se trata da representatividade objetivamente, pode-se tratá-la de maneira relativa, como, por exemplo, pelo aumento do tamanho do *corpus*. Isso porque a probabilidade de um *corpus* extenso apresentar mais instâncias lexicais é maior do que em um *corpus* pequeno, justamente porque será uma amostra mais abrangente da totalidade de traços lingüísticos presentes na língua.

Conforme ressalta Sardinha (2004), o tamanho do *corpus* é representado por três dimensões. A primeira é a quantidade de palavras, no sentido de que quanto maior o número de palavras maior será a probabilidade do *corpus* conter palavras de baixa frequência. A segunda é a

quantidade de textos que compõem o *corpus*, a qual se aplica a *corpus* especializados. Um número de textos maior pode garantir a variedade da tipologia textual ou dos gêneros, sendo assim melhor representado. A terceira é a própria quantidade de tipos textuais ou gêneros, aplicados principalmente aos chamados *Córpus de Referência* de uma língua², criados para representar uma língua em sua totalidade. Nesse caso, um número maior de tipos de textos ou de gêneros representa uma maior abrangência do espectro genérico da língua.

Contudo, apesar da existência de todas essas medidas de extensão, “medir” a representatividade de um *corpus* ainda pode ser difícil, demandando uma avaliação muito abstrata do profissional responsável por sua compilação ou dos usuários do mesmo. Ou seja, a atribuição da representatividade depende da observação dos resultados das pesquisas ou da avaliação subjetiva da qualidade do *corpus* baseada nestes resultados. Esse é, portanto, um dos principais problemas encontrados na tarefa de compilação de *corpus*.

Em Terminologia, o uso de *corpus* é baseado nos chamados *Córpus de Especialidade*. Essa categoria de *corpus* corresponde a um conjunto de textos de uma área ou domínio específico do conhecimento. Compostos pelos textos especializados, geralmente são bem menores do que os *Córpus de Referência*, porém, muito mais representativos de uma área específica da linguagem, o que demonstra uma menor variação lexical do discurso.

Muitos *corpus* de especialidade são compilados por especialistas ou terminólogos com o propósito deles servirem a projetos particulares ou destes pesquisadores estudarem os aspectos lingüísticos da área objeto (como a Terminologia, por exemplo), daí seu tamanho menor e a não disponibilização para a comunidade científica Nascimento (2003). A próxima seção discute o uso de *corpus* de especialidade informatizados no âmbito da terminologia.

3.1.2 O uso de *corpus* de especialidade informatizados em terminologia

O reconhecimento de que é preciso considerar os contextos lingüísticos de ocorrência dos termos para cobrir todo seu funcionamento está se impondo como orientação metodológica do trabalho terminográfico, principalmente aqueles contextos presentes nos textos especializados.

E linhas gerais, o texto especializado pode ser entendido como um instrumento de comunicação profissional e especializada do conhecimento humano. É composto por um conjunto de elementos lingüísticos que comunicam uma determinada realidade específica marcada pela riqueza de termos e enunciados conceituais que carregam consigo o saber especializado. Nesse sentido é indissociável a relação entre os textos especializados e as linguagens de especialidade. De modo geral, as linguagens de especialidade consistem em um subconjunto léxico que em de-

²Como exemplos de *Córpus de Referência* podemos citar o BNC (<http://www.natcorp.ox.ac.uk/>) do Inglês Britânico, o ANC (<http://americannationalcorpus.org/>) do Inglês Americano e o Lácio-Ref, criado no Projeto Lácio-Web, (<http://www.nilc.icmc.usp.br/lacioweb/>) do Português Brasileiro.

terminadas situações comunicacionais adquirem traços peculiares de significado. De acordo com Cabré (1993), as linguagens de especialidade podem ser classificadas de três formas. A primeira classificação considera as linguagens especializadas como um subconjunto léxico diferenciado da língua geral, com traços e regras específicas. Entretanto, delimitar a fronteira existente entre língua geral e língua de especialidade não é trivial, visto que não se sabe quais fenômenos lingüísticos as diferenciam. A segunda classificação considera as linguagens de especialidade como simples variantes da língua geral, fazendo parte do conjunto total da língua. A terceira classifica as linguagens especializadas como subconjuntos léxicos que englobam tanto a língua geral quanto as possíveis variações existentes. A Figura 3.1 mostra um esquema gráfico das três classificações, da esquerda para a direita, respectivamente.

Figura 3.1: Possíveis classificações entre língua geral e língua de especialidade Hernández (2002).

Todavia, como ressalta Lehrberger *apud* Hernández (2002), é difícil identificar quais variações das línguas de especialidade as diferenciam da língua comum, a não ser pela observação dos aspectos pragmáticos ou extralingüísticos, tais como:

1. temática ou assunto específico;
2. restrições léxicas, sintáticas e semânticas;
3. regras gramaticais especiais;
4. alta frequência de determinadas construções léxicas;
5. estrutura textual
6. uso de símbolos morfológicos específicos.

Ainda, em adição a estes seis aspectos, é preciso salientar um fator particular das línguas de especialidades que as diferenciam da língua geral: a transferência e o intercâmbio de informações especializadas. Ou seja, sua função comunicativa é o que diferencia o seu uso, do uso comum da língua geral.

Fica reconhecido então, a partir do descrito na seção anterior, que as línguas de especialidade constituem os textos especializados que por sua vez, em conjunto, formam os *córpus* de especialidade. Além disso, cabe salientar que, embora a especificidade inerente aos textos especializados

torne-se um obstáculo para a compilação e oferta abundante de *córpus* de especialidade, é imperativo compreender que as atividades e pesquisas da terminografia sobre os termos não podem desconsiderar esses *córpus* Sarmiento e Maia (2003). Portanto, a Terminologia interdisciplinar apoiada nos pilares lingüísticos da TCT vê-se motivada para o uso de *córpus* de especialidade devido à dimensão conceitual dos termos ³. Para poder identificar os termos que os especialistas usam no contexto discursivo, os terminólogos precisam estudar as estruturas de conhecimento (os conceitos e suas relações) que os termos representam e comunicam. Isto é, os terminólogos devem adquirir conhecimento sobre o domínio de especialidade para poder sistematizar sua terminologia. Para realizar esse trabalho este profissional tem duas opções: 1) consultar os especialistas da área e/ou 2) estudar e analisar a documentação especializada (*córpus*) da área.

A primeira das opções, a consulta direta aos especialistas, apesar de insubstituível e valiosa, pode apresentar problemas práticos, por outros dois motivos: 1) não é sempre possível tê-los à disposição, devido a outras atividades que executam no cotidiano; e 2), podem ter dificuldades em explicar o significado e o uso da língua que usam, pois muitas vezes o fazem de maneira intuitiva e não lingüística. Logo, essas duas razões fazem com que seja determinante o uso de um *córpus* de especialidade informatizado. Por sua vez, o estudo dos *córpus* de especialidade pode ser útil para facilitar a comunicação entre o especialista e o terminólogo. Na maioria das ocasiões, será mais fácil para o especialista entender as perguntas do terminólogo se essas se referem a um uso, significado ou definição específica de um termo que o especialista pode ver em um texto ou em uma estrutura conceitual Hernández (2002).

Sozinhos, esses dois motivos são suficientes para sustentar a importância do uso dos *córpus* de especialidade no contexto da terminologia. No entanto, adicionam-se nesse conjunto mais dois aspectos imprescindíveis de seu uso: 1) a identificação da variação lexical e conceitual dos termos e 2) o estudo de neologismos.

A análise do *córpus* especializado é uma das tarefas mais difíceis e exigentes do trabalho terminológico. Além da atenção e perspicácia exigida por parte do terminólogo, essa tarefa apresenta dificuldades porque é necessária a pesquisa textual integrada, levando em consideração informações conceituais, lingüísticas e pragmáticas inerentes ao discurso presente no *córpus*. Um exemplo dessa dificuldade é a atividade de procura pelas variações no uso dos termos ou na denominação dos conceitos.

Esse é o principal motivo de uso de *córpus* de especialidade informatizado no âmbito do trabalho terminológico, já que a capacidade de armazenamento, o poder de processamento dos computadores e a Internet aumentaram muito na última década. Nesse contexto, importa salientar

³As dimensões dos termos serão particularmente estudadas na Seção 3.2.1. Brevemente, para o correto entendimento, a dimensão conceitual dos termos se refere ao conteúdo ou ao conceito a que sua denominação se liga.

que existam atualmente muitas ferramentas computacionais para gestão, análise e tratamento de corpú, tanto ferramentas *stand-alone* como é o caso do *WordSmith Tools*⁴, o *XAIRA*⁵ XAIRA (2005) (*XML Aware Indexing and Retrieval Architecture*) e o *Unitex*⁶, quanto as disponíveis na Web como, por exemplo, o *Corpógrafo* Sarmiento et al. (2004) e as criadas no âmbito do Projeto Lácio-Web Aluísio et al. (2003), Aluísio et al. (2004). Neste conjunto, apenas o *Corpógrafo* pode ser utilizado para gestão e análise de corpú de especialidade para fins terminológicos. O restante são ferramentas de gestão de corpú em geral e na maioria das vezes são utilizadas de forma adaptada para este fim.

3.1.3 Compilação e análise de corpú de especialidade informatizado

Conforme visto na seção anterior, as pesquisas e atividades da terminografia estão intimamente ligadas aos corpú de especialidade, de maneira que a construção e implementação de ferramentas computacionais de análise e exploração desses recursos tornaram-se essenciais. Nesse contexto, a importância desses recursos ganha destaque na identificação e caracterização das unidades lexicais que dependendo do uso em contextos textuais específicos adquirem valor terminológico. Dada essa importância, surge naturalmente a necessidade de se estabelecer regras e critérios para o projeto e a compilação de corpú de especialidade, aos quais se destacam a avaliação dos tipos, gêneros e domínios dos textos que compõem o corpú. Isso porque a qualidade de qualquer produto terminológico está diretamente relacionada com a qualidade e utilização desses recursos lingüísticos e computacionais Nascimento (2003).

Dessa forma, e diferentemente do que acontece com corpú da língua geral, que aceita pequenos desvios de compilação devido às grandes dimensões que possui, os corpú de especialidade exigem uma rigorosa atenção e cuidado em relação à qualidade e representatividade dos textos selecionados.

Para os fins terminológicos, o primeiro critério a ser seguido para a construção dos corpú de especialidade é garantir que os textos sejam *completos*, para que não se perca toda a riqueza de termos e de conhecimentos inerentes ao seu conteúdo. Como segundo critério, e também de muita importância, o corpú deve ser *aberto* para suportar constantes atualizações, no qual vão se acrescentando textos “novos” (recentemente produzidos) e excluindo textos “antigos” (obsoletos), de maneira que o corpú possa acompanhar as alterações que se registram nas terminologias na maioria dos domínios de conhecimento.

Além desses dois critérios gerais, há uma série de outros critérios de seleção de textos para compor o corpú. Esses critérios, como bem lembra Hernández (2002) são utilizados para seleção

⁴<http://www.lexically.net/wordsmith/>

⁵<http://xaira.sourceforge.net/>

⁶<http://www-igm.univ-mlv.fr/unitex/>

de textos de corpus gerais, da língua geral, mas também podem ser aplicados na seleção de textos para os corpus de especialidade. São divididos em dois grupos: **critérios externos** e **critérios internos**.

Os **critérios externos** são essencialmente não lingüísticos e classificam os textos pelas características externas do texto. Informações referentes ao emissor, ao seu público (para quem o texto é direcionado), às intenções comunicativas, o tipo de texto, gênero, domínio, origem e outros aspectos socioculturais e profissionais são verificados para determinar a adequação ou não de um texto para um corpus. Em outras palavras, os critérios externos tentam assegurar que um determinado corpus represente uma ampla variedade de contextos da área que pertence.

Muitos dos critérios externos podem seguir as orientações de classificação hierárquica definidas pelas normas internacionais do TEI⁷ (*Text Encoding Initiative*), conforme realizado no *Projeto TermiNáutica* Correia e Andrade (2002), Mineiro (2003), que teve como principal objetivo a construção de um corpus de especialidade relativo ao domínio das ciências náuticas, em português europeu, com cerca de 1,5 milhões de palavras.

Na compilação de um corpus de especialidade, os critérios externos devem ser delineados de acordo com as situações comunicativas dos especialistas do domínio, sejam entre eles próprios ou com o público geral. Segundo Hernández (2002), quatro situações comunicativas podem ser especificadas:

Comunicação entre especialistas : tipo de comunicação cujo conteúdo e vocabulário são altamente especializados e usados de maneira precisa;

Comunicação especialista-iniciante : tipo de comunicação cujo conteúdo é um pouco mais claro, direcionado aos iniciantes da área, apesar da utilização do vocabulário especializado;

Comunicação entre (semi-)especialistas e leigos : tipo de comunicação utilizada em ocasiões de uso menos preciso do vocabulário especializado;

Comunicação professor-aluno : tipo de comunicação utilizada no contexto pedagógico, cujo conteúdo e vocabulário são especializados. Empregada para o ensino de uma determinada área do conhecimento, dos conceitos e os termos utilizados.

Esses quatro tipos de situações comunicativas servem como critérios externos iniciais de seleção e classificação de textos do corpus de especialidade, não sendo, portanto, suficientes para seleção adequada de textos. A esses critérios somam-se outros, mais específicos, que guiam as decisões de inclusão ou exclusão de um texto no corpus, sendo eles:

⁷Mais informações sobre o TEI pode ser encontrada em <http://www.tei-c.org/P4X/index.html>

Quantidade de Textos : embora um corpus de especialidade seja, geralmente, menor do que um corpus de referência é importante não impor limites à quantidade de textos que serão incluídos no corpus. Isso se dá por dois motivos básicos: 1) atualmente, com a Web, é mais fácil encontrar textos de especialidade em formato eletrônico, e 2) uma maior quantidade de textos pode incrementar a riqueza de terminológica;

Qualidade dos Textos : apesar do critério de quantidade admitir a premissa de “quanto mais textos melhor” para a compilação do corpus, os textos devem cumprir critérios de qualidade essenciais, como a referência do(s) autor(es) do texto (deve ser um especialista, ou grupos de especialistas, ou ligados a uma instituição de reconhecido valor na comunidade científica) e ao conteúdo do mesmo. Além desses, pode-se adicionar a condição que o texto a ser incluído deve ter sido publicado, principalmente em revistas ou congressos que possuem corpo de avaliadores de alto nível ou notório prestígio;

Documentação Adicional dos Textos : está diretamente ligada à qualidade dos textos e consiste na quantidade e no tipo de informação de documentação acrescentada ao texto original. Pode ser de dois tipos principais: 1) informações sobre a origem do texto e as circunstâncias de produção, tais como: autor, instituição, data de produção, gênero textual, nível de especialização, etc. e 2) informações morfossintáticas das unidades léxicas que compõem os textos. Estas informações podem ser úteis para a avaliação dos critérios internos, já que permite o terminólogo analisar o texto de forma mais precisa, utilizando conhecimento lingüístico;

Domínio de Especialidade : mesmo que apareça óbvio, os textos devem enquadrar-se no âmbito de especialidade do trabalho terminológico. Este critério é importante porque faz com que o terminólogo defina as fronteiras do domínio de especialidade que é objeto do trabalho, fazendo com que os textos estejam incluídos dentro dessas fronteiras;

Data de Produção do Texto : refere-se à data de publicação dos textos. Esse critério depende do objetivo ou natureza do produto terminológico desejado. Por exemplo, se o objetivo do produto refere-se ao uso terminológico atual de uma área, é preciso incluir textos mais recentes, já que eles representam a feição atual de uso dos termos. Se ao contrário, o objetivo é estudar os termos “antigos”, é melhor selecionar textos com data de publicação mais velha. Em contrapartida, se o objetivo é estudar a evolução dos termos ou identificar o aparecimento de neologismos, é necessário conter textos de várias épocas, tanto os mais recentes quanto os mais antigos, compondo um corpus diacrônico;

Condição de Produção dos Textos : usando este critério, o terminólogo deverá analisar

qual o meio de produção do texto (oral ou escrito) e as características do texto em relação à autoria, a saber: se o texto é original (escrito por autor nativo ou não), ou traduzido.

Condição Factual : ainda que possa parecer estranho ou inútil, esse critério define se os textos são comunicações reais, ou transmitem coisas, fatos, ou processos verdadeiros da área de especialidade. É importante porque permite excluir do *cópus* os textos artísticos (de novela, teatro, romances, etc.), mesmo que estes façam alguma menção sobre a área de especialidade;

Tipo Textual : diz respeito à classificação dos textos do *cópus* em tipologias de caráter especializado, como artigos de divulgação científica (de revistas, seminários, congressos etc.), manuais, teses, dissertações, livros, comunicações especializadas, e alguns tipos de folhetos informativos;

Nível Técnico e Receptores dos Textos : esse critério faz referência à tecnicidade dos textos, pelos propósitos que foram escritos e pelo nível de conhecimento do público a que serão dirigidos. Entretanto, esse critério é raramente analisado de forma isolada, já que está incluído no tipo de comunicação que será escolhida.

Os **critérios internos**, por sua vez, são puramente lingüísticos e procuram classificar os textos de acordo com uma série de categorias linguísticas internas, referentes ao seu conteúdo morfossintático, semântico e pragmático. Entretanto, como afirma Sinclair *apud* Nascimento (2003), não existe nesse tipo de classificação nenhum direcionamento de relevância lingüística que oriente o terminólogo na procura de componentes internos adequados e satisfatórios que contribuam para seleção de textos.

Desse modo, a definição de critérios internos são, na maioria das vezes, dependentes da análise e pesquisa criteriosa do terminólogo ou lingüista sobre os textos, baseado na experiência e conhecimento empírico que esse possui sobre a língua e sobre as línguas de especialidade. Portanto, vale ressaltar que a tarefa de seleção de textos a partir dos critérios internos é bastante subjetiva e subordinada ao trabalho minucioso dos terminólogos.

Contudo, Nascimento (2003) também cita que algumas experiências têm sido feitas no intuito de classificar os textos de acordo com critérios internos, propondo uma análise dos textos por meio da especificação de critérios léxicogramaticais, das nominalizações⁸ e combinatórias lexicais (ou colocações). Portanto, é necessário analisar, de um lado, a densidade terminológica dos textos, para incluir textos de maior ou menor grau de densidade em função da especialização do produto terminológico. Por outro lado, também se deve analisar a densidade conceitual dos

⁸Nominalização é o processo de substituição de uma classe ou estrutura gramatical (tipicamente verbal) por outra nominal. Ou seja, a linguagem científica substitui os processos, expressos normalmente por verbos, por grupos nominais.

textos (que possui relação direta com a densidade terminológica, ainda que não seja assim em todos os casos⁹).

Decorre então que é difícil definir os critérios internos a serem seguidos para compilação dos corpúscos para que a seleção de textos consista em uma combinação equilibrada dos critérios externos e internos. Assim, examinar as ocorrências dos termos em contextos autênticos dos corpúscos de especialidade permite identificar aspectos lingüísticos que lhes estão associados e que, assim como os próprios termos, serão considerados nos trabalhos terminológicos. Isso deixa clara a importância de empregar esses corpúscos nas tarefas terminográficas, e conseqüentemente, a necessidade de utilizar critérios apropriados para a seleção dos textos.

3.1.4 Ferramentas computacionais para análise de corpúscos de especialidade

Notadamente, a análise e exploração manual de corpúscos textuais é uma das tarefas mais difíceis, morosas e de baixo rendimento do trabalho terminológico. Isso porque a análise dos corpúscos exige uma leitura cuidadosa dos textos, na tentativa de identificar o caráter lingüístico e os possíveis termos inseridos no contexto discursivo. Devido a essa dificuldade, juntamente com o fato de que os corpúscos de especialidade estão, em sua maioria, em formato eletrônico (digitalizados), torna-se indispensável o uso de ferramentas computacionais para sua exploração. Tais ferramentas, por sua vez, no âmbito do desenvolvimento dos corpúscos, seja ele de especialidade ou geral, atuam principalmente na análise das informações de caráter lingüístico, pertencentes ao conteúdo do texto. No contexto particular da compilação de corpúscos, as ferramentas informatizadas têm um papel fundamental na percepção e descoberta dos critérios internos de seleção de textos.

Segundo Hernández (2002), há dois tipos de análise de corpúscos que podem ser empregadas com o uso de ferramentas informatizadas. A primeira é a **Análise Qualitativa** que consiste na descrição detalhada e completa do fenômeno lingüístico e do comportamento de uma palavra ou grupo de palavras no corpúscos. A segunda é a **Análise Quantitativa** na qual vários índices estatísticos são associados aos fenômenos lingüísticos observados no corpúscos (pela análise qualitativa), de maneira que podem servir para explicar uma evidência ou particularidade do texto.

É relevante salientar que os dois tipos de análise são complementares, na medida que a análise qualitativa oferece uma grande riqueza e precisão nas observações lingüísticas consideradas, enquanto que a análise quantitativa oferece informação que seja estatisticamente significativa em relação a estas observações.

Existem vários programas, tanto comerciais quanto gratuitos (para fins acadêmicos), que oferecem os dois tipos de análise de corpúscos eletrônicos. Entre os gratuitos aparecem, por ex-

⁹Por exemplo, um texto especializado pode conter muitos termos, mas se for dirigido a um público que não requer explicações sobre eles pode conter pouca informação sobre os termos usados.

emplo, o *Conc* desenvolvido pelo *Summer Institute of Linguistics*¹⁰, o *FreeText Browser*¹¹ da *Universidade de Michigan*. Além destes, existem ainda uma série de ferramentas computacionais que são oferecidas como *Web Services*, via Internet, disponível para qualquer usuário da rede. Apesar desses serem facilmente encontrados na Internet, principalmente para o Inglês, podemos citar, o *TACTWeb*¹² da *Universidade de Toronto*, o *Corpógrafo*¹³ e o *AnELL*¹⁴ ambos da *Linguatca*, o *Lácio- Web*, o *Unitex*¹⁵ Muniz et al. (2005) e as Ferramentas do LAEL¹⁶ da Pontifícia Universidade Católica de São Paulo (PUC-SP), sendo os três últimos para o processamento do Português. Entre os programas ou pacote de ferramentas pagas o principal representante é o *WordSmith Tools*, implementado pelo pesquisador Michel Scott da *University Oxford Press*. Em Sardinha (2004) e Hernández (2002) há uma descrição detalhada de todas ferramentas de análise de cópús que esse pacote possui. As Figuras 3.2, 3.3 e 3.4 apresentam as telas iniciais dos programas *TACTWeb*, *AnELL* e *WordSmith Tools*, respectivamente.

Figura 3.2: Tela inicial do *TACTWeb*

Embora sejam muito úteis para a análise de textos e cópús, é importante salientar que esses programas e/ou ferramentas computacionais não são e não foram desenvolvidos especificamente para o propósito do trabalho terminológico. Isso porque o uso de cópús de especialidade em terminologia também é indispensável para extrair as relações conceituais que se apresentam nos termos e seus contextos. Portanto, é necessário ainda contar com ferramentas adequadas que permitam desenvolver estratégias de busca deste tipo de informação, para melhorar a análise da representatividade e qualidade do cópús. Nesse contexto, destacam-se ferramentas desenvolvidas no âmbito da engenharia do conhecimento, tais como o *Text Analyser* Kavanagh (1995) da

¹⁰Mais informações sobre o *Conc* pode ser conseguida no endereço: http://www.sil.org/computing/catalog/show_software.asp?id=9

¹¹Mais informações sobre o *FreeText Browser*, inclusive *download*, pode ser obtida no endereço: <http://www.georgetown.edu/faculty/ballc/corpora/tutorial3.html#RTFTtoC31>

¹²Mais informações e disponibilização no endereço: <http://tactWeb.mcmaster.ca/tactWeb/doc/tact.htm>

¹³Disponível em: <http://poloclup.linguatca.pt/ferramentas/gc/index.html>

¹⁴<http://acdc.linguatca.pt/AnELL/Service-automatico.htm>

¹⁵O Unitex está disponível no endereço: <http://www.nilc.icmc.usp.br:8180/unitex-pb/>

¹⁶Disponível em: <http://www2.lael.pucsp.br/corpora/index.htm>

Figura 3.3: Tela inicial do *AneLL*

Figura 3.4: Tela inicial do *Word Smith Tools* (versão 4.0)

Universidade de Ottawa, o *IKARUS* Meyer et al. (1997), o *CODE4* Skuce e Lethbridge (1995) e também o *Corpógrafo* descrito em Pinto e Oliveira (2004), Maia e Sarmiento (2005). Um exemplo de uso do *Corpógrafo* nesta tarefa é o experimento *Neurodemo*¹⁷ no qual um corpus de neurologia foi submetido à funcionalidade de extração semi-automática de definições deste

¹⁷Mais informações sobre este experimento/exemplo pode ser obtida no endereço: <http://poloclup.linguatca.pt/Neurodemo.htm>.

sistema para fins de demonstração e teste.

Ainda, sobre esse tema, o trabalho de doutorado de Judit Feliu (Feliu (2004)) trata especificamente das relações conceituais e a sua extração semi-automática no contexto da terminologia, visto que a localização de informação conceitual em um texto deve partir do estudo das estruturas linguísticas que se usam para representar essas relações. A Seção 3.3.1 apresenta alguns exemplos e os tipos de relações conceituais existentes entre os termos e a ontologia.

Realizada essa contextualização sobre as ferramentas de análise de *cópus* para a terminologia, fica clara a necessidade de duas ações pontuais. A primeira é a valorização e aplicação das metodologias e ferramentas já estabelecidas em outros âmbitos do estudo linguístico para a terminologia, e a segunda é desenvolver e delimitar com maior precisão os tipos de ferramentas necessárias para melhorar o trabalho terminológico. Na Seção 4.3.2 do Capítulo 4 tais ferramentas serão tratadas com mais detalhes.

3.1.5 Compilação automática de *cópus* de especialidade

Conforme pôde ser visto na seção anterior, a tarefa de avaliação dos textos dos *cópus* exige do terminólogo uma rigorosa avaliação dos critérios internos e externos para sua seleção. Entretanto, a tarefa de selecionar os textos que irão passar pela avaliação também não é fácil. Do mesmo modo, reunir o conjunto de textos que poderão fazer parte do *cópus* é uma tarefa morosa, que demanda tempo de pesquisa e leitura dos textos candidatos.

Atualmente, a maneira mais fácil de pesquisar textos para compor um *cópus* é utilizando a Internet. Com os atuais portais de busca (por exemplo, *Google* e *Bing*) é possível pesquisar uma grande e variada quantidade de textos, o que facilita a reunião e composição do conjunto textual para avaliação. Entretanto, mesmo utilizando a Internet, devido à enorme quantidade de resultados que os portais de busca produzem, essa tarefa ainda continua morosa, pois exige do terminólogo a visita (virtual) a cada resultado devolvido e, conseqüentemente, a seleção, a cópia e o salvamento do texto para o repositório local (computador).

Mesmo que o terminólogo tivesse a disposição para visitar cada site da Internet para fazer o *download* dos textos, esse trabalho em muito dificultaria a compilação de grandes quantidades de textos, dado o tempo considerável que se gasta com essa tarefa. Para conseguir melhores resultados na retirada de textos da Internet é necessário a utilização de uma ferramenta que executa o *download* sistemático de um *site* inteiro, a partir de uma URL, por exemplo. Essas ferramentas ou programas são chamados de *off browsers* e sua função é copiar, para o computador local, todo conteúdo de um *site* a partir de uma URL fornecida como entrada. Em outras palavras, esses programas fazem uma cópia fiel do conteúdo de um *site*, localizado na Internet, no computador do usuário.

Existem hoje diversos programas dessa categoria, tanto proprietários quanto livres. Como exemplo de *off browser* proprietário, podemos citar o *Teleport Pro*¹⁸ desenvolvido para Windows e que permite a cópia de *sites* inteiros a partir de uma URL de entrada. Como exemplo de programas livres, podemos citar o bem conhecido *WGet*¹⁹ desenvolvido inicialmente para o Linux e depois para Windows. Outro de exemplo de programa *off browser* livre é o *HTTrack*²⁰ desenvolvido para Windows. Uma das grandes vantagens desse tipo de programa é a especificação de um filtro de arquivos que serão recuperados, permitindo ao usuário recuperar apenas arquivos que contenham textos, por exemplo: HTML, PDF, DOC, TXT, PS, RTF, excluindo da busca arquivos de diversas mídias, como, por exemplo, figuras, imagens e sons.

Apesar de facilitar muito o trabalho do terminólogo, os programas *off browsers* possuem duas desvantagens. A primeira está relacionada com a identificação do *site* alvo da cópia. Nesses programas, a responsabilidade de encontrar qual *site* será alvo de cópia é do terminólogo. Ou seja, o terminólogo ainda precisa realizar suas pesquisas e descobrir quais *sites* ele deseja copiar para depois fornecê-los aos programas. A segunda desvantagem está relacionada com os filtros de seleção de textos. Mesmo que o terminólogo informe qual *site* será alvo da cópia, bem como quais tipos de arquivos serão recuperados (textos), esses programas realizam uma cópia cega não executando um refinamento dos textos a serem copiados.

É nesse contexto que aparece uma outra categoria de programas de compilação automática de cópys, os chamados *BootStrap*. Esses programas empregam uma outra maneira de recuperar textos a partir da Internet, pois além de baixar os textos, eles também efetuam as pesquisas nos portais de busca. A partir de um conjunto de sementes (*seeds* - que pode ser entendido como um conjunto de palavras utilizadas nas pesquisas) fornecidas, esses programas efetuam, automaticamente, a pesquisa nos portais de busca e, em seguida, copiam para a máquina local o conteúdo encontrado a partir dos resultados.

A principal vantagem desses programas é o fato de que os terminólogos não precisam efetuar a pesquisa de *sites* para posterior cópia, conforme ocorre com os *off browsers*, pois nesse caso, o terminólogo fornece apenas o conjunto de sementes que deseja pesquisar, e o programa se encarrega de pesquisar. Outra vantagem é a garantia que os textos recolhidos automaticamente satisfazem, ou pelo menos tem relação, com o conteúdo das sementes que foram pesquisadas. Entretanto, uma adequada especificação das sementes de busca é fundamental para execução desses programas, já que se tal especificação for mal elaborada, os resultados também não serão satisfatórios, sendo esta sua principal desvantagem.

¹⁸Mais informações e *download* do programa *Teleport Pro* podem ser obtidos no endereço: <http://www.tenmax.com/teleport/pro/home.htm>

¹⁹No endereço: <http://www.gnu.org/software/wget/> contém um manual completo de funcionamento do *WGet*, bem como informações de *download*.

²⁰Mais informações e *download* do programa *HTTrack* podem ser obtidos no endereço: <http://www.httrack.com/>

Podemos citar pelo menos dois exemplos de programas *BootStrap*. O programa pioneiro de construção de corpus via consultas a motores de busca na Web foi o *CorpusBuilder*²¹ Ghani et al. (2001), Ghani et al. (2005) que iterativamente cria novas consultas para construir um corpus em uma dada língua. Foi escrito em *Perl*²² e utiliza o *TextCat*²³ para definir a língua por meio de um ou mais documentos fornecidos pelo usuário.

O principal deles é o *BootCat* Baroni e Bernardini (2004) que executa consultas na base do *Google* a partir de combinações do conjunto de sementes fornecidas pelo usuário. Basicamente, a execução do *BootCat* pode ser dividida em duas fases. A primeira é a reunião de textos da Web, por meio de um processo iterativo, a partir da combinação das sementes fornecidas. A segunda é a extração de palavras compostas (chamados de *multiterms*) a partir do corpus compilado na primeira fase. Uma das grandes limitações do *BootCaT* é sua restrições de tipos de arquivo que são convertidos e incluídos ao corpus. As URLs retornadas pelo *Google* que se referem a arquivos em formato PDF (*Portable Document Format*) e DOC (*Microsoft Word Document*), por exemplo, são ignorados pela ferramenta. Normalmente, esses arquivos contêm textos mais informativos e confiáveis que os em formato HTML. O trabalho de Genovês e Aluísio (2005) descreve uma experiência de uso do *BootCat* em ambiente acadêmico.

Contudo, vale ressaltar que, independente do programa utilizado ou da maneira que o corpus seja compilado automaticamente, é indispensável executar a atividade de limpeza dos textos e, conseqüentemente, do corpus. Tal limpeza dos textos é necessária devido aos erros ocorridos durante a conversão dos formatos originais (encontrados na Web) para texto puro (extensão TXT). Isso ocorre porque os arquivos que correspondem aos textos oriundos da Web podem ser, na maioria das vezes, de várias extensões (PDF, DOC, HTML, etc.) e mantém formatações específicas que, no momento da conversão para texto, deixam resquícios de “sujeira” como marcações de parágrafos, negritos, itálicos, caracteres tipográficos, etiquetas, etc. que não são cobertos pela extensão TXT. Além disso, mesmo que os textos não estejam “sujos”, é necessária uma adequação correta dos textos ao estilo definido para o corpus, para não influenciar nos resultados das pesquisas.

Para o trabalho de limpeza dos textos pode se usar qualquer editor de texto puro, que não utiliza (ou permita) nenhuma formatação especial. Um bom exemplo é o *Bloco de Notas* (da Microsoft), o *Emacs* ou *Kate* (do KDE/Linux) ou mesmo os editores de textos online disponíveis da Web, como o existente no *Corpógrafo* Sarmiento et al. (2004).

²¹Mais informações, contato com autores e referências sobre o *CorpusBuilder* podem ser obtidas no endereço: <http://www.cs.cmu.edu/TextLearning/corpusbuilder/>

²²<http://www.perl.com>

²³Mais informações sobre o *TextCat* podem obtidas no endereço: <http://www.let.rug.nl/vannoord/TextCat/>

3.1.6 Questões relacionadas com os direitos autorais

Quando se fala em compilação de *cópus* e coleta de textos, principalmente a automática (da Internet), é necessário tratar dos aspectos relacionados aos direitos autorais dos textos. No contexto da terminologia, esses aspectos cobram importância ainda maior, já que as entradas dos verbetes costumam conter trechos (por exemplo: definições, conceitos ou contextos de uso dos termos) extraídos dos textos utilizados durante o trabalho.

Uma das maneiras de garantir a propriedade intelectual dos textos que serão utilizados no *cópus* está fundamentada em uns dos critérios externos especificados na Seção 3.1.3 para seleção dos textos, o qual indica que o texto deve ter sido publicado e, portanto, deve possuir os registros de ISBN (*International Standard Book Numbring*) ou ISSN (*International Standard Serial Number*). Da mesma maneira, embora estejam disponíveis na Internet, dando a entender que são de domínio público, esse critério também deve ser aplicado aos textos que são coletados pela compilação automática. Entretanto, é difícil recuperar tais registros ou números que garantem sua propriedade intelectual, configurando-se assim, uma dificuldade a mais para seleção de textos.

Nesse sentido, a solução adotada para garantir os direitos autorais dos textos pelos projetos *OncoTerm* de Hernández (2002) e pelo *Corpógrafo*, foi garantir o registro completo dos textos dos *cópus*, mesmo que isso signifique uma atividade adicional do terminólogo. Dessa forma, cada texto pertencente ao *cópus* possui um cabeçalho com o registro dos seus autores, da publicação, editora, etc. de maneira que na publicação final do produto terminológico, tais direitos estejam resguardados. Outra maneira, conforme adotada pelo *Projeto Lácio- Web*, é solicitar aos autores dos textos uma autorização formal para utilização dos mesmos.

Todavia, em meio à dificuldade de conseguir a autorização de uso dos textos, o que pode aliviar essa responsabilidade e não violar o direito de propriedade é garantir que os textos serão utilizados no meio acadêmico, para os fins de pesquisa e, ainda, garantir que o produto final não será vendido, ou configure qualquer lucro financeiro. Caso contrário, a questão dos direitos autorais torna-se muito séria e não pode ser deixada de lado.

3.1.7 Problemas referentes à compilação de *cópus* de especialidade

Diante do exposto até aqui, podemos delinear alguns problemas e dificuldades encontrados na utilização de *cópus* de especialidade no trabalho terminográfico, mesmo que seu uso seja indispensável e fundamental para a qualidade do trabalho.

A primeira dificuldade é genérica e está relacionada com a própria tarefa de compilar *cópus* de especialidade. Inicialmente, está a dificuldade de estabelecer um recorte bem definido da área alvo, que muitas vezes é trabalhoso, devido à interdisciplinaridade atualmente estabelecida no

campo das ciências. Da mesma forma, após a definição do recorte, a empreitada de encontrar, seja em qual fonte for, em formato de papel ou eletrônico (usando a Internet), os textos candidatos que provavelmente irão compor o *córpus* é custosa e exige do terminólogo muita atenção e dedicação às pesquisas, visto que é de fundamental importância encontrar fontes confiáveis e de qualidade reconhecida. É por este motivo que a compilação automática de *córpus* é vantajosa para esta tarefa (desde que executada de forma cautelosa), mesmo que a atividade de limpeza e adequação dos textos seja também morosa.

A segunda dificuldade encontrada diz respeito à seleção de textos. Uma vez feita a coleta de textos, inicia-se a tarefa de aplicar os critérios externos e internos para o processo de seleção e avaliação de textos que irão compor o *córpus*. Diante do número de critérios externos, avaliar cada texto de acordo com os tais torna-se um trabalho repetitivo. Nesse mesmo contexto, o exame destes critérios exige também discernimento por parte do terminólogo, visto que é a partir deles que se julga que um texto pode ou não fazer parte do *córpus*.

A terceira dificuldade decorre da segunda, e nos parece o principal problema da compilação de *córpus* de especialidade. Refere-se à avaliação e análise dos critérios internos ou lingüísticos dos textos coletados. Isso porque esta análise implica em medir a representatividade de um texto, levando em conta os aspectos lingüísticos e terminológicos. Conforme dito anteriormente nessa mesma seção, falta ainda um direcionamento que oriente o terminólogo a procurar elementos internos do texto que contribuam para o julgamento da “qualidade” de um determinado texto. Fica então, essa tarefa, condicionada à capacidade empírica do terminológico, baseada principalmente na sua experiência e percepção do comportamento da área alvo de pesquisa. Mesmo que hoje exista um rol de ferramentas computacionais que o auxiliam no exame dos textos, resta ainda a tarefa cognitiva de interpretação e avaliação dos resultados destas ferramentas.

Ainda no bojo do problema de avaliação dos critérios internos, podendo ser encarada como quarta dificuldade, está a carência de ferramentas informatizadas que extraem as relações conceituais dos termos nos textos, ainda que haja esforços para implementação de programas para este fim, como pode ser visto no *Corpógrafo*, com o experimento *Neurodemo* que apresenta um exemplo de extração semi-automática de definições e relações semânticas. Isso porque as relações conceituais dizem muito sobre a qualidade de um texto e seu respectivo valor terminológico para composição do *córpus*, já que nelas são expressas o conhecimento inerente do domínio.

3.2 Etapa 02 - Extração de candidatos a termos

Uma vez que o *córpus* de especialidade foi compilado, segue a segunda etapa do processo de criação dos produtos terminológicos. A extração de candidatos a termos diz respeito à obtenção do conjunto terminológico que comporá a nomenclatura do produto terminológico (do glossário,

dicionário ou base de dados). A fonte a partir da qual serão extraídos os candidatos a termos deve ser o corpus de especialidade.

A extração pode ser feita de duas formas: manual ou automática. Antes de dar prosseguimento a estas duas formas, é preciso entender um pouco mais sobre as dimensões dos termos, visto que é sobre eles que esta tarefa dedica sua atenção.

3.2.1 Unidades terminológicas e suas dimensões

O procedimento de extração de candidatos a termos está fundamentado no entendimento das diferentes faces que constituem as unidades terminológicas. Principalmente, no reconhecimento da relação que existe entre um determinado conceito e o termo que o designa e, consequentemente, da relação existente entre a área do conhecimento específico e os termos que ela utiliza para a comunicação especializada. Esse reconhecimento leva-nos a entender os termos como unidades tridimensionais, integrando ao mesmo tempo uma dimensão fundamentalmente *lingüística*, enquanto são unidades lexicais da língua; uma *dimensão conceitual* que representa conhecimento especializado, e uma *dimensão comunicativa*, já que é por meio deles que os especialistas da área podem transmitir o conhecimento especializado de forma eficiente Hernández (2002).

Portanto, os termos constituem unidades lexicais que possuem uma tripla face: 1) a lingüística, representada por meio da estrutura morfológica do termo e as diferentes possibilidades de formação e combinação lexical; 2) a conceitual, que representa o conceito a que a estrutura morfológica se refere e 3) a comunicativa, que representa a vertente discursiva dos termos na tarefa de comunicar o conhecimento e designar os objetos de uma realidade. Logo, um perfeito ajuste entre essas três dimensões garante que os termos cumprirão sua função de transmissão do conhecimento especializado.

3.2.1.1 A dimensão lingüística dos termos

De acordo com Hernández (2002), a dimensão lingüística das unidades lexicais se caracteriza sob quatro aspectos diferentes: a forma, a função, a semântica e a procedência lingüística, descritos abaixo.

Critério Formal : consiste do número de morfemas que compõe os termos, podendo ser simples ou complexo. Os termos simples são aqueles formados por um único morfema como, por exemplo: *sangue*, *vídeo*, etc. e também são chamados de unigramas. Os termos complexos podem ser **derivados**, se formados por base léxica e um ou mais afixos, como acontece em: *cardiologia*, *fisiologia*, *lexicologia*, etc. ou **compostos** quando formados por dois ou mais morfemas juntos como, por exemplo: *videoconferência*. Além disso, os termos

complexos podem ser formados por combinações de palavras, os chamados *N-grams*, em que *n* representa o número de palavras do termo, por exemplo: *hipertensão arterial* (bigrama), *ciclo de queima* (trigrama), etc. Também atendendo a critérios formais, existem termos complexos que são formados pelos processos de redução ou truncamento, como é o caso das siglas como, por exemplo: *AVC* (Acidente Vascular Cerebral) e *EAD* (Educação a Distância), os acrônimos, por exemplo: *terminótica* (terminologia + informática) e as formas abreviadas como: *auto* (automóvel) e *moto* (motocicleta);

Critério Funcional : consiste na função que os termos desempenham no discurso, de acordo com sua classe gramatical em um dado momento. Podem ser: substantivos, verbos, adjetivos e advérbios. Entretanto, também ocorrem formações com as preposições e artigos, principalmente nos termos com valor de *n* maior que três (trigramas, tetragramas, pentagramas, etc.);

Critério Semântico : refere-se à formação de termos baseada nas classes de conceitos que os mesmos denominam. Esses conceitos são classificados de acordo com as relações estabelecidas entre eles e podem ser:

1. Objetos ou entidades: substantivos;
2. Processos, operações, ações: verbos;
3. Propriedades, estados ou qualidades: adjetivos, substantivos;
4. Relações: verbos, adjetivos.

Procedência Lingüística : os termos podem ser formados usando recursos lingüísticos já existentes, atribuindo um novo significado a uma palavra já existente como, por exemplo: *rato* para *mouse* em informática. Também se encaixa nesse aspecto a formação de neologismos que servem como termos, sendo eles unidades novas ou empréstimos de outras línguas.

3.2.1.2 A dimensão conceitual dos termos

A norma ISO 704 de 1987 estabelece a definição de conceito como “*construções mentais que servem para classificar os objetos individuais do mundo exterior ou interior através de um processo de abstração mais ou menos arbitrário*”. Entretanto, no âmbito da Terminologia a norma ISO DIS 12620 ISO 12620 (1999), que é mais recente, estabelece a definição de conceito como um elemento do pensamento, geralmente rotulado por um termo ou símbolo lexical.

Essa última definição é mais relevante para o estudo terminológico, pois estabelece que os termos são usados para expressar ou denominar elementos do pensamento. Em outras palavras,

o principal papel dos termos é denominar um conceito, fornecendo-lhe um nome para que ele possa ser referenciado nas comunicações.

Por sua vez, as características do conceito são representadas por cada uma das propriedades que os descrevem, de maneira que devem ser suficientes para diferenciar entre um conceito e outro, no mesmo domínio de conhecimento. Dessa forma, as propriedades que definem um conceito e o diferenciam dos demais constituem sua intenção, enquanto os objetos e entidades do mundo real as quais faz referência se denomina de extensão.

Decorre então que a função comunicativa do termo, cumprindo seu papel de transmissor do conhecimento, é dependente de uma correta especificação do(s) conceito(s) com que(m) se relaciona(m), feita, por sua vez, pela definição terminológica. É nesse contexto que destacamos a importância da dimensão conceitual dos termos, já que constitui a ligação entre a dimensão lingüística (denominação) e dimensão comunicativa, encarregada de transmitir o conhecimento inerente ao conceito.

3.2.1.3 A dimensão comunicativa dos termos

Do que foi exposto sobre as dimensões dos termos, fica clara a função da dimensão comunicativa: a de comunicar o conteúdo especializado. Assim, tomando como ponto de partida esta função, as unidades terminológicas têm o propósito eminentemente referencial de denominar a realidade especializada, dentro de uma temática específica, delimitada pelo domínio de especialidade em que estão inseridos.

Em outras palavras, podemos dizer que a dimensão comunicativa das unidades terminológicas se insere no contexto real de seu uso nos textos especializados, e representa a simbiose das dimensões lingüística e conceitual dos termos. É a dimensão comunicativa que concebe o resultado da junção do trabalho terminográfico e o lingüístico na função de produzir obras de referência que transmitem o saber especializado.

3.2.2 O uso de cópulas de especialidade para extração automática de termos

O uso de cópulas textuais eletrônicas nas pesquisas lingüísticas tornou-se comum, devido à proliferação destes recursos informatizados, mostrando na prática a aplicabilidade das ferramentas e a confiabilidade dos resultados obtidos.

O processamento automático ou semi-automático dos textos especializados é um dos ideais da Lingüística de Cópulas. Para concretizá-lo, é preciso combinar os atuais avanços da Terminologia, como disciplina, com os progressos alcançados pela prática do Processamento de Língua Natural (PLN) e, conseqüentemente, da *Lingüística de Cópulas*. Nesse sentido, o desenvolvimento da *Lingüística de Cópulas* no contexto da Terminologia tem tornado viável a utilização de cópulas

de especialidade, principalmente agilizando a identificação e extração de termos em grandes volumes de textos Maciel (2000).

Embora tenha inúmeras aplicações na terminografia como, por exemplo, a criação de bases de dados terminológicas, a elaboração de definições e a edição de verbetes de dicionários, uma grande vantagem de uso da informática é a identificação de candidatos a termos. De fato, se nos apoiarmos em um *cópus* de especialidade confiável, é possível com uso de algumas ferramentas gerar uma lista de palavras que representam um conjunto prévio de itens candidatos a integrar a nomenclatura de um futuro produto terminológico. Entretanto, como bem salienta Krieger e Finatto (2004), a análise de um *cópus* de especialidade para fins de extração de candidatos a termos envolve diferentes estágios e refinamentos de processamento. O trabalho automático com o *cópus* deve vincular a pesquisa terminológica, a teoria lingüística, e a capacidade dos recursos informatizados que fazem buscas de um determinado tipo de palavras em um texto. Dessa maneira, esses recursos devem estar munidos de conhecimentos lingüísticos para se obter resultados mais expressivos.

De posse desses conhecimentos, o reconhecimento terminológico é o segundo passo concreto na elaboração de um produto terminológico. O conhecimento lingüístico, nesse contexto, corresponde principalmente aos aspectos sintáticos e semânticos dos candidatos a termos no contexto de seus registros no *cópus*. Segundo Maciel (2001), as formas nominais (substantivos) têm sido privilegiadas nos estudos e na extração de termos destinados aos repertórios terminológicos, pois o objetivo de denominar conceitos explica essa primazia concedida à categoria nominal. No entanto, apesar desse costume, outras categorias também podem representar termos, principalmente os adjetivos e verbos.

No entanto, mesmo que se tenha o conhecimento lingüístico para melhorar a qualidade das ferramentas de reconhecimento de candidatos a termos, ainda é possível obter resultados insatisfatórios que demandam a intervenção dos terminólogos ou profissionais que fazem as correções necessárias depois da extração. Isso porque, utilizando as palavras de Alves (1999), a delimitação dos termos de uma área de especialidade é problemática devido à dificuldade no estabelecimento das fronteiras entre um segmento frásico livre e um que se tornou, ou está se tornando, um novo termo.

Do ponto de vista sintático-semântico, os critérios de inseparabilidade e de comutação podem, em alguns casos, contribuir para a distinção, ou identificação de fronteiras entre os sintagmas livres e os candidatos a termos. O tamanho dos termos e dos sintagmas livres é um critério sintático para a tarefa de delimitação das unidades lexicais. Outro critério sintático na delimitação dos termos que também merece ênfase é a ausência de artigos entre seus elementos, revelando uma coesão interna do sintagma.

Um critério de caráter semântico refere-se à imprevisibilidade do sintagma do ponto de vista de seu significado. Essa imprevisibilidade vai ocorrer nos casos em que em um dos elementos do sintagma tem uso metafórico. Uma vez que grande parte dos sintagmas em vias de lexicalização possui um significado transparente e explicativo, o critério de frequência de um determinado sintagma deve ser considerado, já que, coordenado com os demais critérios (sintático e semântico), permite avaliar a aceitabilidade desse elemento entre os usuários (especialistas ou não) de uma língua de especialidade.

Dessa forma, para empreender a atividade de localizar e extrair termos em um corpus de especialidade, quando se pretende uma análise abrangente da língua, principalmente com apoio informatizado, necessariamente é preciso ir além da feição peculiar das terminologias e considerar também o valor constitutivo das estruturas e especificidades das diferentes áreas do conhecimento. Há ainda nesse bojo o reconhecimento dos termos compostos, já que estes termos também fazem parte do conjunto terminológico de uma área de especialidade, na qual se destaca nesta tarefa o trabalho de Ha (2004).

Vale salientar que os possíveis erros encontrados durante o processo de extração, principalmente os que usam conhecimento lingüístico, podem fazer referência aos erros em cascata, ou erros advindos de erros de outras ferramentas de cunho lingüístico. Um exemplo claro desse caso é quando se usa um etiquetador morfossintático para etiquetar o corpus antes de se aplicar os extratores. Um erro do etiquetador pode gerar um erro do extrator, sem que este último tenha qualquer responsabilidade.

O estudo dos pressupostos teóricos e práticos acerca da extração de termos se assenta em princípios que colocam em evidência o uso de corpus especializados, de modo que a construção deste recurso torna-se de grande importância. Portanto, ressaltamos que o valor deste recurso lingüístico conjugado com as ferramentas computacionais adequadas à extração de termos tem-se destacado no controle da identificação e caracterização das unidades lexicais, decorrendo daí a atenção especial que deve ser dispensada ao projeto e construção de corpus especializados, bem como da avaliação dos tipos de texto que os constituem Nascimento (2003).

Dado o exposto acima, é direta a conclusão de que a qualidade da extração de candidatos a termos e, conseqüentemente, a do projeto terminológico como um todo, está diretamente relacionada com a qualidade do corpus de trabalho. Assim, tendo em vista o trabalho em terminologia, somente poderá se assegurar que o resultado das análises e a frequência de ocorrência sejam fiáveis baseando-se em um corpus especializado de boa qualidade.

3.2.3 A extração automática de candidatos a termos

Um dos maiores problemas encontrados no trabalho terminológico diz respeito à extração manual dos termos do *cópus* de especialidade. Além de atrasar a pesquisa, o procedimento manual muito dificulta a utilização do critério de *frequência* para a seleção de termos, bem como que sejam facilmente armazenados vários contextos relevantes de cada termo selecionado, inibindo a inclusão desses nos verbetes do produto terminológico final.

Nesse sentido, em relação à extração manual de termos, os principais inconvenientes são: 1) dispêndio de tempo, visto que esta tarefa exige a leitura cuidadosa do *cópus* inteiro; 2) uso do critério semântico para a seleção dos termos, já que a extração manual inviabiliza a utilização do critério de frequência; e 3) impossibilidade de se armazenar contextos relevantes sobre cada termo, o que dificulta o processo, pois em um trabalho terminográfico é de extrema relevância o acesso a contextos para a elaboração da definição.

De acordo com Bagot (2001), o reconhecimento das unidades terminológicas de um texto é uma das tarefas centrais do trabalho de terminologia. Entretanto, devido ao grande volume de textos e informação técnica geralmente disponíveis nos *cópus* de especialidade, a extração manual de termos, além de ineficaz, tornou-se um gargalo para todo o trabalho de Terminologia, pois requer muito tempo e aplicação de critérios sistemáticos. Devido à esta dificuldade, nota-se, no início dos anos oitenta, as primeiras tentativas de se extrair automaticamente unidades terminológicas dos textos especializados, buscando automatizar, ou pelo menos semi-automatizar, essa tarefa do trabalho terminográfico. Nesse sentido, a criação de *cópus* textuais de especialidade, no final da década de oitenta e no decorrer dos anos noventa, fez com que os primeiros programas de extração automática de terminologia começassem a apresentar resultados positivos.

Um extrator automático de termos pode ser definido como um conjunto de programas dedicados ao reconhecimento e extração de termos que aparecem em um *cópus* de especialidade. Esses programas, por sua vez, têm grande notoriedade para as aplicações de PLN que trabalham com domínios especializados Bagot (2001).

Especificamente para a pesquisa terminológica, a extração automática de candidatos a termos é de fundamental importância para sua eficácia, pois permite a utilização de mais de um critério para a seleção dos termos que se tornarão verbetes do dicionário, tais como o critério *frequência* e as várias medidas estatísticas de associação de palavras, como pode ser visto em Daille (1996). Além disso, a extração automática permite que sejam armazenados vários contextos relevantes sobre cada termo, possibilitado a inserção desses contextos nas fichas terminológicas e nos verbetes (conforme será estudado posteriormente nesse capítulo).

No universo do discurso dos textos especializados, os termos são inicialmente classificados com um dos dois tipos: 1) os **termos de autor** e 2) os **termos de dicionários**. Os **termos de**

autor correspondem àqueles que são explicitamente ou implicitamente definidos pelos autores e usados em seus textos. É importante ressaltar que esses termos devem abandonar seu escopo local de aplicabilidade e, conseqüentemente, ao serem listados pelos extratores automáticos podem passar a compor os dicionários. Os **termos de dicionários** são aqueles que já deixaram o status de termos de autor e já pertencem ao universo dicionarístico.

Empiricamente, existe uma fronteira tênue entre os termos de autor e termos de dicionários. Inicialmente, cada termo é definido como termo de autor e, à medida que vai sendo utilizado em vários textos (frequência de uso), torna-se termo de dicionário Teline (2004). Segundo Bolshakova *apud* Teline (2004) há três formas de identificar um termo: 1) o termo é explicitamente definido; 2) o termo é indefinido, estando sua definição ausente, mas visualmente exposto; 3) o termo não é nem definido nem exposto, sendo então escondido.

Portanto, essas três formas de identificação devem ser consideradas pelos métodos de extração automática. No entanto, existe uma grande dificuldade de identificação dos termos indefinidos, já que os extratores geralmente utilizam padrões morfológicos e morfossintáticos para reconhecer as unidades terminológicas. Assim, se forem utilizadas padrões referentes somente à forma da unidade, a maioria dos candidatos a termos serão errados. Por esse motivo, os extratores também devem considerar conhecimentos lingüísticos para delimitar automaticamente as unidades terminológicas de maneira mais precisa Teline (2004).

Na Terminografia, a área de investigação que mais recebeu atenção foi a extração automática de candidatos a termos Hernández (2002). Segundo Bagot (1999), conforme a metodologia utilizada para extração de termos, os sistemas extratores de candidatos a termos podem se dividir em três categorias: 1) *Sistemas baseados em conhecimento estatístico*; 2) *Sistemas baseados em conhecimento lingüístico* e 3) *Sistemas híbridos*, baseados em conhecimento estatístico e lingüístico.

A função de todos os sistemas, independentes da metodologia utilizada, é de identificar e extrair os candidatos a termos. Os de conhecimento estatístico identificam os termos apoiados na frequência de ocorrência dos mesmos no corpus de especialidade. A principal desvantagem desses sistemas é a dependência do tamanho do corpus que utilizam, pois podem gerar **silêncio** ou **ruído**. O **silêncio** corresponde ao falso-negativo e consiste no número de termos não encontrados do total de termos válidos de um corpus. O **ruído**, por sua vez, corresponde ao falso-positivo e consiste no número de candidatos a termos selecionados em um corpus que não correspondem a um termo válido. Por outro lado, a independência de língua é a principal vantagem destes sistemas.

Já os *sistemas de conhecimento lingüístico* identificam os termos a partir das informações lingüísticas pertencentes ao texto. Essas informações dizem respeito a informações morfossintáticas

e às vezes, dependendo da análise metodológica realizada, semântica, como foi experimentado no trabalho de Junior (2008). A desvantagem destes sistemas fica por conta da dependência da língua em que o produto terminológico será gerado, já que envolve conhecimento lingüístico particular de cada língua. Como vantagem, podemos citar a independência do tamanho do corpus visto que análise realizada é sobre as estruturas do texto. Entretanto, vale frisar que mesmo no conhecimento lingüístico a representatividade do corpus é uma variável importante.

Os *sistemas híbridos* utilizam o conhecimento estatístico juntamente com o lingüístico para tornar a extração de termos mais eficiente, dado que sua abordagem pode combinar os dois resultados. Há dois tipos de sistemas híbridos: 1) aqueles que aplicam o conhecimento estatístico primeiro e depois o lingüístico e 2) os que usam a estatística como complemento do lingüístico. No primeiro caso, os problemas de **silêncio** também aparecem, assim como nos sistemas puramente estatísticos. No segundo caso, os resultados podem ser melhorados pelo uso da estatística para negar ou reafirmar a existência de um termo, conforme dito por Bagot (1999).

Nesse contexto, o projeto *ExPorTer*²⁴ Teline (2004) reúne um conjunto de informações significativas a respeito das vantagens e desvantagens de vários métodos de extração automática de termos, segundo as abordagens estatística, lingüística e híbrida, avaliados com um corpus de textos escritos em Português do Brasil da área de Revestimentos Cerâmicos. Além disso, este trabalho apresenta uma avaliação inicial de tais métodos, favorecendo futuras melhorias dos mesmos. A seguir faremos uma breve descrição destes métodos.

3.2.3.1 O método estatístico

Os sistemas estatísticos de extração de candidatos a termos consistem naqueles que utilizam dados referentes às frequências de ocorrência das unidades lexicais de um corpus para selecionar o conjunto de termos candidatos. Em sua grande maioria, não exigem nenhuma preparação ou tratamento prévio do corpus e são relativamente simples de aplicar.

Devido à sua característica estatística, dada a possibilidade de realizar cálculos com as frequências das palavras do corpus, esses sistemas utilizam várias medidas estatísticas de associação de palavras para melhorar seu desempenho. Tais medidas, em sua maioria, tentam quantificar o grau de proximidade entre duas ou mais unidades léxicas no texto na intenção de identificar valores que representam a possibilidade de termos candidatos poliléxicos.

Para isso, utilizam desde a *frequência simples* Daille (1996) até as estatísticas mais complexas, como *Informação Mútua* e *Coeficiente Dice*, *Loglikelihood* conforme descrito em Pantel & Lin *apud* Teline (2004), e *C-value*. Uma breve descrição dos cálculos e interpretação dessas medidas

²⁴ExPorTer significa *Evaluation of Automatic Terminology Extraction Methods for Portuguese Text* e foi desenvolvido no NILC. Mais informações no endereço: <http://www.nilc.icmc.usp.br/nilc/projects/termextract.htm>

pode ser vista em Teline (2004) e Hernández (2002).

Além disso, um recurso muito utilizado pelos extratores estatísticos é a *stoplist/ stopwords*, que consiste em uma lista de palavras ou colocações da língua geral que não possuem valor terminológico e que são excluídas no momento da identificação automática dos candidatos a termos. Dessa forma, o uso da *stoplist* evita que “palavras” da língua geral, sejam consideradas na lista de candidatos a termos.

Os principais problemas dos sistemas estatísticos é a ocorrência dos erros de *silêncio* e *ruído*. Nesse contexto, calcular o impacto do *silêncio* no resultado final da extração é uma tarefa difícil, pois este erro fica “contido” no *cópus* (já que representa os termos não recuperados) e é necessário analisá-lo minuciosamente para descobrir quais termos que não foram extraídos. Diferentemente do *silêncio*, o *ruído* está “contido” na lista de candidatos a termos e é, naturalmente mais palpável. De acordo com Bagot (2001), os extratores baseados em conhecimento lingüístico geralmente possuem altas porcentagens de *ruído*, podendo atingir até 75%. A Figura 3.5 mostra a localização dos erros de *silêncio* e *ruído* no processo automático de extração de termos.

Figura 3.5: Localização dos erros de *silêncio* e *ruído* no processo de extração automática de termos.

Como exemplo de extrator estatístico podemos citar o Pacote NSP (*N-gram Statistics Package*)²⁵, escrito em Perl. O Pacote NSP foi implementado por Ted Pedersen juntamente com uma equipe de pesquisadores da *Universidade de Minnesota*. O NSP é formado por um conjunto de programas que auxilia na análise de *n-grams* em arquivos texto.

Esse pacote é encontrado em diversas versões, sendo que a mais atual é a 1.09. Entre os programas do pacote, há dois programas principais (*Count* e *Statistic*) que são utilizados para identificação e cálculo de medidas estatísticas. Essa versão proporciona dezenas de medidas de associação para utilização na análise e extração de candidatos a termos.

3.2.3.2 O método lingüístico

Os sistemas de conhecimento lingüístico identificam os termos a partir das informações lingüísticas pertencentes ao texto. Estas informações dizem respeito a informações morfológicas,

²⁵Para mais informações e *download* do Pacote NSP acesse o endereço: <http://www.d.umn.edu/~tped-erse/nsp.html>

morfossintáticas, sintáticas, semânticas e pragmáticas e podem ser:

- Lexicográficas: dicionários de termos e lista de palavras auxiliares (*stoplist*);
- Morfológicas: padrões de estrutura interna da palavra;
- Morfossintáticas e Sintáticas: categorias morfossintáticas e sintagmas nominais;
- Semânticas: classificações semânticas;
- Pragmáticas: representações tipográficas e informações de disposição do termo no texto.

Esse tipo de conhecimento, quando empregado nos sistemas lingüísticos, faz com que estes sejam aplicados somente em uma língua e às vezes até mesmo a uma única variante, pois a sua utilização em textos em uma língua diferente exige um estudo lingüístico prévio.

De acordo com Bagot (1999), a grande quantidade de **ruído** gerada (entre 55% e 75%) é um dos problemas principais dos sistemas que trabalham apenas dados morfológicos, morfossintáticos, sintáticos e/ou léxicos. Nem todas as palavras que são consideradas pelo sistema como unidades terminológicas polilexicais o são, já que a maioria dos mesmos padrões corresponde também a unidades léxicas e fraseológicas que não apresentam uso especializado. Em alguns casos elas correspondem a unidades de uso especializado, mas não são terminológicas, como as unidades fraseológicas especializadas ou as combinações que apresentam muitas recorrências (colocações); e em outras vezes tais padrões são expressões discursivas (como por exemplo: “o objetivo deste trabalho”, “nesta seção”, etc.), sem caráter especializado. Por esta razão, os extratores lingüísticos também fazem uso de uma *stoplist*.

Na literatura de terminografia, há vários exemplos de extratores que utilizam conhecimentos lingüísticos. Podemos citar o extrator DEFINDER Klavans e Muresan (2000), Klavans e Muresan (2001), que consiste na combinação de técnicas de PLN de análise gramatical, com o intuito de explorar textos da área médica para identificar expressões linguísticas típicas e indicadores estruturais que introduzem definições dos termos. Também há o TERMS, citado em Terminometro (1994), que se baseia na formação morfossintática e sintática das unidades lexicais do discurso especializado. Outro importante sistema que utiliza conhecimento lingüístico para extração de candidatos a termos é o LEXTER também citado em Terminometro (1994). Este sistema baseia-se exclusivamente em técnicas linguísticas que representam o conhecimento negativo de uma unidade léxica, ou seja, na identificação dos elementos que não podem formar um termo ou parte de um termo.

Ainda há outros sistemas que utilizam o conhecimento lingüístico para extrair candidatos a termos nos corpúsculos de especialidade, por exemplo, o NODALIDA, FASTR e NAULLEAU. Uma

análise mais aprofundada destes sistemas pode ser encontrada em Bagot (1999) e Terminometro (1994).

3.2.3.3 O método híbrido

Por último, existem os sistemas híbridos que utilizam o conhecimento estatístico juntamente com o lingüístico. A aplicação do conhecimento híbrido torna o sistema mais eficiente, visto que ele condiciona os resultados.

Existem dois tipos de métodos híbridos: aqueles que aplicam o conhecimento estatístico primeiro e depois o lingüístico, e aqueles que utilizam a estatística apenas como um complemento da lingüística. No primeiro caso, acontecem os mesmos problemas de **silêncio** encontrados nos sistemas puramente estatísticos. Já no segundo, os resultados finais podem ser melhores em razão de a estatística auxiliar no momento do processo de detecção, reafirmando ou recusando a condição de termo de uma unidade lingüística Bagot (2001), Daille (1996), Frantzy (1997), Ananiadou (1994).

Um exemplo de extrator híbrido que podemos citar é o ACABIT resultado da tese de doutorado de Daille *apud* Teline (2004) e Bagot (1999). A idéia básica deste trabalho é combinar o conhecimento lingüístico com cálculos estatísticos. No sistema, utilizou-se um *cópus* de meio milhão de palavras da área de telecomunicações etiquetado morfossintaticamente, do qual criou-se uma lista de candidatos a termos de acordo com padrões de formação de termos informados previamente que, por sua vez, foi submetida a métodos estatísticos para filtrar a primeira seleção.

Outros exemplos de aplicação de critérios lingüísticos e estatísticos nas tarefas de extração de candidatos a termos são os trabalhos de Frantzy e Ananiadou (1997), que discutem o reconhecimento de termos utilizando informações de contexto; os mesmos aparecem e Dias (2000), que estuda a combinação de conhecimento lingüístico com dados estatísticos para a extração de termos compostos.

Destacamos aqui a utilização do Pacote NSP (visto na Seção 3.2.3.1) também para o método híbrido, principalmente na parte estatística. Isso porque, uma vez aplicados os critérios lingüísticos para identificação dos candidatos a termos, as atribuições deste pacote são muito úteis para a análise estatística. O uso do NSP na abordagem híbrida mostrou-se proveitosa na extração de candidatos a termos, em relação aos métodos estatístico e lingüístico, como pode ser verificado no trabalho de Teline (2004).

3.2.4 Padrões de Formação dos Termos

Em Terminologia os termos também são vistos como componentes lexicais ordinários, os quais foram adaptados a um sentido técnico ou científico através de um processo de terminologização a partir da necessidade de um desenvolvimento mais profundo em um determinado domínio do conhecimento. No entanto, é preciso observar que esta necessidade surge somente após o aparecimento de uma idéia, conceito, objeto ou processo, pois não se denomina o que não existe.

Conforme inicialmente dito na Seção 3.2.1.1, os termos, no entanto, diferentemente da língua geral, caracterizam-se por apresentarem predominantemente formações de sintagmas em que um, dois ou mais elementos lexicais constituem uma unidade composta e correspondem a um conceito, referente a um campo particular do conhecimento.

Essas unidades compostas são representadas por distintas estruturas formais. A estrutura de um termo é composta por n palavras, ou *n-grams*, onde $n = 1, 2, 3 \dots m$. Considerando a variabilidade da formação dos termos, as estruturas mais comuns possuem $n \leq 3$, embora termos com valor de n alto também sejam considerados Alves (1999), Ananiadou (1994). Empiricamente, termos com $n = 1$ são chamados de *unigramas*, $n = 2$ de *bigramas* e com $n = 3$ de *trigramas*. Entretanto, analisar a formação dos termos apenas como uma composição de *n-grams* de tamanho variado é insuficiente, visto que suas estruturas formais sintagmáticas devem ser também consideradas. Assim, acatando esta idéia e com base nos estudos de Alves (1999), a estrutura formal mais freqüente é constituída por um substantivo genérico, que é especificado por adjetivo determinante (bigrama):

substantivo + adjetivo

Outro tipo de estrutura, também bastante freqüente, é a formada por um substantivo determinado seguido de uma preposição e de outro substantivo ou um substantivo seguido por dois adjetivos (trigramas):

substantivo + preposição + substantivo

substantivo + adjetivo + adjetivo

Outras possibilidades de formação de *tetragramas* são ainda observadas entre as formações sintagmáticas:

substantivo + adjetivo + preposição + substantivo

substantivo + preposição + substantivo + adjetivo

Essas formações resultam de um processo de lexicalização de segmentos frásicos pertencentes aos textos da área de especialidade em questão, em que o conceito usualmente precede a criação do termo correspondente.

Os sintagmas lexicalizados ou as unidades terminológicas são, com muita frequência, sujeitas a uma expansão (à esquerda ou à direita) de caráter lexical que, na verdade, coincide com a denominação correspondente a uma especialização de um termo genérico. De acordo com Alves (1999), exemplos de expansão à direita de um termo genérico são bastante comuns nas línguas de especialidade e correspondem a um novo conceito, ou uma nova tecnologia que determina a especialização do conceito expresso pelo termo genérico (anterior à expansão). A expansão à esquerda é também observada nas línguas de especialidades, mesmo que com frequência menor. Os termos são, geralmente, de tamanho limitado e são formados por dois (bigramas) ou três (trigramas) elementos.

3.2.5 Problemas relacionados com a extração automática de termos

Observando os extratores automáticos de um ponto de vista prático, sua função primordial é extrair candidatos a termos de um corpus de especialidade. Uma vez que a lista de candidatos a termos tenha sido produzida, é necessário avaliar a sua qualidade para verificar se realmente possuem valor terminológico. Um dos gargalos da tarefa de extração automática é a avaliação da lista de candidatos a termos extraída, pois exige a opinião de especialistas, sendo esse processo caro e demorado. Por outro lado, contar com recursos como glossários ou dicionários, isto é, com listas de referências, também traz seus riscos, uma vez que tais recursos são incompletos, dada a constante produção de novos termos Zavaglia et al. (2004). Isso porque a principal dificuldade da tarefa de extração automática é o reconhecimento terminológico. Ou seja, julgar se uma determinada unidade lexical, independente do seu tamanho, possui ou não valor terminológico, em um determinado contexto, é uma tarefa difícil, mesmo usando conhecimentos lingüísticos e as mais variadas medidas estatísticas. Em outras palavras, o desafio dos extratores automáticos é diminuir ao máximo tanto o **silêncio** quanto o **ruído**.

Acreditamos que para reduzir os valores de **silêncio** e **ruído** no contexto da extração automática de candidatos a termos é necessário investir em pesquisa e estudo que consideram tanto os aspectos lingüísticos de ocorrência dos termos nas línguas de especialidade, quanto os aspectos computacionais. Aos aspectos lingüísticos interessam estudos sobre: 1) as categorias gramaticais suscetíveis ao valor terminológico em diferentes áreas de especialidade, 2) as relações semânticas entre os constituintes das unidades terminológicas, bem como a própria relação semântica dos termos, 3) as possíveis características pragmáticas que podem ser associadas à identificação dos termos, 4) a disposição dos termos no corpus e individualmente nos textos e 5) o estudo das especificidades peculiares das línguas de especialidade.

Os aspectos computacionais, por sua vez, devem dar conta da automatização dos aspectos e conhecimentos lingüísticos acima mencionados, bem como interpretar empiricamente os resulta-

dos das medidas de associação, combinar tais medidas com intuito comparativo e de julgamento entre as medidas e, ainda, melhorar as interfaces para favorecer a interação com os usuários.

Embora os resultados dos extratores pareçam, em uma primeira análise, insatisfatórios, eles revelam que a aplicação de métodos automáticos para a identificação e captura de termos, após serem refinados e especializados para o domínio tratado, assim como para o tamanho de córpus de especialidade em uso, poderá auxiliar de maneira eficaz o trabalho de extração do terminólogo, dado que são capazes de identificar diversos itens lexicais que são considerados efetivamente termos de forma bastante rápida. Em uma extração de termos humana com base no critério semântico, essa captura de unidades lexicais demandaria tempo e seria certamente muito mais lenta e subjetiva, se comparada à feita pelo computador.

3.3 Etapa 03 - Criação da Estrutura Conceitual do Domínio (Ontologia)

Os resultados produzidos pelos extratores são as listas de candidatos a termos que, em uma tarefa posterior à extração, são julgados quanto ao seu valor terminológico. Ou seja, cada termo candidato é avaliado a respeito do seu valor terminológico e, conseqüentemente, sobre seu rótulo de termo ou não-termo.

Para auxiliar na tarefa de avaliação e julgamento dos termos candidatos, é comum inseri-los em uma estrutura conceitual da área de especialidade (como será visto na Seção 3.4). A estrutura conceitual do domínio, ou comumente chamada de árvore de domínio Krieger e Finatto (2004), Pavel e Nolet (2002), é um arranjo hierárquico dos conceitos de uma área do conhecimento. Este tipo de arranjo, ou esquema, serve para organizar os campos nocionais de uma especialidade ou ciência, de modo que o terminólogo e o especialista possam, baseado nele, compreender as hierarquias básicas e também situar um recorte do conhecimento objeto do trabalho terminológico.

Por sua vez, uma **ontologia** é uma organização semântica da área-objeto, semelhante ao que se entende por árvore de domínio, com a diferença que os conceitos/termos estão ali armazenados como instâncias e pode haver explicitamente a especificação das relações conceituais entre eles. Dessa forma, organiza-se uma estrutura constituída de campos nocionais, de forma que essa estrutura reflita os conceitos da área-objeto juntamente com os termos, bem como as suas relações. A ontologia deve ser elaborada pelos terminólogos com assessoria dos especialistas.

Conforme mencionam Gillam et al. (2005), Novello (2002), Feliu et al. (2004) a palavra ontologia foi herdada da filosofia, significando uma explicação sistemática da existência das entidades de um determinado universo.

Segundo Novello (2002), baseadas na definição de Gruber (1993), muitas definições de ontologias foram propostas na literatura. Exemplificando algumas dessas definições, podemos

dizer que uma ontologia é uma hierarquia estruturada de um conjunto de conceitos e termos para descrever um determinado domínio, podendo ser usada como estrutura de conhecimento, fornecendo meios para descrever explicitamente as relações dos conceitos de uma área.

Uma ontologia corresponde, portanto, a uma representação de um domínio a partir de seus conceitos abstratos e a forma como esses conceitos se relacionam entre si. É um modelo consensual de um mundo particular, no sentido de que é reconhecido da mesma forma pelas pessoas desse mundo.

3.3.1 Ontologias e a terminologia

O uso de ontologias visa capturar o conhecimento declarativo do domínio e fornecer uma compreensão deste, possibilitando o reuso e o compartilhamento. Muitos trabalhos na área de inteligência artificial vêm explorando o uso de ontologias como uma maneira formal para especificar informações conceituais de um domínio, compartilhar e reusar conhecimento. Em Terminologia, as ontologias fornecem uma estrutura formal que representa em parte o vocabulário comum de uma área e ajuda a definir o significado dos termos e dos relacionamentos entre eles. Uma ontologia de um domínio é determinada através de conceitos (classes ou campos nocionais), suas relações e valores (instâncias).

Nesse contexto, a ontologia na pesquisa terminológica é fundamental para: 1) possibilitar um mapeamento mais sistemático de um campo de especialidade; 2) circunscrever a pesquisa, já que todas as ramificações da área-objeto, com seus campos, são previamente mapeadas; 3) delimitar o conjunto terminológico; 4) determinar a pertinência dos termos, pois separando cada grupo de termos pertencentes a um determinado campo, poder-se-á apontar quais termos são relevantes para o trabalho e quais não são; 5) prever os grupos de termos pertencentes à área-objeto, como também os que fazem parte de matérias conexas; 6) definir as unidades terminológicas de maneira sistemática e, finalmente, 7) controlar a rede de remissivas Almeida (2000).

Na teoria terminológica os conceitos têm de ser organizados de forma sistemática e caracterizados de acordo com as relações que estabelecem uns com os outros no seio de um sistema conceitual. Tradicionalmente, tais relações conceituais podem ser de quatro tipos Hernández (2002):

Relações genérico-específico : relação hierárquica na qual se identificam os conceitos genéricos e específicos. Um conceito genérico pode ser chamado de superordenado de outros conceitos mais específicos. Os conceitos subordinados, embora compartilhem características do conceito genérico, possuem particularidades próprias que os diferenciam e os fazem mais específicos. Essa relação conceitual alcança dois sentidos diferentes: *vertical*, a que se estabelece entre um conceito específico (subordinado) e seu conceito genérico (superor-

denado) e *horizontal*, a que sustentam vários conceitos específicos que possuem o mesmo conceito genérico e que se diferenciam entre si por possuir alguma característica diferente. Também conhecida pela relação verbal “**É um**” (“*Is a*”, *em inglês*).

Relações parte-todo : é a relação existente entre conceitos que são formados por mais de uma parte, chamada de partes constituintes. Também conhecida pela relação verbal “**Parte de**” (“*Part of*”, *em inglês*).

Relações polivalentes : relação que representa a possibilidade de um conceito se localizar em posições (ramos) diferentes em uma mesma ontologia. Em Terminologia tal relação também pode ser identificada como polissemia, quando um determinado termo se liga a dois conceitos em posições diferentes.

Relações complexas : englobam uma série de inter-relações que se estabelecem entre conceitos em uma hierarquia, mas que não podem se considerar nem com genérico-específico nem parte-todo. Exemplos desse tipo de relações são as de causa-efeito e de processo-produto.

Dos quatro tipos de relações mencionados, as mais estudadas e usadas na prática terminográfica foram relações genérico-específico e parte-todo Hernández (2002). Isso porque são as relações mais fáceis de serem representadas em estruturas arbóreas. No entanto, principalmente com o desenvolvimento da TCT, as representações de ontologias em formatos de árvores têm sido insuficientes, já que, essa estrutura dificulta a representação das relações polivalentes e complexas entre conceitos de uma mesma estrutura conceitual.

Nesse sentido, uma ontologia é precisamente o tipo de recurso que permite uma concreta realização conceitual de um domínio e deve ser feita de forma explícita e detalhada. Na prática terminológica, a criação de uma ontologia define algumas restrições sobre a metodologia de trabalho, assegurando, por exemplo, que todos os candidatos a termos estejam atribuídos a um determinado conceito.

Cada um dos conceitos, por sua vez, deve compor uma estrutura bem definida, possibilitando a especificação de diversos tipos de relações entre eles, que ajudem a descrever formalmente o conceito específico no qual os termos se encaixam. Além disso, conforme menciona Cabré *apud* Hernández (2002) o uso de uma ontologia exige que o terminólogo represente explicitamente a perspectiva do trabalho terminográfico, já que esta estará refletida na estrutura da ontologia.

Ainda nesse contexto, o aspecto fundamental de uso de ontologias no âmbito da Terminologia se refere à possibilidade de que um conceito possa se localizar em mais de uma classe de conceitos, principalmente sob os pressupostos da TCT. Isso é possível graças aos mecanismos de herança, já que uma das vantagens das ontologias de conceitos é a possibilidade de herança múltipla,

por meio da qual um conceito filho pode se ligar a mais de um pai e aparecer, desse modo, em lugares diferentes da ontologia Feliu et al. (2004).

É por esse motivo que atualmente há uma tendência em se utilizar estrutura de grafos direcionados, ao invés de árvores, para representar ontologias. Isso é justificado porque o uso de grafos permite a ocorrência das relações polivalentes e complexas entre os conceitos, concebida pelos aspectos polissêmicos presentes no domínio especializado. Além disso, o uso de grafos para estruturar as ontologias no contexto terminológico admite a complexidade inerente em cada área de especialidade.

3.3.2 Linguagens e formatos de representação de ontologias

A maneira mais simples de representar ontologias é utilizar o formato texto puro, em um arquivo com extensão TXT, legíveis em qualquer tipo de programa ou arquiteturas computacionais. Entretanto, apesar da vantagem da facilidade de leitura, o formato texto é muito limitado e permite, basicamente, a representação das relações conceituais genérico-específico, não permitindo a representação de outras relações. Mesmo que se crie um formato particular para representar tais relações, este formato será utilizado em contextos reduzidos de um único trabalho, não interessantes para os aspectos de intercâmbio que permeiam a terminologia.

Dada essa limitação, diversas linguagens ou formatos de representação vêm sendo utilizados para formalizar ontologias. Os formatos mais representativos são *Ontolínqua*, RDF, OIL e OWL. *Ontolínqua* é uma linguagem baseada em KIF (*Knowledge Interchange Format*) e no *Frame Ontology*, e permite que as ontologias sejam construídas, editadas e salvas em grandes bases de conhecimento em uma linguagem formal de representação, oferecendo algum grau de consistência do conhecimento. A *Ontolínqua* foi implementada no sistema *Ontology Server*²⁶ Farquhar et al. (1996), Farquhar et al. (1997) cuja função é favorecer a criação de ontologias por meio do compartilhamento dos exemplos de ontologias de várias áreas, a maioria delas desenvolvidas em trabalhos e escopos particulares.

O *World Wide Web Consortium*²⁷ (W3C) desenvolveu o XML (*Extensible Markup Language*) W3C (2005) que permite o armazenamento de informações usando etiquetas (tags) de marcação. Desde sua criação em 1998, o XML tem sido utilizado em diversos projetos de pesquisa e desenvolvimento, principalmente naqueles que requerem uma massiva necessidade de intercâmbio e troca de dados. No entanto, apesar do XML possibilitar a marcação de metadados e caracterização dos atributos de um dado, ele tem capacidade limitada de descrever relações conceituais, principalmente na representação de ontologias Feliu et al. (2004).

Por este motivo, a W3C concentrou esforços no desenvolvimento de um outro formato,

²⁶Mais informações podem ser encontradas no endereço: <http://www-ksl-svc.stanford.edu>

²⁷Mais informações sobre o consórcio W3C acesse o site: www.w3c.org.

baseado em XML, mas que pudesse representar tais relações, que foi denominado de RDF (*Resource Description Framework*). RDF então é definido com uma linguagem de marcação declarativa que fornece um formato padrão para representar informação de recursos da Internet (Web). Esse formato foi recomendado pela W3C e tem como principal objetivo criar um modelo simples de dados, com uma semântica formal baseada em uma sintaxe XML, que represente as propriedades dos dados armazenados e suas relações.

Novamente, em fevereiro de 2004, a W3C lançou o RDF-S²⁸ (*Resource Description Framework Schema*) como sendo uma linguagem para representar a estrutura hierárquica das ontologias, principalmente no contexto da Web Semântica. O RDF-S define propriedades, classes e valores de recursos contendo inclusive mecanismos para descrever grupos de recursos assim como as relações entre eles Pereira e Baptista (2004).

Por sua vez, de acordo com Pereira e Baptista (2004), a linguagem OIL possui dois significados distintos, sendo eles: *Ontology Inference Layer* ou *Ontology Interchange Language*. Independente do significado, a OIL é particularmente direcionada para a criação e edição de ontologias. Utilizando a sintaxe do RDF, essa linguagem explora o modelo de marcação do RDF-S, permitindo tratar as ontologias como extensões RDF, podendo assim disponibilizá-las em diversas aplicações.

Outra linguagem de grande importância que podemos citar no contexto de representação de ontologias é a OWL (*Web Ontology Language*). A OWL é uma linguagem semântica de anotação, criada com o objetivo de publicar e partilhar ontologias na Internet. Foi desenvolvida como uma extensão da RDF, proveniente da linguagem de ontologias OIL e tem a capacidade de descrever classes e relações entre documentos e aplicações Web.

O trabalho de Almeida (2003b) faz uma compilação de linguagens, formatos, ferramentas e metodologias para construção de ontologias. A Tabela 3.1 mostra uma lista dessas linguagens com uma breve descrição de cada um delas.

3.3.2.1 Ontologias e modelos de banco de dados relacionais

Uma maneira diferente de representar ontologias, dependendo do contexto em que serão utilizadas, é utilizar banco de dados relacionais. Isso porque os esquemas de banco de dados relacionais servem como ontologias por meio da especificação das relações que podem existir entre os dados e as restrições de integridade que devem ser manipuladas pelo banco.

Para a representação da ontologia em banco de dados, é necessário esclarecer a seguinte abstração, em nível conceitual: o domínio do conhecimento é construído por dois componentes – as *estruturas do domínio* e as *bases de conhecimento*. As *estruturas do domínio* descrevem

²⁸Mais informações sobre a linguagem RDF-S visite: http://www.w3schools.com/rdf/rdf_schema.asp

Tabela 3.1: Linguagens para representação de ontologias.

CycL http://www.cyc.com/	Linguagem formal cuja sintaxe é derivada da lógica de primeira ordem. Expressa o conhecimento por meio de um conjunto de constantes semânticas, variáveis, números e/ou seqüências de caracteres os quais são combinados em expressões, sentenças e, finalmente, em bases de conhecimento.
F-Logic Kifer et al. (1995) (Frame Logic)	Consiste numa linguagem que integra <i>frames</i> , lógica de primeira ordem e orientação a objetos permitindo a representação de conceitos, taxonomias, relações binárias, funções, instâncias, axiomas e regras
LOOM http://www.isi.edu/isd/LOOM/LOOM-HOME.html	Baseado em lógica descritiva e regras de produção. Permite a representação de conceitos, taxonomias, relações n-árias, funções, axiomas e regras de produção.
CARIN Levy e Rousset (1996)	Trata-se de uma combinação da <i>Datalog</i> (linguagem baseada em regras) lógica descritiva. Uma ontologia CARIN é construída por dois componentes terminológicos: um conjunto de conceitos e um conjunto de regras que usam os conceitos.
OCML http://kmi.open.ac.uk/projects/ocml/	Linguagem que permite a especificação de funções, relações e classes, instâncias e regras. Muito utilizada em aplicações de gerenciamento do conhecimento, desenvolvimento de ontologias, comércio eletrônico e sistemas baseados em conhecimento.
OML (Ontology Markup Language) http://www.ontologos.org/OML/.%COML%5COML%200.2.html	É uma linguagem de especificação de ontologias baseada em lógica descritiva e grafos conceituais. Permite a representação de conceitos organizados em taxonomias, relações e axiomas em lógica de primeira ordem.
NKRL Zarri (1997)	Linguagem especificamente desenvolvida para descrever modelos semânticos para documentos multimídia. Baseada em <i>frames</i> , permite definir conceitos organizados em taxonomias e seus relacionamentos.
SHOE (Simple HTMLOntology Extensions) http://www.cs.umd.edu/projects/plus/SHOE/	Utiliza extensões adicionadas ao HTML, com marcações para inserir dados semânticos em páginas Web. As marcações podem ser: para a construção de ontologias e para anotações em documentos da Web, permitindo representar conceitos, taxonomias, relações, instâncias e regras.
XOL http://www.ai.sri.com/pkarp/xol/	Consiste numa linguagem restrita, na qual apenas conceitos, taxonomias e relações binárias podem ser especificadas.
DAML + OIL http://www.daml.org/	Linguagem de marcação da Web Semântica que apresenta extensões a linguagens como o DAML (<i>DARPA Agent Markup Language</i>), RDF e RDFS, permitindo representar conceitos, taxonomias, relações binárias e instâncias.

de forma esquemática os tipos abstratos do domínio e como eles se relacionam através das hierarquias. As *bases de conhecimento*, por sua vez, contêm as instâncias do conhecimento desses tipos abstratos, ou seja, são os elementos do conhecimento encaixados na hierarquia.

Do exposto é possível implementar uma ontologia em modelos de bancos de dados de forma que as *estruturas do domínio* são mapeadas para o nível conceitual de definição do banco (tabelas, atributos, relacionamentos, restrições de integridade, *views*, etc.), enquanto que a *base de conhecimento* é mapeada pelos dados armazenados no banco, estando sujeitos às regras impostas pelo modelo conceitual.

Assim, segundo Novello (2002), existe uma similaridade entre ontologia e modelo conceitual de banco de dados relacionais por dois motivos: 1) é possível fazer uso da hierarquia da ontologia para a representação do modelo de dados, e 2) as instâncias são concebidas pelos dados propriamente ditos, armazenados no banco.

3.3.3 Metodologia para construção de ontologias

Do exposto até aqui, podemos dizer que a existência de ontologias pode ser um fator determinante para a organização da informação e do conhecimento de um domínio. Entretanto, apesar da existência de vários formatos e linguagens de representação das ontologias, sua construção ainda é uma tarefa difícil, exigindo dos especialistas ou terminólogos um trabalho dedicado.

De acordo com Almeida (2003b), a construção de ontologias tem sido estudada por diversos autores na literatura. Do mesmo modo, as técnicas e metodologias para construção de ontologias têm se desenvolvido por iniciativas de especialistas de diversas áreas como, por exemplo, a de inteligência artificial (no desenvolvimento de sistemas baseados em conhecimento), a de PLN (como a lexicologia, tradução e terminologia), a de recuperação da informação e da ciência da informação.

Segundo Gillam e Tariq (2004), a tarefa de criar uma ontologia consiste basicamente em identificar a área do conhecimento que se deseja atacar, levantar os conceitos dessa área, associar esses conceitos com outros e demarcá-los com palavras-chave que representam o sentido nocional de cada um.

Especificamente no âmbito do trabalho terminológico, destacam-se os trabalhos de Gillam e Tariq (2004), Gillam et al. (2005), que utilizam a extração automática de candidatos a termos para a construção de ontologias. Para produzir ontologias, estes autores utilizam técnicas estatísticas para extração de termos candidatos a partir de *corpus*, incluindo para esta tarefa considerações sobre as estatísticas de colocação e agrupamento dos termos (*term clustering*) Bourigault e Jacquemin (1999), produzindo uma lista de termos candidatos hierarquizada. Utilizando este resultado preliminar, técnicas linguísticas de refinamento são aplicadas para in-

crementar a hierarquia, seguindo esta atividade até uma consolidação inicial da ontologia. Esta estrutura primária de termos hierarquizados é organizada de acordo com dois padrões internacionais da ISO, especificamente a ISO 12620 (1999) que trata das categorias de dados e a ISO 16642 que define o formato de marcação de dados terminológicos. Uma vez tendo organizado os dados extraídos no formato da ISO, esta estrutura é utilizada como base para o trabalho terminológico, sendo passível de refinamento sempre que necessário.

3.3.4 Ferramentas para construção de ontologias

Embora necessária, a tarefa de construção de ontologias, seja no campo da engenharia do conhecimento, inteligência artificial ou mesmo no trabalho terminológico, é dispendiosa, e qualquer apoio informatizado pode representar ganhos significativos.

Entretanto, o apoio informatizado o qual fazemos referência não é a construção automática de ontologias, já que este trabalho é notoriamente complexo para ser feito automaticamente, sem a interferência humana. Na verdade, ainda que existam pesquisas sobre a criação de algoritmos para construção automática, principalmente na área de inteligência artificial, como destacam os trabalhos de Junior (2008) e Moura et al. (2008), nossa menção ao apoio informatizado se refere ao auxílio para a edição das ontologias eletronicamente, principalmente na criação de ferramentas informatizadas que ofereçam suporte necessário para facilitar sua montagem, permitindo a definição das relações entre conceitos, pesquisas de conteúdo, visualização da ontologia e, sobretudo, que ofereça a estrutura de dados adequada e satisfatória para o correto armazenamento.

No contexto do trabalho terminológico, há atualmente várias ferramentas com o propósito de editar ontologias. Entretanto, de acordo com Hernández (2002) a maioria das ferramentas são separadas dos sistemas de gestão terminológica, e funcionam como um programa à parte, não permitindo em um único ambiente de trabalho a construção de uma ontologia. Esta característica torna-se uma desvantagem considerável, pois existe, devido à separação dos programas, a necessidade de intercâmbio entre a ferramenta de edição de ontologia e o sistema de gestão terminológica, o que nem sempre é uma tarefa fácil.

Um dos sistemas que fogem a esta regra é o *OntoTerm* Hernández (2002), Feliu et al. (2004). O *OntoTerm* é um sistema de gerenciamento de bases terminológicas composto por quatro módulos de trabalho, sendo eles: 1) Editor de Ontologias (*Ontology Editor*), 2) Editor da Base de Termos (*TermBase Editor*), 3) Navegador de Ontologias (*Ontology Navigator*) e 4) Gerados de Relatórios HTML (*HTML Report Generator*). Embora a principal característica do *OntoTerm* seja o gerenciamento de bases terminológicas (aplicação do módulo *TermBase Editor*), principalmente nas atividades de edição dos termos, a primeira tarefa a fazer no uso deste sistema é criar uma ontologia (aplicação do módulo *Ontology Editor*), visto que nenhum termo pode ser

inserido na base sem antes estar acoplado à ontologia.

No conjunto de ferramentas específicas de criação de ontologias não pertencentes a um ambiente terminográfico específico, mas que também podem ser usadas para o trabalho terminológico, podemos citar a *Ontolândia*, o *Protégé*, o *CODE4* o *IKARUS* e o *pOWL*.

A *Ontolândia* Farquhar et al. (1995) é o nome genérico de *Ontology Server*, e consiste em uma ferramenta para a construção colaborativa de ontologias, oferecendo um conjunto de serviços que possibilitam a construção de ontologias compartilhadas, por meio do acesso a uma biblioteca de ontologias e um editor de criação e navegação.

O *Protégé* foi desenvolvido pelo Grupo de Modelagem de Conhecimento da *Universidade de Stanford* e consiste em uma ferramenta de código aberto que forma um ambiente interativo para projeto de ontologias. Oferece uma interface gráfica para edição de ontologias e uma arquitetura para a criação de ferramentas baseadas em conhecimento.

O *CODE4* (*Conceptually Oriented Description Environment*) Skuce e Lethbridge (1995) é uma ferramenta de propósito geral para o gerenciamento de uma base de conhecimento. Possui diferentes modos de herança e inferência, interface gráfica de fácil uso, modo de hipertexto para navegação e utilitários para leitura de documentos e gerenciamento léxico. Este sistema surgiu das atividades do projeto *Cognitern*, iniciado em 1991 Hernández (2002) e é resultado da evolução do sistema *IKARUS* (*Intelligent Knowledge Acquisition and Retrieval Universal System*) Skuce (1996) que, por sua vez, foi escrito em *Applets Java*, para explorar as capacidades do ambiente Web. O *IKARUS* explora as capacidades cooperativas do ambiente Web utilizando uma representação hierárquica gráfica que permite herança múltipla.

O *pOWL* Auer (2005) é uma plataforma Web para a criação e desenvolvimento de ontologias voltadas para a Web Semântica. Escrito na linguagem PHP seu principal objetivo é servir uma solução baseada na Web para o gerenciamento e edição de várias ontologias em todo seu ciclo de vida, seguindo o padrão OWL.

Existem ainda várias outras ferramentas, fora do contexto da Terminologia que são utilizadas para a criação de ontologias. A Tabela 3.2, baseada no trabalho de Almeida (2003b), embora não exaustiva, mostra algumas dessas ferramentas.

3.3.5 Problemas relacionados com criação da estrutura conceitual do domínio

A principal dificuldade relacionada à criação da ontologia de um domínio específico é a dificuldade existente no momento de definição dos campos nocionais que correspondem à estrutura da ontologia. Isso porque, geralmente, o entendimento da organização estrutural dos conceitos de uma determinada área objeto é muito subjetiva e dependente da compreensão da mesma por parte do responsável pela sua elaboração.

Tabela 3.2: Ferramentas de edição de ontologias.

GKB-Editor (Generic Knowledge Base Editor)	Ferramenta para navegação e edição de ontologias através de sistemas de representação baseados em frames. Oferece interface gráfica, onde os usuários podem editar diretamente a base de conhecimento e selecionar a parte que é de seu interesse.
JOE (Java Ontology Editor)	Ferramenta para construção e visualização de ontologias que proporciona gerenciamento do conhecimento em ambientes abertos, heterogêneos e com diversos usuários. As ontologias são visualizadas como um diagrama entidade-relacionamento, em estrutura em árvore como na visualização <i>folder-tree</i> .
APECKS (Adaptive Presentation Environment for Collaborative Knowledge Structuring)	É um servidor de ontologias que permite trabalho cooperativo através da criação de ontologias pessoais pelos usuários. Estas ontologias podem ser comparadas com outras e é possível a discussão sobre as diferenças e similaridades entre elas.
OilEd	É um editor de ontologias de código aberto que permite construir ontologias utilizando a linguagem OIL. Entretanto, não é um ambiente para desenvolvimento de ontologias e a verificação da consistência e completo classificação deve ser executada por outra ferramenta.
OntoEdit	É um ambiente gráfico para edição de ontologias que permite inspeção, navegação, codificação e alteração de ontologias. O modelo conceitual é armazenado usando um modelo de ontologia que pode ser mapeado em diferentes linguagens de representação (XML, RDF, OIL) e em bancos de dados relacionais.
WebODE	Ambiente que fornece suporte à maioria das atividades de desenvolvimento de ontologias baseado na disponibilização de uma API (<i>Application Programming Interface</i>) que facilita a integração com outros sistemas, por meio da importação e exportação de ontologias em linguagens de marcação.
WebOnto	Ferramenta que possibilita a navegação, criação e edição de ontologias, representadas na linguagem de modelagem OCML. Permite o gerenciamento de ontologias por interface gráfica, inspeção de elementos, verificação da consistência da herança e trabalho cooperativo.

Visto que o terminólogo, na maioria dos casos, não detém conhecimento suficiente do domínio para definir a estrutura da ontologia, sua criação conta com a participação ativa dos especialistas da área. Mesmo entre os especialistas da área objeto tal compreensão pode ser distinta, provocando opiniões diferentes sobre a estrutura da ontologia, e consequentemente, ser difícil chegar em um consenso. Nestes casos, o que deve ser feito é um acordo entre as partes para definir uma única ontologia, e considerá-la como certa.

Outro problema que merece destaque é em relação às ferramentas de criação e edição de ontologias. Conforme dito anteriormente, estas ferramentas, em sua grande maioria, estão separadas dos sistemas de gestão terminológica, dificultando a comunicação de dados entre as etapas do trabalho. O ideal seria que tais ferramentas fizessem parte do conjunto de aplicações desses sistemas, principalmente para facilitar a integração dos dados entre as etapas do trabalho. Além disso, estando em um mesmo conjunto, as ferramentas de edição de ontologias seriam mais uniformes.

A estrutura da ontologia em formato de árvore que as atuais ferramentas assumem é outro problema que merece ser evidenciado. O fato é que a estrutura em árvore desampara, até certo ponto, as características comunicacionais que são inerentes à teoria da TCT. Isso porque a estrutura arbórea mapeia com multiplicidade a ocorrência da polissemia e da pluralidade conceitual que podem ocorrer nas dimensões dos termos no âmbito da TCT. Tal fato acontece porque na estrutura de árvore não é possível que um determinado “nó filho” tenha dois ou mais “nós pais” diferentes, fazendo com se tenha na estrutura da ontologia dois ou mais nós com o mesmo nome (ou identificador), ligados, por sua vez, em nós pai diferentes.

Essa deficiência seria contornada se a estrutura da ontologia fosse um grafo, ao invés de árvore. Diferentemente da estrutura em árvore, os grafos permitem a ligação entre os nós que representam a polissemia e/ou a diversidade conceitual que pode ocorrer na ontologia, através das ligações entre um “nó filho” e dois ou mais “nós pais”. Dessa maneira, a representação de ontologias em grafos suporta as atuais relações entre os termos e as múltiplas facetas nocionais que ele pode assumir.

3.4 Etapa 04 - Inclusão dos candidatos a termos na ontologia e a validação do especialista

Uma vez terminadas as etapas 2 e 3, teremos como resultado a lista de candidatos a termos e a ontologia da área objeto. Estes resultados serão utilizados na próxima etapa do processo de criação dos produtos terminológicos que compreendem, respectivamente, as tarefas de inclusão dos candidatos a termos na ontologia e a validação destes pelo especialista do domínio.

A atividade de inclusão dos candidatos a termos consiste em alocar todos os termos extraídos do corpus na estrutura conceitual da área objeto, representada pela ontologia. Ou seja, esta tarefa compreende em analisar cada termo extraído e encaixá-lo na estrutura hierárquica da ontologia, atribuindo a cada termo uma herança nocional. Ao executar esta tarefa, cada termo ganha uma carga de informação cognitiva referente ao domínio do conhecimento, sendo esta primordial para a transferência do saber especializado.

O resultado final desta inserção pode ser visto como uma intercalação entre a ontologia e a lista de candidatos a termos, sendo representado pela própria ontologia com os termos candidatos alocados, representado na maioria das vezes pelo último nível hierárquico da estrutura do conhecimento.

A partir do momento que os termos estão alocados na ontologia, pode-se proceder à sua validação pelos especialistas. A validação de termos pelos especialistas é a segunda fase desta etapa e é feita seguindo dois critérios: 1) *critério semântico* e 2) *critério estatístico*. A validação segundo os *critérios semânticos* ocorre da seguinte maneira: selecionam-se da ontologia determinados campos nocionais e pede-se que cada especialista assinale os termos considerados semanticamente relevantes em cada campo.

Depois da validação feita pelos especialistas, procede-se à complementação da validação utilizando os *critérios estatísticos* que, por sua vez, é feita pelo terminólogo. Isso significa que tanto as informações de frequência quanto as semânticas estão sendo consideradas, e é o ideal em uma pesquisa terminológica. Dessa maneira, os termos obtidos por meio da validação de cada um desses critérios irão constituir as fichas terminológicas (como será descrito na próxima seção, na Etapa 05).

3.4.1 Problemas relacionados às tarefas de inclusão e validação dos termos candidatos

A tarefa de inclusão de termos na ontologia é sempre um trabalho demorado e difícil, com grande carga cognitiva. Isso porque tal inclusão exige do profissional um abrangente conhecimento da área objeto, bem como das características e peculiaridades das classes e elementos envolvidos na categorização.

Em Terminologia, este trabalho é ainda mais complicado, pois o conteúdo informativo que carrega o conhecimento a ser transmitido pelos termos depende diretamente de uma correta inclusão do termo na ontologia. Além disso, uma correta inserção dos termos nos trabalhos terminográficos propicia um ganho considerável na tarefa de definição terminológica, já que a localização dos termos na estrutura conceitual ajuda a entender a feição conceitual que o mesmo irá assumir.

No entanto, apesar da dificuldade inerente desta tarefa, existem pesquisas e experimentos na

literatura que tentam automatizar este processo, conforme descrito em Avancini et al. (2002). Estas pesquisas utilizam as técnicas de aprendizado de máquina e recuperação de informação para classificar um conjunto de termos em classes pré-definidas, da mesma forma como são utilizadas para a classificação de textos e documentos automaticamente. Instanciando essas técnicas para o domínio da Terminologia, os elementos a serem inseridos são representados pelos os termos e as classes por cada nível hierárquico da ontologia.

Os resultados destes trabalhos têm mostrado baixos valores de *escores* na classificação de termos, e apontam que a tarefa de classificar termos é muito mais difícil do que a de classificar textos, embora o método utilizado seja viável e necessite de melhoramentos, conforme salienta o próprio autor no trecho abaixo:

“Our experiments suggest that the approach is viable although large margins of improvement still exists: F1 scores values are still low at least if compared to the F1 scores values that have been obtained in text categorization research ... Anyway the very fact that the F1 scores are much lower than the ones usually obtained in text categorization by the same kind of extensional representation supervised learning approach proves that term categorization is a harder task than text categorization.”
(Avancini et al. (2002), 5)

Neste sentido, vale salientar que este trabalho de doutoramento não investiu esforços para criar metodologias de classificação automática de termos em ontologias. No entanto, ainda cabe no escopo deste trabalho o emprego de algoritmos e/ou métodos de confirmada relevância e resultados satisfatórios para esta tarefa, bem como dedicar um empenho futuro no desenvolvimento de ferramentas que forneçam dados para facilitar a atividade de inserção dos termos, tais como recuperação de contextos semânticos, dados estatísticos de frequências e inclusive acesso ao *cópus*.

Por sua vez, a tarefa de validação dos termos também carece de ferramentas informatizadas para empreender seu propósito. Atualmente, não existem programas dedicados exclusivamente para este fim, sendo utilizado, na maioria das vezes, diversos programas adaptados, como editores de textos e aplicativos gerenciadores de dados Almeida (2003a). Tais ferramentas, por exemplo, não permitem a validação colaborativa e simultânea das listas de termos candidatos inseridos na ontologia, exigindo do terminólogo um trabalho extra de junção e comparação das validações efetuadas, principalmente quando esta é feita por diversos especialistas.

Dessa forma, podemos afirmar que a carência de ferramentas informatizadas dedicadas exclusivamente às tarefas de inserção e validação dos termos é o principal problema relacionado com esta etapa do processo terminográfico.

3.5 Etapa 05 - Criação da ficha terminológica e elaboração da base definicional

Terminada a etapa de validação dos candidatos a termos, obtemos como resultado os termos considerados relevantes para fazer parte do conjunto terminológico do produto final, que a partir desse momento, não serão “candidatos a termos”, e sim, “termos”, com reconhecido valor especializado. Deste modo, a próxima etapa é a criação e da Ficha Terminológica e a elaboração da Base Definicional.

3.5.1 Criação da Ficha Terminológica

A Ficha Terminológica é um elemento crucial na organização de produtos terminológicos e um dos itens fundamentais para geração destes. Pode ser definida como um registro completo e organizado de informações referentes a um dado termo, consistindo em um verdadeiro dossiê que contém todos os dados e elementos pertinentes ao objetivo da pesquisa e trabalho focalizados. Nela, estão contidas informações indispensáveis, tais como: código de identificação do termo, contextos de uso e ocorrência (do *corpus*), dados sobre as variantes denominativas, sinônimos e antônimos, informações sintáticas, marcas de uso, equivalência(s) em língua(s) estrangeira(s), definição, informações enciclopédicas, construções recorrentes que o acompanham, além de informações de registro operacionais como nome do responsável pela coleta, datas de registro e revisão.

Portanto, a Ficha Terminológica é uma coleção de informações acerca de um termo, usada para extrair os dados que irão compor o verbete final do produto terminológico almejado. Desse modo, o registro dos dados na Ficha Terminológica dever ser o mais completo possível, consistindo em uma etapa imprescindível para o trabalho terminológico Krieger e Finatto (2004).

Nesse contexto, importa salientar que cada trabalho terminológico, em suas especificidades, pode exigir um tipo diferente de Ficha Terminológica que, em linhas gerais, alimentará diferentes produtos terminológicos. Assim, não há um modelo único e ideal de Ficha Terminológica que possa atender a todos detalhes e exigências dos diversos trabalhos, de maneira que é fundamental que cada Ficha seja bem planejada para refletir as necessidades do projeto, isto é: “*para quê*” e “*para quem*” se faz um determinado produto. Isso auxilia o terminólogo a prever quais campos deverão constar do protocolo de preenchimento da Ficha.

Seja em papel ou sob a forma de um registro informatizado em um banco de dados, a Ficha Terminológica deve permitir que todas as informações nela contida possam ser recuperadas. Em um banco de dados terminológico, cada Ficha Terminológica é um registro do banco, encabeçado pelo termo. Esse registro é composto por campos, interligados ou não, contendo dados relevantes

para a identificação do termo, seus valores e sua significação.

Apesar da diversidade decorrente das exigências de cada situação, os critérios principais para criar uma Ficha Terminológica são: validade, concisão, atualidade e complementaridade dos dados Pavel e Nolet (2002). Além disso, segundo Krieger e Finatto (2004), uma vez determinados os objetivos e a natureza do trabalho terminológico, há um consenso sobre quais informações seriam essenciais para cada registro de termo. São elas:

Entrada : corresponde ao termo propriamente dito e denomina o(s) conceito(s) referente(s) a ele;

Informações pertinentes ao conceito : corresponde à área ou subárea temáticas, definição, remissivas (conceitos relacionados, hiperônimos, hipônimos), e notas relativas ao conceito;

Informações relativas à denominação : corresponde à categoria gramatical do termo, bem como os sinônimos, antônimos, abreviaturas, formas truncadas, fraseologia e informações morfológicas;

Fontes : consiste no registro das fontes textuais, bibliográficas ou pessoais consultadas;

Dados da gestão da Ficha : corresponde aos dados operacionais da ficha, tais como a data de registro, da última alteração/atualização, do nome do revisor ou usuário responsável pelo registro ou atualização, etc.

São estas informações que precisam estar organizadas e armazenadas de maneira satisfatória para que sejam fáceis e rapidamente recuperadas.

3.5.2 Elaboração da Base Definicional

Concomitantemente à criação das Fichas Terminológicas, é realizada a elaboração da Base Definicional. A Base Definicional consiste em um repositório de excertos definicionais referentes aos termos, compilados de diversas e variadas fontes, tais como: livros, manuais, revistas científicas, dicionários de áreas conexas, dicionários de língua geral, *sites* da Internet e demais fontes que se mostrarem úteis para a obtenção de contextos explicativos e/ou definitório sobre os termos que serão definidos.

Em outras palavras, a Base Definicional é um recurso de análise que reúne os termos e as provas textuais (definições, contextos, exemplos de uso, unidades fraseológicas, observações do terminólogo ou pareceres de especialistas da área) resultantes da pesquisa dos termos. A compilação dos excertos para compor a Base Definicional é uma etapa delicada da análise terminológica, pois envolve a avaliação da informação coletada referente aos traços semânticos do

conceito e a seleção de provas textuais mais pertinentes que serão registradas na Ficha Terminológica para posterior definição dos termos, bem como sua harmonização e publicação.

A Base Definicional é essencial para o trabalho terminológico, pois depois de compilada, é a partir da leitura desses excertos que o terminólogo redige, então, o texto da definição. A redação final da definição, por sua vez, deve estar de acordo com padrões lingüísticos pré-estabelecidos obedecendo a uma tipologia definicional. Somente após a redação da definição é que esse campo pode ser preenchido na Ficha Terminológica.

Apesar de complexa e custosa, a fase de redação da definição terminológica é muito importante para o trabalho terminológico que objetiva a elaboração de obras especializadas, já que um bom produto terminológico (seja um dicionário ou glossário) se avalia, principalmente, pela qualidade das suas definições.

Segundo Pinto e Oliveira (2004), a definição é toda a frase ou segmento de frase que possa descrever e caracterizar um dado termo. Da mesma forma, Sager *apud* Krieger e Finatto (2004), considera a definição como um produto que representa a descrição lingüística de um conceito, baseada em uma lista de características particulares que transmitem o significado desse conceito, por meio de um texto definitório que possui a forma de uma predicação sobre um termo. Mais especificamente, vale notar que as definições especializadas, no contexto da Terminologia, descrevem um conceito dentro de um campo temático especializado.

Conforme apontado no Capítulo 2 o gênero próximo e a diferença específica têm sido tomados com parâmetros de qualidade para observação do enunciado definitório. Em Terminologia, estes parâmetros também desencadearam uma tradição do estudo da definição, principalmente devido à função de verificação da qualidade do enunciado definitório, pela análise do gênero e da diferença.

Entretanto, no desenvolvimento da Terminologia sob a luz da TCT, a definição terminológica não se reduz apenas a estas categorias, ainda que pareçam muito úteis Krieger e Finatto (2004). De fato, a definição terminológica ultrapassa a observação do enunciado definitório terminológico somente pelo gênero próximo e a diferença específica, e deve considerar também a área de conhecimento objeto da terminologia e as necessidades de informação do público alvo a que se pretende atingir com a obra especializada.

Dessa maneira, para contemplar essas exigências, costuma-se redigir também as informações enciclopédicas referentes ao termo, juntamente com a redação da definição. Ambas são tratadas de modo diferente porque normalmente a definição é um campo obrigatório do verbete, e a informação enciclopédica não. Todavia, vale destacar que tanto as definições quanto as informações enciclopédicas são validadas pelos especialistas para que, se for o caso, serem inseridas nos campos correspondentes da Ficha Terminológica, o que permitirá a edição final dos verbetes

(visto na Seção 3.6.1).

Existem outros aspectos de grande importância que devem ser considerados no momento de formular as definições terminológicas. O terminólogo no ofício de definir, não só deve estar consciente das possíveis ambigüidades e conotações inerentes ao conteúdo lingüístico do texto definitório, mas também estar ciente que as definições são o reflexo da comunidade científica sobre as estruturas do conhecimento que são representadas pelas definições. Neste sentido, como indica Hernández (2002), os terminólogos e especialistas devem ser cuidadosos para encontrar o equilíbrio entre a precisão e a excessiva restrição denominativa. A primeira é necessária para estabelecer a identificação entre o termo e a definição. A segunda pode ser imprópria para dar conta do caráter dinâmico das estruturas de conhecimento.

Além disso, o enunciado da definição deve seguir padrões lingüísticos pré-estabelecidos, o que não ocorre com a informação enciclopédica. Importa destacar que esses padrões lingüísticos podem ser sistematizados, por meio de um *Editor de Definição*, sendo possível criar uma espécie de tutorial que auxilie o terminólogo, ou editor, a redigir a definição. Assim, é possível semi-automatizar esse procedimento extremamente relevante para o trabalho terminológico, já que a definição é a essência de qualquer produto terminológico.

A fixação de padrões lingüísticos para a redação da definição favorece o estabelecimento de uma medida de controle dos usos do vocabulário, pois atitudes nessa direção, como bem lembra Finatto (1998), estão relacionadas à concepção de que a comunicação especializada requer um elevado grau de precisão comunicativa, aspecto para qual o emprego de terminologias contribui largamente.

Considerando ainda que a finalidade maior de uma obra de referência especializada é atender às necessidades de informação dos consulentes, este trabalho de doutorado seguiu as orientações descritas em Almeida (2000) que afirma que as definições são elaboradas levando-se em consideração orientações como: 1) o produto terminológico (dicionário e glossários) tem a função primária de facilitar a comunicação, de maneira que o texto definitório deve ser suficientemente claro e completo para que o consulente entenda. Assim, ainda que os tipos de definição sejam utilizados como orientação, eles não devem subjugar o texto; e 2) não se pode estabelecer com exatidão que a(s) característica(s) intrínseca(s) deve(m) figurar no texto da definição e a(s) extrínseca(s) na informação enciclopédica, pois nem sempre se pode classificar com segurança o que são características (ou traços distintivos) intrínsecas e extrínsecas do conceito cujo termo está sendo definido.

Entretanto, mesmo obedecendo tais orientações, salientamos que uma série de procedimentos no que se refere à redação da definição terminológica podem ser sistematizados, facilitando tanto a elaboração quanto a validação.

3.5.2.1 Normas ISO para redação das definições

Embora a imposição de padrões rígidos de definições seja contestável no âmbito da redação do enunciado definitório, as normalizações da ISO a respeito da definição terminológica são diretrizes úteis para a organização do texto da definição, principalmente para que o termo possa ser compreendido pelo usuário consulente.

Um exemplo dessa norma é a ISO 704 (1987), que traz orientações básicas para redação de definições que figuram em dicionário ou repertórios terminológicos. Em linhas gerais, as recomendações das normas da ISO tratam da objetividade de formulação do enunciado e procuram garantir um adequado fluxo de informação.

Umas das recomendações indicadas nessas normas são chamadas *definições por intenção* e as *definições por extensão*. A *definição por intenção* é a definição clássica com a simples indicação e relevância clara do gênero próximo e diferença específica. A *definição por extensão* consiste na relação de entidade ou exemplos que representam o significado ou conceito. Um exemplo desse último tipo de definição pode ser:

Linguagens para Web: PHP, ASP, JavaScript, Perl, Java, Python, JSP.
--

Outra recomendação das normas ISO estabelece que a apresentação das características mais essenciais da entidade que se define devem ser priorizadas e redigidas primeiro, de maneira direta. Comentários ou explicações adicionais devem vir em segundo plano, possivelmente nas informações enciclopédicas, visto que o enunciado deve informar com primazia o que é o termo, deixando em segundo lugar outras informações ou indicações.

Além dessas recomendações, podemos salientar também que os textos das definições devem ser coerentes e completos, assim como devem ser evitadas definições que iniciem por uma negação ou por erros de concordância de número, por exemplo, formular uma definição no plural para um termo no singular.

3.5.3 Problemas relacionados com a criação da Ficha Terminológica e a elaboração da Base Definicional

Em relação à criação da Ficha Terminológica, o principal problema diz respeito à correta definição dos campos que irão compor as Fichas. Conforme dito anteriormente, tal definição é importante porque a etapa de edição dos verbetes e a difusão dos termos dependem diretamente da composição dos campos da Ficha Terminológica. É nesse momento, então, que o terminólogo precisa estar atento e sintonizado com os objetivos do trabalho terminológico para criar uma Ficha que atenda completamente tais objetivos. Além disso, uma Ficha Terminológica que não contempla os propósitos do trabalho empobrece o produto terminológico.

Esse problema ganha bastante importância se a Ficha Terminológica não seguir o formato eletrônico, já que uma possível correção das Fichas Terminológicas (no final do processo), em papel é um processo caro. Contudo, isso não significa dizer que se a Ficha estiver em formato eletrônico tal trabalho não é difícil ou moroso. Entretanto, vale ressaltar que a Ficha em formato eletrônico oferece mais flexibilidade para possíveis correções e ajustes futuros, em caso de erro, embora ainda dependam muito do tipo de correções a serem feitas. Por exemplo, se for adicionado um campo cuja informação seja peculiar a cada termo individualmente, é necessário verificar a Ficha de cada termo, mesmo que eletronicamente, o que se torna uma tarefa trabalhosa. Portanto, é imperativo que a Ficha Terminológica seja eletrônica e atenda a todos os objetivos do trabalho terminológico em questão, bem como contemple todas as informações disponíveis referentes aos termos.

O principal problema da Base Definicional, por sua vez, é o armazenamento e consulta dos excertos e conteúdos que compõem a base, encontrados nas mais diversas fontes de dados. Atualmente, não existem programas dedicados exclusivamente à elaboração da Base Definicional, principalmente aqueles integrados aos sistemas de gestão terminológica, que sejam multiusuários e forneçam funcionalidades de consulta e avaliação colaborativa de excertos. Isso porque, geralmente, o incremento da Base Definicional é feita por mais de um usuário, geralmente o editor da base, que compõe a equipe de trabalho, e que necessita de um correto armazenamento e rápida consulta dos conteúdos armazenados durante a compilação dos excertos. Além disso, é comum, entre os editores da equipe, a troca de excertos para avaliação, correção e validação do conteúdo armazenado.

Os programas e/ou ferramentas que são utilizadas para a elaboração da Base Definicional são, geralmente, editores de textos (por exemplo: *Microsoft Word*, *Microsoft WordPad*, *Bloco de Notas*) ou planilhas eletrônicas, com o *Microsoft Excel*. Estes programas²⁹ além de não serem explicitamente criados para o trabalho terminológico, possuem várias limitações que impedem um adequado manejo das tarefas inerentes ao incremento da Base Definicional. Um exemplo dessas limitações é a dificuldade do controle de versões da Base Definicional editadas por vários usuários. Isso porque tais programas trabalham sobre arquivos de dados em formato próprio, e mesmo não trabalhando simultaneamente, fica difícil saber qual usuário fez a última alteração, bem como quais alterações foram feitas, impossibilitando o trabalho colaborativo. Além disso, se o trabalho fosse feito de maneira simultânea, haveria o risco de perda de informações.

Nesse contexto, seria ideal se a Base Definicional fosse incrementada utilizando uma ferramenta inserida em um sistema de gestão terminológica, com objetivos explícitos para terminologia, implementada sobre um banco de dados relacional, que suportasse a edição simultânea dos

²⁹Estes programas são utilizados pelo Grupo de Estudos Terminológicos, GETerm, do Departamento de Letras (DL) da Universidade Federal de São Carlos (UFSCar).

dados, o controle de versões, bem como o armazenamento seguro e a consulta dos dados.

A tarefa de elaboração da definição terminológica é, por si só, uma atividade difícil e lenta, cuja virtude e presteza muito depende da Base Definicional. Dessa maneira, a existência de uma boa Base Definicional, que seja não somente volumosa, mas que também contenha excertos de qualidade auxilia a tarefa de definir deixando-a menos trabalhosa. De outra forma, dado que a ação de definir é diretamente ligada com ação de redigir um texto (da definição), o uso de um editor de definição, voltado estritamente para o trabalho terminológico de redação, é muito bem-vindo, principalmente se vierem acoplados a este corretores ortográficos (*spelling*) e algumas tipologias definicionais, conforme dito anteriormente.

3.6 Etapa 06 - Edição dos verbetes e uso de formatos para facilitar o intercâmbio e difusão dos produtos terminológicos

Após elaboração da Base Definicional e da redação das definições do conjunto de termos, começa a sexta e última etapa do trabalho terminológico, que corresponde à edição dos verbetes e o uso de formatos para representação, publicação, intercâmbio e difusão dos produtos, quando estes forem objetivo ou intenção da equipe responsável.

3.6.1 Edição dos verbetes

A edição dos verbetes nada mais é do que a seleção de determinados campos da Ficha Terminológica para constarem do modelo do verbete final, ou seja, na apresentação dos termos ao consulente. Via de regra, há nos verbetes informações sistemáticas (obrigatórias em todos os verbetes) e não-sistemáticas (informações não recorrentes). O trabalho de doutoramento de Almeida (2000) utiliza um determinado modelo de verbete para a versão impressa de dicionários especializados. Do mesmo modo, o modelo de verbete que nos pareceu bastante satisfatório é o apresentado no *Glossário de termos neológicos da economia*, projeto coordenado pela Profa. Dra. Ieda Maria Alves, USP-SP, 1998. A Figura 3.6 mostra uma tela com verbetes exemplos deste glossário.

O importante nesse contexto é permitir que o terminólogo ou, algumas vezes até o próprio consulente, selecione, entre os campos da Ficha Terminológica, quais campos ele deseja expor na apresentação do verbete, atendendo suas necessidades ou objetivos do trabalho terminológico.

No entanto, é notório que a edição dos verbetes depende muito da cobertura e do formato da Ficha Terminológica, pois caso a ficha contenha poucos campos de dados, ou não atenda os objetivos do trabalho, a edição dos verbetes fica prejudicada. Da mesma forma, é nesta direção que salientamos a importância de desenvolver a Ficha Terminológica em formato eletrônico,

Figura 3.6: Exemplos de verbetes do *Glossário de termos neológicos da economia*.

possibilitando a inclusão, exclusão ou alteração dos campos de dados, já que o ideal na tarefa de edição dos verbetes é a flexibilidade de escolha dos campos da Ficha Terminológica que irão compor o verbete final. Por isso, é muito relevante a participação da informática nesta tarefa, principalmente na criação da Ficha Terminológica e seleção dos campos dos verbetes.

3.6.2 Intercâmbio e difusão dos termos

O importante papel que a terminologia assume para a sociedade atual é devido ao desenvolvimento da comunicação moderna, especialmente nos âmbitos especializados que, por sua vez, deve ir além da fronteira de um trabalho particular, passando a exigir um intercâmbio de conhecimentos e um fluxo dinâmico de informações. Dessa forma, diante da funcionalidade operada pelos termos especializados na transmissão do conhecimento, justifica-se a necessidade de sua divulgação em um contexto de integração da sociedade Krieger (1998).

Portanto, facilitar as condições de intercâmbio e difusão de conhecimento é uma estratégia que esbarra no acesso aos repertórios terminológicos, visto que o interesse não se restringe mais aos especialistas, que como usuários diretos, sempre compreenderam a necessidade de dominar as terminologias, mas principalmente para os usuários consulentes, que cada vez mais consomem estes produtos comunicacionais, especialmente no cotidiano das notícias que abarca o público não especializado.

Assim, a difusão e o intercâmbio de informações especializadas, tanto entre os profissionais entre si como entre os profissionais e os leigos se tornam mais urgentes. Nessa interface, o indivíduo não especializado carece do apoio da obra de referência terminológica capaz de orientá-lo nos intricados caminhos da área científica e técnica. O produto terminológico, portanto, ao se originar e se dirigir para o uso da linguagem de especialidade se constitui em facilitador da comunicação e difusor de conhecimento Maciel (1998).

Desse modo, o intercâmbio de dados rápido e eficiente entre os diferentes bancos ou bases de dados de terminologia é uma necessidade atual, principalmente se considerarmos que a terminologia moderna deve fazer frente de forma inevitável à necessidade de reutilizar a informação terminológica.

Entretanto, a falta de homogeneidade e sistematicidade na organização e no modo de registrar informação sobre uma dada terminologia pode impedir a comunicação entre os usuários, bem como entre os diferentes sistemas de gestão terminológica. Ou seja, quando cada equipe faz o trabalho de um jeito, sem critérios de organização para a informação coletada, a ausência de um padrão para organização tanto no formato de apresentação quanto no armazenamento de dados em um repertório ou acervo terminológico pode dificultar a transferência de dados entre acervos que tenham sido produzidos por equipes diferentes.

3.6.2.1 Formatos e padrões de intercâmbio e modelo de dados

Seria de esperar, portanto, que em uma atividade que tem como principal objetivo a organização e armazenamento estruturado de um corpo de informação complexo (uma base de dados terminológica) fosse feito um esforço por adotar determinadas técnicas e recursos direcionados para garantir o intercâmbio e reutilização de dados. Entretanto, alguns dos esforços de normalização, em sua vertente de regulação metodológica, foram e estão sendo elaborados por parte de organismos de padronização internacionais, tais como a ISO³⁰, TEI³¹ (*Text Encoding Initiative*), e o LISA³² (*Localization Industry Standards Association*).

Atualmente, a maioria dos sistemas de gestão terminológica existentes no mercado permite a exportação de seus dados para o formato texto ou texto enriquecido (com extensão RTF), mas quase nenhum permite a importação de dados de outras fontes que não sejam arquivos em seu próprio formato “proprietário”. Desde que o uso dos recursos seja voltado para uso interno, tal fato não causa nenhum problema. Entretanto, o uso de formatos proprietários não permite o intercâmbio e a reutilização efetiva por parte de outros sistemas ou consumidores, característica desejável e necessária para os trabalhos e profissionais da terminologia.

De acordo com Budin e Melby (2000a), um grande número de padrões de intercâmbio tem surgido, principalmente nos trabalhos desenvolvidos nas áreas de tradução automática, sistemas de autoria e alinhamento de textos. Alguns exemplos desses são os padrões da ISO tais como o ISO 12200 (*MARTIF - MACHine-Readable Terminology Interchange Format*) e 12620, bem como os derivados destes padrões como: 1) o LISA TMX (*Translation Memory eXchange*) e TBX (*TermBase eXchange*), 2) o Geneter (*Modèle GENERique pour la TERminologie*) Melby (1999),

³⁰<http://www.iso.org/>

³¹Para mais informações acesse: <http://www.tei-c.org/>

³²Para mais informações acesse: <http://www.lisa.org/sigs/terminology/>

3) o IIF (*Interval Interchange Format*) no âmbito do Projeto INTERVAL³³ PROJECT (1998), 4) o OLIF (*Open Lexicon Interchange Format*) no âmbito do Projeto Otelo, 4) o TRANSTERM Transterm (1996) e 5) o TEI-Term (*Terminology Interchange Format*) uma iniciativa do grupo TEI. A Tabela 3.3 agrega importantes informações sobre estes e outros formatos.

Tabela 3.3: Formatos de intercâmbios de dados terminológicos e lexicográficos.

Terminology Interchange Format	TEITerm	Formato baseado em SGML para o intercâmbio de dados usando cabeçalhos TEI, usados para codificar textos em formato eletrônico.
MAchine-Readable Terminology Interchange Format	MARTIF	Formato baseado em SGML para o intercâmbio de dados terminológicos, baseado na norma ISO 12200. Foi derivado e ligeiramente modificado a partir do formato TEITerm, e posteriormente envolvido no desenvolvimento do TBX.
Terminology Markup Framework	TMF	Meta-modelo de dados baseado na ISO 16642 que define as estruturas necessárias para a representação computacional dos dados terminológicos.
Modèle GENERique pour la TERminologie	Geneter	Linguagem de marcação que fornece um modelo genérico baseado em SGML para representação de dados terminológicos servindo de formato intermediário entre diferentes aplicações.
Open Lexicon Interchange Format	OLIF	Formato padrão utilizado para o intercâmbio de dados entre bases e recursos lexicais, principalmente para os sistemas de tradução automática.
TermBase eXchange	TBX	Formato baseado em XML para o intercâmbio de dados terminológicos, baseado na norma ISO 16642.

Embora existam muitos formatos ou padrões de intercâmbio, há uma lacuna na definição de qual formato deve ser utilizado nos trabalhos terminológicos Budin e Melby (2000a), decorrente da falta de uma coordenação central entre as iniciativas, resultado da ausência de interfaces entre as aplicações que utilizam estes formatos ou padrões já definidos. Dessa maneira, conforme salienta Hernández (2002) o crescimento desordenado dos padrões de intercâmbio é já por si só razão suficiente para tratar de encontrar um ponto comum.

Todavia, não faltam iniciativas para a definição de um formato único Melby (2001). Um exemplo dessas iniciativas está descrito em Melby (1999), um comunicado informal sobre a semelhança dos padrões *Geneter* e *MARTIF* na tentativa de fazer um único formato a partir de uma intercalação dos dois. Nesta mesma linha, a norma ISO 12200 que representa o padrão *MARTIF* foi desenvolvida com a finalidade de facilitar o intercâmbio de recursos terminológicos em formato eletrônico. O objetivo do padrão *MARTIF* é servir de “interlíngua terminológica”, garantindo o intercâmbio entre usuários (indivíduos ou empresas) que utilizam sistemas de gestão terminológica com formatos internos diferentes.

Outra iniciativa que tenta criar um único formato padrão de intercâmbio é o Projeto SALT³⁴

³³Mais informações sobre o Projeto Interval acesse o endereço: http://www.computing.surrey.ac.uk/ai/new_interval/

³⁴Mais informações sobre o projeto SALT podem ser obtidas no endereço: <http://www.ttt.org/salt/index.html>

(*Standards-based Access to multilingual Lexical and Terminological Resources*) Budin e Melby (2000a). O Projeto SALT (Translation e Technology (2000)) é um consórcio formado por grupos acadêmicos, governamentais, associações e representantes comerciais dos Estados Unidos e da Europa cujo objetivo é criar um meta-formato para o intercâmbio de bases de dados terminológicas e léxicos para tradução automática. Este meta-formato é baseado nos formatos *MARTIF*, *OLIF* e *TBX* com incorporação *TMX*, *Geneter*, *IIF* e *TEI*, e sua função é incorporar as características dos formatos existentes para criar um único formato.

Outro aspecto importante que deve ser levado em consideração ao utilizar um formato padrão de intercâmbio de dados é a perda de informação. Neste sentido, destaca-se o formato (ou modelo de dados) utilizado na aplicação ou sistema gestor de terminologia. Segundo Hernández (2002), nos sistemas de gestão terminológica a disparidade de modelos de dados é maior do que em princípio se pode imaginar, dado que muitos deles utilizam um modelo de dados particular para cada uma de suas bases de dados. Consequentemente, não é possível criar um algoritmo de conversão de um modelo de dados a outro, dado que não se conhece de antemão qual é o modelo de dados.

Nessa mesma linha, podemos citar o *CLS Framework* Hardman (1999), baseado na ISO 12620 ISO 12620 (1999). O principal objetivo do *CLS Framework* é a criação de um modelo genérico de base de dados que sirva de padrão para o desenvolvimento de diferentes tipos de bases de dados terminológicas, sejam estas textuais, relacionais ou orientadas a objetos; visando possibilitar o desenvolvimento de rotinas de conversão de informação entre diferentes repositórios terminológicos para que o intercâmbio de informação seja possível.

Nesse contexto, o intercâmbio de informações terminológicas entre dois sistemas pode ocorrer sob três situações:

1. Tanto o sistema A como o B se baseiam em um modelo de dados único, ainda que diferentes o um do outro. Neste caso o intercâmbio é factível, ainda que custoso, pois tanto os fabricantes dos sistemas A quanto do B podem criar os algoritmos de conversão. A perda de informação acontecerá naqueles casos em que os modelos de dados forem muito diferentes;
2. O sistema A baseia-se em um modelo de dados único, enquanto o sistema B baseia-se em um modelo de dados particular. Neste caso o fabricante do sistema A pode construir um algoritmo de conversão que gere os dados no modelo de dados idêntico ao do sistema B. A perda de informação neste caso é variável, podendo até ser nula, e depende dos campos previstos no sistema A e B forem de alguma forma coincidentes;
3. Ambos sistemas, A e B, se baseiam em um modelo de dados particular. Neste caso é preciso verificar quais campos de dados são presentes nos dois modelos para verificar o que

pode ser feito, embora o índice de perda de informação possa ser grande.

Estas três possibilidades acontecem para o caso individual entre dois sistemas de gestão terminológica. O trabalho de conversão aumenta exponencialmente caso existam vários sistemas. Diante desta situação, duas ações podem ser tomadas para garantir o intercâmbio de informação entre os sistemas de gestão terminológica:

1. Antes da implementação dos sistemas, os fabricantes devem decidir qual formato ou modelo de dados será utilizado. Entretanto, neste caso, é melhor utilizar e desenvolver o mesmo sistema, ao invés de dois ou mais;
2. Utilizar um formato de intercâmbio cego, isto é, um formato amplamente aceito e conhecido pelos fabricantes dos sistemas.

Um formato de intercâmbio é chamado de cego quando não há a necessidade de determinado sistema gestor conhecer a origem dos dados vindos de outro sistema Hardman (1999). Ou seja, para o efetivo intercâmbio, basta que um determinado sistema exporte os dados neste formato para que outro sistema os importe, independente da procedência dos mesmos. A criação de um formato cego, segundo Hernández (2002), foi trabalho do Comitê Técnico 37 (TC 37) da ISO o que culminou na criação dos formatos *MARTIF* e o recente *TBX* (*TermBase eXchange*). Outro formato cego é o chamado de *BlindMARTIF* resultado da pesquisa de doutoramento de Alder (1998).

3.6.3 Problemas relacionados com a edição dos verbetes e o intercâmbio de produtos terminológicos

Conforme dito anteriormente, não há problemas com a edição dos verbetes se a Ficha Terminológica estiver em formato eletrônico. O que é necessário nesta tarefa de edição é oferecer ao terminólogo melhores condições de seleção dos campos da ficha que irão compor o verbete final. Apresentar os campos elegíveis bem como suas informações³⁵ (metadados) facilitam muito a atividade de seleção. Permitir que a edição do verbete seja flexível também é outro aspecto que ajuda esta tarefa, pois permite que os campos de apresentação dos verbetes sejam alterados sempre que necessário.

Quanto aos formatos ou padrões de intercâmbio, o principal problema é a escolha de formatos que sejam amplamente aceitos e que possibilitem as tarefas de importação e exportação de dados terminológicos. Entretanto, dada a grande quantidade de formatos existentes, é necessário realizar uma avaliação crítica de cada formato, para decidir qual(is) será(ão) usado(s) para as

³⁵As informações sobre os campos correspondem os metadados armazenados no momento da criação da Ficha Terminológica, que servem como um conjunto de informações referentes aos próprios campos, como sua função, tipo de dados, tamanho etc.

funções de intercâmbio. Esta avaliação deve analisar a aceitação do formato na comunidade acadêmica, comercial/industrial e governamental, a facilidade de conversão dos dados, tanto na importação quanto na exportação, a linguagem de marcação em que foi implementada (Texto, XML, RDF, SGML, etc.), e as rotinas de atualização do formato. Nesse sentido, há uma tendência em escolher aqueles formatos que são indicados pelos órgãos nacionais (no caso a Associação Brasileira de Normas Técnicas - ABNT) e internacionais (ISO) de normalização.

No âmbito dessa escolha do formato de intercâmbio, faz sentido também selecionar o padrão que o modelo de dados seguirá para o armazenamento dos dados terminológico. Esta seleção deve se basear principalmente nas categorias de dados que cada padrão abarca, inclusive para facilitar o posterior intercâmbio dos dados.

Como um bom exemplo dessas escolhas, sugerimos selecionar o formato *TBX* para o intercâmbio de dados, visto que este é implementado em XML e também segue as normas da ISO. No entanto, outras avaliações podem ser feitas, principalmente em relação ao Projeto SALT, que busca uma integração de vários formatos.

O Ambiente colaborativo Web e-Termos

O Capítulo 3 descreveu as tarefas e os problemas relacionados a todas as etapas do processo de gestão e criação de produtos terminológicos. Conforme pôde ser visto, estas tarefas, excluindo algumas como a redação da Base Definicional, a criação da Ficha Terminológica e a edição dos verbetes, podem ser executadas utilizando-se vários programas informatizados. Entretanto, os programas que executam estas tarefas, salvo a extração automática de candidatos a termos, raramente são desenvolvidos exclusivamente para o âmbito terminológico, sendo então adaptados para este fim. Além disso, a maioria destes programas é independente e não pertencem a um sistema integrado de gestão terminológica Hernández (2002), Cabré (2005).

Especificamente no Brasil, existe uma carência de recursos computacionais que sistematizam o trabalho terminológico. Os terminólogos brasileiros que coordenam grupos de pesquisa que desenvolvem produtos terminológicos operam na maioria das vezes com ferramentas adaptadas e incompatíveis, não elaboradas especificamente para a finalidade das suas pesquisas. Portanto, o uso de programas distintos para as tarefas terminológicas torna o trabalho dos terminólogos muito moroso, trabalho esse que geralmente é feito em equipe multidisciplinar que deve ter acesso, a todo o momento, à versão mais atual das várias etapas da geração de um produto terminológico. Além disso, programas distintos deixam o gerenciamento dos dados mais complicado, pois cada um possui seu próprio formato de entrada e saída de dados.

Ainda nesse contexto, a existência de um sistema integrado e via Web com a base de dados compartilhada e totalmente dedicada às tarefas terminológicas permite que os membros de uma equipe possam trabalhar remotamente, sem a necessidade de estarem no mesmo lugar e nem se preocuparem com a junção ou merge de dados posterior.

Como fundamento do projeto e desenvolvimento do **e-Termos**, entende-se que fazer Terminologia na era da Informática significa criar um conjunto de procedimentos automatizados ou semi-automatizados que dêem suporte àquelas tarefas do trabalho terminológico citadas no Capítulo 3. Deste modo, agregamos à nossa proposta metodológica a contribuição advinda da Informática, já que o crescimento de estudos e pesquisas na área de Lingüística Computacional e Lingüística de Córpus e o conseqüente aprimoramento e desenvolvimento de ferramentas computacionais voltadas para o processamento automático do português começam a interferir diretamente na prática terminográfica.

Sob os pressupostos teóricos da terminologia de orientação descritiva, o **e-Termos** pode ser visto com um sistema informatizado que contempla as atividades de desenvolvimento de terminologias. Sendo uma aplicação *Computer Supported Collaborative Work – CSCW*, o **e-Termos** é um Ambiente Colaborativo Web, composto por seis módulos de trabalho inter-relacionados, cujo propósito é automatizar ou semi-automatizar as tarefas de criação e gestão do trabalho terminológico.

4.1 Suporte teórico

Como base teórica para esta terminologia descritiva, os pressupostos de Cabré *apud* Almeida (2003a) foram norteadores e serviram de diretrizes para a esta pesquisa, sendo eles:

1. Os objetos terminológicos (conhecimento especializado, textos especializados e termos) devem ser estudados sob três perspectivas: a) a **social**, que se refere às características que um trabalho terminológico deve ter para atingir os fins a que se propõe sendo adequado às necessidades comunicativas dos profissionais e do público em geral; b) a **cognitiva**, que diz respeito ao conhecimento do domínio especializado em questão; e c) a **lingüística**, que se refere ao uso de um modelo que abrange a competência e a atuação;
2. O conjunto de termos não constitui uma língua marginal, eles pertencem ao sistema lingüístico e, por isso, estão sujeitos às mesmas regras que regem o funcionamento da língua natural;
3. A variação é considerada como fenômeno natural, inerente à linguagem e deve estar adequadamente descrita na terminologia não devendo ser reduzida para diminuir as ambigüidades;
4. Os conceitos atribuídos aos termos não possuem um funcionamento diferente dos significados da língua geral, podendo sim aparecer em níveis distintos de especialização;

5. Os conceitos são influenciados diretamente pelos aspectos socioculturais e lingüísticos das comunidades, bem como pelos canais por onde eles circulam e se difundem.

Finalmente, a escolha da TCT como base teórica de desenvolvimento do **e-Termos** se deu em função da consideração dos aspectos lingüísticos e comunicacionais das línguas de especialidade, além de tomar os textos, representados pelos *cópus*, com base de análise.

4.2 Aspectos gerais do **e-Termos**

O **e-Termos** como um Ambiente Colaborativo Web é, sob uma perspectiva genérica, dividido em duas grandes partes:

1. o portal do **e-Termos** e
2. o sistema **e-Termos**

O portal é de acesso livre e irrestrito e está direcionado ao público geral da Internet (usuário final) que deseja conhecer o **e-Termos**, pois reúne todas as informações gerais do sistema, como: a descrição, o projeto que originou sua construção, os objetivos, as etapas de trabalho definidas, as funcionalidades oferecidas, os perfis de usuários previstos, a equipe responsável pelo seu desenvolvimento e um conjunto de *links* relacionados ao tema de terminologia. Além disso, o portal também oferece três serviços: 1) um *Fale Conosco* no qual o internauta pode contatar a equipe responsável, 2) um formulário de cadastro, no qual o usuário pode realizar seu registro gratuito para ter acesso ao sistema (segunda parte) 3) um serviço de autenticação que permite um determinado usuário cadastrado acesse o sistema (segunda parte). Outra importante seção do portal é de **Acesso aos Produtos**, que corresponde a uma área específica dedicada exclusivamente à disponibilização dos produtos terminológicos criados pelo sistema, após a sua publicação final. Ou seja, além de reunir as informações do próprio **e-Termos**, o portal também destina um espaço gratuito para a difusão dos produtos desenvolvidos. A Figura 4.1 mostra página principal do portal.

A segunda parte corresponde ao *Sistema e-Termos*, também de acesso livre e público, mas restrito a apenas aos indivíduos que fazem cadastro. Pode ser visto como um sistema Web cuja entrada principal é um *cópus* de especialidade de um determinado domínio do conhecimento, e a saída um produto terminológico e/ou lexicográfico (por exemplo, glossário, dicionário, vocabulário, lista de termos, ontologia, etc.) do domínio em questão. Todas as funcionalidades que o sistema fornece têm a função primária de servir como suporte às tarefas envolvidas na criação destes produtos. No entanto, tais funcionalidades são abrigadas em um único ambiente computacional sendo transparente aos usuários. Os usuários, por sua vez, têm a participação

Figura 4.1: Página principal do Portal do e-Termos.

colaborativa, interagindo entre eles e com o sistema em diferentes fases do processo de elaboração do produto, a partir das interfaces Web que o Ambiente oferece, tendo acesso compartilhado aos dados e subprodutos gerados em cada módulo do sistema, fato que facilita e uniformiza as tarefas de gestão terminológica. A Figura 4.2 mostra uma das telas do sistema.

Figura 4.2: Exemplo de tela do Sistema e-Termos.

Observe que os usuários podem acessar qualquer Etapa de trabalho clicando nas abas localizadas na parte superior da tela. Veja que o nome e a descrição de cada Etapa é mostrado no balão amarelo com o simples repousar do *mouse* sobre a aba. No exemplo, a Etapa Principal é mostrada escrita em azul no centro superior, juntamente com nome do projeto e o perfil o qual está representando localizado no canto superior esquerdo da tela. Note também que o sistema oferece um menu com várias opções e funcionalidades que podem ser acessadas a qualquer momento. Conforme será visto adiante, cada Etapa possui um menu específico, de acordo com

as tarefas que disponibiliza. Já na parte central maior, situa-se a área de trabalho, neste caso mostrando os status atuais de cada Etapa do projeto e se existe ou um novo Recado para o usuário autenticado. No lado esquerdo da tela é mostrada uma lista com todos os membros da equipe junto com a indicação se estão on-line ou off-line no sistema em um determinado momento (por meio das bolinhas vermelhas e verdes). Especificamente, esta parte é comum a todas as Etapas e é partir dela que se inicia o bate-papo instantâneo entre os membros.

É sobre o sistema **e-Termos** que o restante deste capítulo é dedicado, concentrando-se principalmente na apresentação das suas características e descrição das funcionalidades.

4.2.1 Perfil dos usuários do **e-Termos**

A terminografia é caracterizada por um trabalho interdisciplinar. Assim, estudar a linguagem e os textos de uma determinada área do conhecimento sempre envolverá a colaboração de vários profissionais, principalmente os especialistas e os linguistas. Dessa maneira, o trabalho terminológico promove uma relação de troca entre os especialistas da linguagem e os especialistas do domínio, favorecendo a competência textual do lingüista em uma determinada área, bem como a reflexão do especialista sobre a feição e a importância da própria linguagem.

No **e-Termos** estão previamente definidos quatro perfis de usuários. Cada perfil possui uma visão específica do Ambiente, que varia de acordo com a atividade que desenvolve. Estes perfis são:

Gerente de Projeto : é o usuário criador e gestor de um projeto terminológico. É responsável por gerenciar o projeto e suas etapas, bem como criar e coordenar a equipe de trabalho (aprovar um usuário, convidar, incluir, excluir, etc.), embora algumas atividades ocorram fora do seu controle, como é o caso da colaboração natural entre os membros da equipe. Na maioria dos casos, o gerente de projeto é o próprio terminólogo/lingüista, pois é ele quem administra projetos e orienta sua equipe de trabalho durante todas as etapas de desenvolvimento de um produto. Entretanto, um especialista de um dado domínio pode exercer esse papel se o produto final for, por exemplo, uma estrutura conceitual (ontologia) de sua área de especialidade.

Especialista do Domínio : é o usuário representante da área de conhecimento que é objeto de trabalho da Terminologia. É responsável por várias atividades, como, por exemplo, pelo julgamento dos candidatos a termos quanto a sua relevância, pela validação e correção da ontologia, e pela verificação dos dados constantes da Ficha Terminológica e/ou no verbete final;

Terminólogo e/ou Linguista : é o responsável por todo o processo de criação do produto

terminológico e participa de todas as etapas do trabalho. É ele quem faz, por exemplo: 1) a compilação do corpus de especialidade a ser utilizado e, consequentemente, sua análise e utilização, tanto na extração automática de candidatos a termos como na edição da Ficha Terminológica, 2) o preenchimento da Base Definicional, 3) a avaliação das listas de candidatos a termos (dado que pode excluir palavras e siglas reconhecidamente da língua geral), 4) a observação e crítica das listas intermediárias de termos, 5) a relação e contato com os especialistas, 6) a redação da definição e a maioria das tarefas inerentes ao processo de preparação do produto terminológico;

Usuário Final (consulente) : é o usuário final, consulente e consumidor dos produtos terminológicos criados no Ambiente. Pode ser entendido como o “cliente” que interage com o **e-Termos** para consulta e/ou utilização do produto terminológico final como um recurso de suas pesquisas em outro sistema computacional ou de gestão terminológica.

Outro usuário importante do Ambiente é o Administrador do Sistema, responsável por sua administração e manutenção. São suas atribuições a configuração e monitoramento do sistema, o gerenciamento do banco de dados, a liberação e bloqueio de usuários, a definição das visões do Ambiente, a produção das estatísticas de acesso e o *back-up* dos dados armazenados.

Independente do perfil, o **e-Termos** foi desenvolvido para atender as necessidades desses usuários, propondo soluções automatizadas ou semi-automatizadas para eliminar ou amenizar os problemas e dificuldades presentes em cada etapa do trabalho terminológico, de maneira que a eficiência de suas atividades seja garantida.

4.2.2 O ambiente modular

Fruto da aliança entre a lingüística computacional, a lingüística de corpus, a informática e a terminografia, o **e-Termos** é composto por seis módulos de trabalho, que juntos cobrem todas as etapas do processo de criação e gestão dos produtos terminológicos. Neste contexto, a finalidade de cada módulo é empreender as tarefas inerentes a cada etapa do trabalho terminográfico vistas no Capítulo 3.

A estrutura macro **e-Termos** pode ser dividida em duas partes: 1) **Parte Administrativa** e 2) **Parte Operacional**. A **Parte Administrativa** consiste num conjunto de programas e recursos computacionais voltados para o completo gerenciamento do Ambiente, envolvendo todas as tarefas pelas quais o usuário administrador é responsável.

A **Parte Operacional** é a parte mais importante do **e-Termos**, pois acolhe os seis módulos de trabalho, comportando todas as atividades de elaboração e gestão do produto terminológico a partir da interação colaborativa dos usuários. Desse modo, cada módulo definido possui: 1) funcionalidades específicas a serem oferecidas para suportar as tarefas dos usuários em cada etapa;

2) interfaces dedicadas a diferentes usuários para a execução de suas tarefas e da própria colaboração, 3) padrão de entrada e saída de dados bem definidos, e 4) quando relevante, ferramentas de PLN adicionais que servem de suporte as suas atividades.

É importante salientar que a **Parte Operacional** corresponde à grande maioria do trabalho e da pesquisa deste doutoramento, já que ela compreende os aspectos relevantes e as motivações deste trabalho, principalmente os relacionados ao desenvolvimento terminológico e sua automatização. Apesar da **Parte Administrativa** ser de suma importância, contém apenas as ferramentas de “administração”, sem necessariamente possuir as interfaces gráficas e Web para as mesmas, já que o único usuário desta parte é o administrador, geralmente um profissional da área de informática, conhecedor dos comandos de controle de administração de um sistema. A Figura 4.3 apresenta o esquema geral modular da **Parte Operacional** do **e-Termos**. Na Seção 4.5.1 sua arquitetura computacional será mais detalhada.

Figura 4.3: Esquema geral modular da **Parte Operacional** do **e-Termos**.

Observando a Figura 4.3, cada caixa rotulada com a letra maiúscula “M”, seguida de um número decimal, representa um Módulo de trabalho. Apesar de estarem dispostos de maneira sequencial, as execuções dos módulos podem acontecer de maneira independente e, devido à sua característica colaborativa, diferentes usuários podem interagir com diferentes módulos ao mesmo tempo, dependendo somente da tarefa que cada um desempenha. Por exemplo, enquanto um Editor da Ficha trabalha no Módulo 5, outro usuário, como o Terminólogo, pode estar no Módulo 1, compilando cópus de outro produto, ou um consultante, pode estar no Módulo 6 consultando produtos finalizados. A caixa rotulada de “MP” representa o *Módulo Principal* do sistema. De forma resumida, cada Módulo de trabalho é responsável pelas seguintes atividades:

Módulo Principal : este módulo é responsável pela gestão terminológica do trabalho e contém funcionalidades e ferramentas que permitem o controle e a gestão: 1) do projeto terminológico, 2) da equipe de trabalho, 3) das etapas de trabalho e 4) de usuários e perfis, funcionando como componente essencial para a gerência;

Módulo 1 – Compilação Automática de Corpus : este módulo é responsável pela compilação automática do corpus de especialidade. É responsável pelo estabelecimento das relações de confiança inter-sistemas, principalmente aqueles que representam a busca, distribuição e compartilhamento de corpus. Corresponde à primeira etapa do trabalho terminológico;

Módulo 2 – Compilação e Suporte para Análise de Corpus : módulo responsável pela compilação e análise quantitativa e qualitativa do corpus de especialidade. Oferece um conjunto variado de ferramentas lingüísticas que auxiliam o terminólogo a avaliar a qualidade do corpus. Também serve para a compilação de corpus feita pelo próprio usuário (sem passar pelo Módulo 1), através das ferramentas de inclusão e junção de textos. Corresponde à segunda etapa do trabalho terminológico;

Módulo 3 – Extração Automática de Candidatos a Termos : é o módulo responsável pela extração de candidatos a termos a partir dos corpus de especialidade compilados nos Módulos 1 e/ou 2. Embora tenha a capacidade para a expansão no nível da análise lingüística, atualmente, o **e-Termos** disponibiliza o método estatístico de extração automática de termos. Corresponde à terceira etapa do trabalho terminológico;

Módulo 4 – Edição do Mapa Conceitual e Categorização de Termos : este Módulo abriga as ferramentas de criação, edição e visualização da ontologia da área de especialidade, além dos recursos computacionais para a categorização dos termos na ontologia e avaliação pelos usuários especialistas dos termos candidatos na estrutura ontológica. Também abriga as funcionalidades de controle e manutenção das *Relações Conceituais* específicas de cada projeto. Corresponde à quarta etapa do trabalho terminológico;

Módulo 5 – Criação e Gerenciamento de uma Base de Dados Terminológica : este módulo é responsável pela criação e preenchimento da Ficha Terminológica e da elaboração da Base Definicional que serve de suporte para a redação das definições terminológicas. Neste contexto, o Módulo 5 abriga um conjunto de ferramentas cuja finalidade é a gerência de uma base de dados que representa os produtos terminológicos que estarão disponíveis. Corresponde à quinta etapa do trabalho terminológico;

Módulo 6 – Edição dos Verbetes e Intercâmbio dos Produtos Terminológicos : módulo responsável pela edição dos verbetes e pela difusão, intercâmbio e consulta dos produtos terminológicos finalizados e disponíveis no **e-Termos**. Para tal, agrupa um conjunto de ferramentas de exportação de dados terminológicos para os fins de disseminação, bem como interfaces para edição e consulta dos verbetes armazenados na base terminológica.

Também é responsável pelas tarefas de pré-publicação e publicação do conteúdo no portal do **e-Termos**. Corresponde à sexta etapa do trabalho terminológico.

Observando ainda a Figura 4.3, importa dizer que as ferramentas colaborativas estão sempre disponíveis, em todo o sistema e em cada módulo, não havendo, portanto, disponibilização específica. Isso porque as ferramentas colaborativas são fundamentais para garantir o processo de colaboração dos membros da equipe de trabalho e devem sempre estar facilmente acessíveis em toda atividades do sistema. Já as ferramentas lingüísticas também estão contempladas em todos os módulos, sendo, entretanto, pertinentes a cada um deles, de acordo com as atividades e tarefas desempenhadas no módulo. Assim, por exemplo, as ferramentas de análise da qualidade do corpus estão contempladas no Módulo 2 do sistema, os extratores de termos no Módulo 3, a categorização dos termos no Módulo 4, a consulta de termos no corpus no Módulo 5, etc. A Tabela 4.1 apresenta uma lista de ferramentas de PLN atualmente implementadas e disponibilizadas no **e-Termos** e seus respectivos módulos de vinculação.

Tabela 4.1: Conjunto de ferramentas de PLN disponíveis no **e-Termos**

Módulo	Nome da Ferramenta	Referência
2	Contadores de frequência de palavras	Aluísio et al. (2003)
2	Contadores de frequência de uma única palavra ou expressão	Aluísio et al. (2003)
2	Concordanceadores	Aluísio et al. (2003)
2	Identificação e separação de lexias simples e complexas	Aluísio et al. (2003)
2, 5	Identificação e recuperação de termos do corpus	—
3	Extrator Automático de Termos (estatísticos)	Teline (2004)

Vale ressaltar, entretanto, que esta lista não é exaustiva, de maneira que outras ferramentas ou novos recursos informatizados podem ser adicionados no futuro, principalmente as mostradas na Tabela 4.2.

Tabela 4.2: Conjunto de futuras ferramentas de PLN

Módulo	Nome da Ferramenta	Referência
2	Segmentadores Sentenciais	Disponíveis no NILC
2	Lematizadores	Disponível no NILC Os lemas (ou formas canônicas de substantivos, verbos e adjetivos) podem ser obtidos do dicionário para o português no formato DELA, gerado pelo projeto Unitex-PB
2, 5, 6	Corretores gramaticais (identificação de erros ortográficos e sintáticos <i>parsers</i>)	Nunes e Junior (2000)
5	Editor de definição (tipologia de definição)	Aluísio (1995)
2, 5	Alinhadores de palavras e sentenças	Silva (2004), Caseli et al. (2005), Caseli (2003)
5	Extratores de Definição	Pinto e Oliveira (2004)

A seguir apresentamos uma explicação mais detalhada de cada ferramenta e o caso de uso de cada uma delas.

Definição e Caso de Uso das Ferramentas Lingüísticas

Contador de Frequência : programa ou ferramenta que calcula a frequência de palavras de um determinado corpus ou texto. A frequência de palavras pode ser entendida como a quantidade de vezes que determinada palavra aparece em um corpus/texto. Tem como entrada um corpus ou texto e como saída uma lista de palavras e suas respectivas frequências, podendo ser ordenada alfabeticamente ou pela frequência. É muito utilizada para análise da qualidade do corpus, principalmente para recuperação dos valores que corresponde aos critérios externos;

Contador de Frequência por Palavra ou Expressão : é um contador de frequência para uma única palavra ou expressão pertencente ao corpus ou texto. Também utilizado na análise de corpus e pode ser útil para se saber especificamente a frequência de uma determinada palavra ou a posição (*ranking*) que ela aparece no corpus;

Concordanceador : é uma ferramenta que gera concordâncias. Em PLN, concordância é uma listagem das ocorrências de uma determinada palavra ou expressão, disposta de tal modo que a palavra de busca (aquela que se tem interesse em investigar) aparece em evidência acompanhada do seu contexto original, isto é, das palavras que ocorrem ao seu redor no conteúdo do corpus Sardinha (2004). As concordâncias de um corpus também são úteis para análise da qualidade de um corpus, principalmente a avaliação de critérios internos;

Identificador de Lexias Simples e Complexas : é um programa baseado e uma regra de formação de palavras que identifica as lexias simples e complexas. Uma lexia complexa pode ser entendida como uma sequência de unidades lexicais que ocorrem lado a lado com forte índice de coesão, cujo sentido depende da totalidade do conjunto no contexto do discurso. A lexia complexa também é utilizada na análise de corpus, principalmente para identificação de nomes próprios ou termos compostos;

Identificador de Termos : consiste em uma ferramenta cuja função é identificar termos em um corpus ou texto. Não pode ser confundido com os extratores de termos, pois neste caso, o usuário já sabe o(s) termo(s) a ser(em) buscado(s). Também pode ser utilizado na análise de corpus, principalmente para uma avaliação prévia, antes da extração de candidatos a termos, para verificação se alguns termos comuns da área estão presentes no corpus;

Etiquetador (tagger) : programa que identifica e insere automaticamente em um corpus etiquetas que indicam a classe gramatical de cada palavra presente nos textos. O corpus etiquetado pode ser usado para análise de corpus, principalmente para identificação dos

padrões lingüísticos dos termos e para a extração automática de candidatos a termos pelo método lingüístico ou híbrido;

Segmentadores Sentenciais : programa que “quebra” um corpus ou texto em uma sequência de sentenças. Uma sentença pode ser definida como uma sequência de palavras entre dois sinais de ponto final “.”, retirando desse conjunto, as ocorrências das siglas e abreviações que ocorrem no texto. É muito utilizado pelos etiquetadores, nas tarefas de pré-etiquetagem;

Lematizador : programa que identifica a forma canônica, ou lema, de uma determinada palavra, agrupando duas ou mais formas diferentes numa mesma unidade lexical. Por exemplo, as ocorrências do verbo “*defendem*” e “*defenderam*” são agrupados sob o lema “*defender*”. O lematizador pode ser usado para lematizar um determinado corpus cuja extração de candidatos a termos pode ser melhorada;

Corretor Gramatical : programa cujo objetivo é corrigir um texto em nível ortográfico (*spell checker*) e sintático (*parser*). Este programa é muito útil para tarefa de redação das definições dos termos, e na maioria das vezes evita erros de gramaticais e sintáticos do Português. Ganha destaque nessa categoria de programas o corretor gramatical **ReGra** Nunes e Junior (2000), desenvolvido no NILC, e acoplado em 2000 ao Pacote Office da *Microsoft* e implementado para o ambiente Web no projeto SciPo¹ Antikeira et al. (2003);

Editor de Definição : programa de auxílio à autoria da redação de definições de domínios técnicos. Este programa mantém um conjunto de tipos definicionais que podem ser utilizados no momento da tarefa de definir um determinado termo e pode ser muito útil na tarefa de redação das definições no Módulo 5 do **e-Termos**;

Alinhador de Palavras e Sentenças : programa que estabelece correspondências entre sentenças (alinhador sentencial) ou palavras/multipalavras (alinhador lexical) em textos paralelos². Mais especificamente, um alinhador de sentenças determina qual(is) sentença(s) no texto fonte é(são) tradução de qual(is) sentença(s) no texto alvo, enquanto que um alinhador de palavras/multipalavras determina qual(is) palavra(s) no texto fonte é(são) tradução de qual(is) palavra(s) texto alvo. É muito útil para a tarefa de preenchimento da Ficha Terminológica, caso seja necessário a busca de um determinado termo em outra língua;

Extrator de Definição : programa que identifica e extrai candidatos a definição de um determinado termo pertencente ao corpus ou texto. A partir de padrões definitórios, como por

¹<http://www.nilc.icmc.usp.br/nilc/projects/scipo.htm>

²Textos paralelos são textos que possuem uma relação de tradução, ou seja, um texto fonte que possui uma ou mais traduções, chamados de textos alvo.

exemplo: “é um”, “é definido como”, “consiste em”, “pode ser entendido”, etc. este extrator pode indicar uma possível definição do termo, presente no corpus. Essa ferramenta é muito utilizada na elaboração da Base Definicional no Módulo 5.

Todavia, importa salientar que pode haver ciclos de execução entre os módulos, conforme mostrado na Figura 4.3 pelos arcos azuis. Isto é possível porque o trabalho terminológico pode iniciar com um corpus de especialidade (Módulo 1 ou 2) que atenda a um mapa conceitual inicial do domínio, e a partir extração de candidatos a termos (Módulo 3) e a sua subsequente avaliação (Módulo 4) pode indicar a necessidade de um detalhamento maior deste mapa, o que, por sua vez, acarretará o incremento do corpus. Particularmente neste aspecto, a pesquisa de Pardo (2004) fundamenta a criação de corpus de especialidades através de uma metodologia cíclica baseada no refinamento do mapa conceitual e do corpus que o apóia. O Módulo 1 participa da execução cíclica porque o corpus de especialidade pode ser inicialmente compilado automaticamente, pelo fornecimento de palavras-chaves que representam os campos nocionais do mapa conceitual inicial.

É necessário registrar ainda os benefícios que esse projeto pode receber do laboratório NILC. Durante os 17 anos de existência do NILC (desde 1993), várias ferramentas de PLN foram desenvolvidas. Por exemplo, no Projeto Lácio-Web³ foram criados: contadores de frequência de palavras que fazem um tratamento de lexia complexa, contadores de frequência de uma única palavra ou expressão, Concordanceadores, etiquetadores (*taggers*); no Projeto *ExPorTer*⁴ foram avaliados e disponibilizados quinze métodos de extração de candidatos a termos sob a luz das três abordagens: estatística, lingüística e híbrida; no Projeto ReGra⁵ foi criado um corretor gramatical do português do Brasil que já foi incorporado em outros projetos do NILC, como por exemplo, o SciPo; nos Projetos PESA⁶ e PEWA⁷ foram criados alinhadores de sentença e de palavras, respectivamente, a partir de corpus paralelos. Outras ferramentas como segmentadores sentenciais, lematizadores e identificadores de neologismos também estão disponíveis ou em desenvolvimento no NILC.

4.3 Descrição dos módulos de trabalho do **e-Termos**

Baseando-se na Figura 4.3, nas subseções seguintes, além da apresentação daquilo que já foi implementado, serão descritas todas as propriedades, características e funcionalidades dos sete módulos de trabalho pertencentes ao **e-Termos**.

³<http://www.nilc.icmc.usp.br/lacioWeb/>

⁴<http://www.nilc.icmc.usp.br/nilc/projects/termextract.htm>

⁵<http://www.nilc.icmc.usp.br/nilc/projects/regra.htm>

⁶<http://www.nilc.icmc.usp.br/nilc/projects/pesa.htm>

⁷<http://www.nilc.icmc.usp.br/nilc/projects/PEWA.htm>

4.3.1 Módulo Principal - Gestão terminológica

Dito no próprio nome o Módulo Principal é a base do sistema e concentra as funcionalidades da gestão terminológica e dos aspectos colaborativos, principalmente no abrigo das ferramentas de comunicação disponíveis no Ambiente. Conforme pode ser visto na Figura 4.2, embora exista uma divisão particular para o módulo principal, representada pela aba “*Principal*” no canto superior direito da tela, suas tarefas e funcionalidades estão presentes em todo sistema, indo desde o cadastro e acesso autenticado do usuário (realizado na página do Portal) até as ferramentas de suporte a colaboração, sempre disponíveis em todos os módulos.

4.3.1.1 Gestão terminológica

A gestão terminológica no **e-Termos** acontece a partir da simbiose de 3 componentes: 1) o projeto, 2) a equipe e 3) as etapas de trabalho. Dessa maneira, a criação de um determinado produto terminológico é estabelecida a partir da ocorrência simultânea desses fatores, ou seja, todo produto é fruto de um projeto que possui uma equipe de trabalho que perfaz uma seqüência de etapas e suas tarefas. Nesse contexto, o “projeto” pode ser entendido como uma área de trabalho oficial de elaboração do produto na qual os membros de uma determinada equipe atuam. A equipe, por sua vez, é composta por usuários necessariamente cadastrados no sistema que pertencem a um determinado perfil e pode executar tarefas específicas em cada etapa. As etapas são um conjunto de atividades que elaboram um produto terminológico.

Contudo, embora exista a obrigatoriedade de uma equipe legitimamente formada, não há uma determinação exigindo que toda equipe possua membros que representam todos os perfis de usuários. A única exigência do sistema é que toda equipe tenha pelo menos 1 membro e que este membro seja necessariamente um *Gerente de Projeto*. Por outro lado, também não há limite de membros e nem perfis numa determinada equipe, podendo possuir quantos membros forem necessários sob diferentes perfis.

Da mesma forma não há obrigatoriedade que todo projeto terminológico execute todas as etapas de trabalho para elaboração de um produto, havendo, portanto, uma flexibilidade em sua execução. A flexibilidade no **e-Termos** pode ser classificada em 3 níveis: 1) de Etapa, 2) de funcionalidades e 3) de ferramentas. A flexibilidade em nível de etapa corresponde a possibilidade de não executar uma determinada etapa durante a elaboração de um produto, ocorrência em que o Gerente (ou a equipe) ignora as tarefas e atividades de uma determinada etapa passando para a próxima. Um exemplo desta possibilidade é a decisão de não executar a Etapa 1 (Módulo 1) de compilar cópulas automaticamente. A flexibilidade em nível de funcionalidade é a mais comum no sistema e corresponde a liberdade da equipe em não executar determinadas funções do sistema, simplesmente por achar que não é necessário para o projeto em questão. Exemplo disso seria a

não inclusão de *relações conceituais específicas* (visto na Seção 4.3.5) ou não preenchimento da Base Definicional (visto na Seção 4.3.6.3). Já a flexibilidade de ferramentas é o nível mais baixo e corresponde a decisão da equipe de não usar uma determinada ferramenta, seja lingüística, colaborativa ou de apoio. Assim, um determinado membro da equipe pode decidir não utilizar o *identificador de termos* ou não *enviar recados* para seus colegas.

Essa flexibilidade é uma das principais vantagens do **e-Termos** pois o torna um sistema sensível a outras metodologias de trabalho e criação de produtos terminológicos ou lexicográficos, podendo atingir nichos diferentes de pesquisa e habilitando-o para receber variadas equipes de trabalho que possivelmente podem não seguir ou executar todas as etapas, funcionalidades ou ferramentas.

Dessa maneira, a gestão terminológica corresponde ao controle e a gerência dos seguintes aspectos: 1) de usuários e perfis, 2) de projetos, 3) de equipes e 4) das etapas de trabalho. A gerência de usuários e perfis corresponde ao controle e manutenção do cadastro e do acesso dos usuários, bem como das permissões de cada perfil no sistema. As funções de cadastro e o acesso estão localizados na página principal do portal do **e-Termos** pelo link *Cadastre-se!* e pela área de acesso, respectivamente. Já a verificação das permissões dos perfis é uma função “invisível” e está disponível em todo o sistema, em cada funcionalidade, pois é a partir dela que se organiza e inspeciona o quê cada perfil pode executar durante o trabalho.

A gerência de projetos engloba as funções de controle dos projetos criados e mantidos pelo sistema, e correspondem às funcionalidades de abertura e fechamento⁸ de um projeto por um determinado membro da equipe e a atribuição de status do mesmo, podendo ser: *Ativo, Finalizado, Interrompido ou Suspenso*. Importa ressaltar ainda que todo projeto é criado a partir da formulação de uma “*proposta de projeto*”, elaborada por um determinado usuário e julgada por uma equipe técnica.

A gerência de equipe corresponde as atividades desempenhadas pelo *Gerente de Projeto* no controle de sua equipe de trabalho. As ações desse âmbito vão desde o envio de um convite de participação a novos membros até o bloqueio e/ou desbloqueio de um determinado usuário-membro. Além disso, o sistema mantém um registro das tarefas desempenhadas por cada usuário, bem como as interações colaborativas dos membros de uma equipe. Porém, vale frisar que neste caso, somente o próprio usuário-membro tem acesso às suas interações (conversas no chat, e-mails e recados), garantindo a privacidade.

Por sua vez, a gerência das etapas de trabalho corresponde ao controle de execução das etapas de um determinado projeto e envolve a atribuição de diferentes status às etapas, podendo ser: *Liberada, Bloqueada e Finalizada*, permitindo que o usuário com perfil de *Gerente* possa

⁸Neste caso “abertura” quer dizer a entrada de um determinado membro da equipe na área de trabalho do projeto, e “fechamento”, a saída.

controlar o tempo e execução das atividades. Esta possibilidade também é primordial para a natureza colaborativa, pois possibilita a fixação de *pontos de controle (check-points)* na evolução do trabalho.

4.3.1.2 Suporte a colaboração

Como parte da implementação do Módulo Principal está o suporte a colaboração representada pelas possibilidades de edição colaborativa e as ferramentas de comunicação síncrona e assíncrona disponibilizadas no Ambiente.

A edição colaborativa no **e-Termos** corresponde à permissão de uso de uma mesma ferramenta ou funcionalidade de maneira concomitante e a edição ou alteração de dados ou recursos por diferentes usuários simultaneamente. A utilização concomitante de ferramentas é natural do ambiente Web e corresponde ao fato de dois ou mais usuários utilizarem a mesma funcionalidade ao mesmo tempo. Por outro lado, a edição simultânea pode ser síncrona, quando um usuário pode visualizar o que outro faz em tempo real, ou assíncrona, quando os usuários editam os recursos de maneira individual, e depois visualizam as alterações. Nesse sentido, a edição de dados disponibilizada no sistema é assíncrona, baseada no conceito das operações de trancamento de recursos (*lock*), não executando, portanto, nenhum tipo de merge de alterações.

Dessa forma, a edição colaborativa está presente em vários módulos do sistema, desde a compilação automática e simultânea de corpus (vista na Seção 4.3.2) até a tarefa de pré-publicação do produto, explicada na Seção 4.3.7.

Também vinculado ao suporte colaborativo estão as ferramentas comunicacionais que permitem interação entre os membros de um determinada equipe de trabalho e/ou entre todos os usuários do **e-Termos**. Tais ferramentas são onipresentes no Ambiente e podem ser acessadas de qualquer módulo do sistema, independente da tarefa ou funcionalidade a ser utilizada, e se dividem em duas categorias de comunicação: 1) as assíncronas e 2) as síncronas.

As assíncronas são representadas pelas ferramentas de *e-Mails Categorizados*, *Recados* e o *Fórum de Discussão*. Destes, os dois primeiros são utilizados apenas no âmbito de cada projeto, estando disponível apenas para a comunicação entre os membros de uma mesma equipe de trabalho. De outro modo, o *Fórum de Discussão* está disponível para todos os usuários do **e-Termos**, independentemente de projeto, servindo como um espaço único e público para compartilhamento de experiências, troca de informações e auto-ajuda. Já a ferramenta de comunicação síncrona é representada por um *Chat Instantâneo* na qual os usuários membros de uma mesma equipe podem bater papo em tempo real por meio da troca de mensagens *on-line*. Observe na Figura 4.2 que tais ferramentas estão disponíveis pelos menus *Mail*, *Recado* e *Fórum* (assíncronas) e o *Chat* na tabela do lado direito da tela, mostrando em tempo real os usuários

on-line (bolinha verde) e *off-line* (bolinha vermelha) em um determinado momento.

O aspecto colaborativo do **e-Termos** e as ferramentas de comunicação serão vistas em detalhes adiante neste capítulo na Seção 4.4.1.

4.3.2 Módulo 1 - Compilação automática de corpus

Conforme visto no Capítulo 3, o uso de corpus na área da lingüística computacional, envolvendo Processamento de Língua Natural (PLN) no âmbito da Tradução, da Lexicografia e da Terminologia tem crescido progressivamente. Atualmente, o conceito de corpus tem sido usado como sinônimo de “corpus eletrônico”, isto é, aqueles em que os dados textuais são armazenados eletronicamente, podendo ser processados por computador e servir para vários propósitos.

No mesmo compasso de desenvolvimento, a criação de grandes corpus, constituídos de vários milhões de palavras, tem ganhado consideráveis esforços nos últimos anos. Estes esforços são direcionados tanto para os fins genéricos de pesquisa de corpus como para posterior disponibilização dos mesmos, seja em formato de CD-ROM ou via Web. Como exemplo deste tipo de iniciativa, podemos citar os corpus BNC⁹ e ANC¹⁰ para a língua inglesa nas variantes britânica e americana, respectivamente, e o CETEMPúblico¹¹ e Lácio-Web¹², para a língua portuguesa nas variantes européia e brasileira, respectivamente.

Embora já exista um número razoável de corpus genéricos (ou de referência, como são chamados) para várias línguas, a quantidade de corpus de especialidade disponíveis para suporte à pesquisa terminológica ainda é deficiente. Essa deficiência dá-se pela própria especificidade de tais corpus que são muitas vezes construídos para serem utilizados por um período curto de tempo, durante um único trabalho, levantando o questionamento sobre o custo de investimento de grandes esforços para sua compilação. Por outro lado, a Internet atualmente é uma importante fonte de recursos lingüísticos e é considerada por alguns autores como o maior corpus disponível Kilgariff e Grefenstette (2003).

Para atender a necessidade específica da criação de corpus de especialidade, existem atualmente diversas ferramentas que auxiliam a compilação automática, tanto como resultado de pesquisas em corpus de referência como a partir de consultas a um motor de busca na Web. É neste espaço que o primeiro o módulo de trabalho do **e-Termos** se encaixa, sendo responsável pela tarefa de construção automática de corpus de especialidade para os propósitos da pesquisa terminológica.

Conforme visto na Seção 4.3.2 uma das maneiras de compilar corpus automaticamente pela Web é utilizar as ferramentas de *BootStrap*. Entretanto, devido às desvantagens encontradas

⁹<http://www.natcorp.ox.ac.uk/>

¹⁰American National Corpus (<http://americannationalcorpus.org/>)

¹¹<http://www.linguateca.pt/>

¹²<http://www.nilc.icmc.usp.br/lacioWeb/>

neste método, a saber, a dificuldade de conhecer a origem dos textos e o número de resultados falsos-positivos encontrados, este método não foi desenvolvido neste Módulo.

Em contrapartida, para atender a necessidade de compilação automática, foi estabelecida a relação inter-sistemas entre o *e-Termos* e o *Portal de Córpus*¹³, um repositório de textos livre e gratuito na Web componente do **Projeto PLN-Br**¹⁴.

4.3.2.1 O *Portal de Córpus* do **Projeto PLN-Br**

O projeto Recursos e Ferramentas para a Recuperação de Informação em Bases Textuais em Português do Brasil (PLN-BR) tem como principal objetivo a criação de um espaço interinstitucional para o intercâmbio de práticas de análise e pesquisa lingüístico-computacional sobre a representação e recuperação de informações de natureza semântica e pragmático-discursivo veiculada por enunciados produzidos no português do Brasil Bruckschen et al. (2008).

O projeto está subdividido em 7 subprojetos relativamente autônomos que compartilham a idéia do tratamento da informação localizada em um mesmo *cópus*, e envolve pesquisadores vinculados à Universidade de São Paulo (USP), *campus* de São Carlos; à Universidade Federal de São Carlos (UFSCar); à Universidade Estadual Paulista (UNESP), *campus* de Araraquara; à Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS); à Pontifícia Universidade Católica do Rio de Janeiro (PUCRJ); à Universidade do Vale do Rio dos Sinos (UNISINOS); e à Universidade Presbiteriana Mackenzie.

A arquitetura geral dos 7 subprojetos vinculados à produção de recursos (*corpus*, dicionário, thesaurus, bem como padrões e protocolos), à construção de ferramentas (etiquetadores, *tokenizadores*, lematizadores) e ao desenvolvimento de aplicações (de sumarização, de acesso ao *corpus*, de recuperação e extração de informações) está ilustrada na Figura 4.4.

Embora cada um dos subprojetos possua problemas que lhe são específicos, todos compartilham o mesmo ambiente interativo de desenvolvimento, sendo representado por: 1) um grande *cópus* e os dois outros derivados deste, 2) a construção colegiada de padrões e protocolos, e 3) a disposição para o reaproveitamento e a adaptação dos vários recursos já existentes desenvolvidos por várias instituições proponentes. Particularmente, este grande *corpus* do gênero informativo, subgênero jornalístico, chamado PLN-BR FULL gerou dois outros menores, chamados de PLN-BR CATEG e PLN-BR GOLD e motivou o desenvolvimento da aplicação **Portal de Córpus**.

Genericamente, o **Portal de Córpus** é uma aplicação de consulta a *cópus* e fornece uma estrutura básica para manter, disponibilizar e trabalhar com *cópus* em ambiente Web por meio de duas interfaces. A primeira é voltada para o usuário final, que pode acessar e criar sub*cópus* a partir de pesquisas que atendam a características específicas e particulares. A

¹³<http://www.nilc.icmc.usp.br:8180/portal/index.jsp>

¹⁴<http://www.nilc.icmc.usp.br/plnbr/>

Figura 4.4: Subprojetos vinculados ao Projeto PLN-Br.

base para montagem dos subcórpus são metadados armazenados no cabeçalho dos textos que compõem os córpus, criados e mantidos pelo **Editor de Cabeçalhos** Aluísio et al. (2007). Outra ferramenta disponível nesta interface é o **PEx-Corpus Tool** Muniz et al. (2007), uma adaptação do projeto **Projection Explorer (PEx)**¹⁵ que permite inspecionar visualmente um subcórpus para explorar o seu conteúdo e criar outros subcórpus com base numa seleção de tópicos relacionados. A segunda interface é administrativa e voltada para manipular o conteúdo do Portal, sem a necessidade de serviços manuais diretamente no servidor Aluísio e Muniz (2007).

Nesse contexto, a atual implementação do Módulo 1 corresponde a relação inter-sistemas existente entre o **e-Termos** e o **Portal de Córpus**. Nesta relação, o **e-Termos** passar a ser um cliente de “confiança” do **Portal de Córpus**, de maneira que os subcórpus montados possam ser automaticamente transferidos para o **e-Termos**, sem a interferência do usuário. Dessa maneira, é possível afirmar que cada usuário do **e-Termos** também é considerado um usuário indireto do **Portal de Córpus**, podendo utilizá-lo como fonte de pesquisas de textos para a compilação automática de córpus. A Figura 4.5 exemplifica esta relação.

Observe que a relação de confiança estabelecida é baseada numa autenticação segura (representado pelo cadeado amarelo) de modo que o usuário do **e-Termos** ganha uma seção de trabalho automaticamente no sistema **Portal de Córpus**. Esta seção é mostrada na Figura 4.6 que mostra o nome do usuário autenticado e o campo de pesquisa de textos. No exemplo foi

¹⁵<http://www.lcad.icmc.usp.br/paulovic/pex/>

Figura 4.5: Relação de confiança entre os sistemas **e-Termos** e **Portal de Córpus**.

usada a palavra “política” como chave de busca.

Figura 4.6: Tela de pesquisa de texto do **Portal de Córpus**.

O sistema então pesquisa textos que aparecem a palavra de busca e monta um subcorpus com as ocorrências encontradas, conforme mostrado na tela da Figura 4.7. Neste exemplo, foram encontrados 2 textos. Em seguida, se desejar, o usuário pode enviar o subcórpus de resultado para o **e-Termos**, que o recebe, verifica, calcula o número de palavras e o disponibiliza para uso, como pode ser visto na tela da Figura 4.8. Observe que os dois textos encontrados na pesquisa foram transferidos para o **e-Termos** e disponibilizados no Módulo 2 para posterior análise.

A possibilidade de estabelecer as relações de confiança com outros sistemas é outra vantagem de fundamental importância e caracteriza-se pela possibilidade singular de formação de parcerias e/ou acordos de cooperação entre sistemas diferentes, mas que possivelmente possuem algum ponto em comum. Além disso, tal possibilidade representa a utilização de ferramentas de terceiros de forma transparente incrementando a capilaridade do **e-Termos**.

Figura 4.7: Tela de resultado da montagem de subcórpus e de envio para o e-Termos.

Figura 4.8: Tela de recebimento de subcórpus do e-Termos.

4.3.3 Módulo 2 - Compilação e Suporte para Análise de Corpus

O estudo do texto, considerando a ocorrência e distribuição de elementos lingüísticos, tais como fraseologias, repetição de palavras, colocações, concordâncias, análise morfossintática entre outras, com o auxílio do computador, revelam o quão produtiva pode ser a investigação das particularidades e regularidades de um corpus de especialidade, seja o objetivo a descrição e análise do próprio corpus ou a criação de produtos terminológicos Krieger e Finatto (2004).

É no contexto do uso do computador para a análise textual, e consequentemente, dos corpus de especialidade, que foi inspirado o Módulo 2, tendo a função de empreender a tarefa de verificação da qualidade e relevância do corpus de especialidade, antes da submissão do mesmo aos extratores automáticos de termos (tarefa do Módulo 3).

Além oferecer um conjunto de ferramentas computacionais (automáticas) de PLN, úteis para avaliação dos critérios internos, este módulo também implementa condições de avaliação e julgamento da maioria dos critérios externos utilizados para julgar um determinado corpus quantitativa e qualitativamente. A disponibilização e implementação dos critérios internos e externos têm o objetivo de fornecer subsídios palpáveis de verificação, comparação e análise dos

mesmos, para facilitar o julgamento da qualidade de um *córpus* favorecendo a tarefa de avaliação empírica realizada pelo terminólogo.

Ferramentas lingüísticas como contadores de frequência, identificador de *N-grams*, Concordeadores, identificador de lexias complexas (sob a aplicação de um conjunto de regras específicas de formação de palavras), recuperação de contextos e termos entre outras, são de extrema relevância para esta análise, pois com elas é possível identificar e extrair várias informações específicas sobre o *córpus* para julgar sua relevância.

Todas as ferramentas citadas, se bem harmonizadas, podem ajudar a avaliar a qualidade dos *córpus*, tanto aqueles criados com ajuda do Módulo 1, quanto aqueles já compilados e trazidos para o ambiente do **e-Termos**. Na implementação atual do **e-Termos** a combinação das ferramentas citadas está disponível a partir do Módulo 2. Assim, por exemplo, ao aplicar um contador de frequência a um *córpus* podemos ver as palavras mais frequentes, o índice vocabular¹⁶ de um *córpus* (*token/type ratio*), as ocorrências dentro de um *córpus* que não são palavras da língua em questão (e podem se caracterizar como erro de digitação) e outras informações relacionadas.

Dessa maneira, o *córpus* de especialidade é a entrada principal do Módulo 2, e representa a matéria-prima das atividades desempenhadas neste módulo, e a saída, por sua vez, é o mesmo *córpus* de especialidade analisado e avaliado pelo(s) terminólogo(s).

4.3.3.1 A interface Web de compilação de *córpus* e aplicação de ferramentas lingüísticas

Nesse contexto, fora as ferramentas de comunicação e colaborativas, presentes em todos os módulos, a implementação atual do Módulo 2 pode ser subdividida em 3 partes: 1) a Gerência de Textos, 2) a Gerência de *Córpus* e 3) as ferramentas de análise lingüística.

A Gerência de Textos consiste num conjunto de funcionalidades que possibilita o controle e a manutenção de vários textos no sistema. Tais textos podem ser utilizados para compilar um determinado *córpus*, sempre de acordo com a necessidade ou desejo dos usuários. Desse modo, estabelece-se a regra que antes dos *córpus* existem os textos, de forma que todo *córpus* é formado por textos disponíveis no sistema. Nesse cenário, a partir dos textos, o usuário tem a possibilidade de criar quantos *córpus* desejar, sem limite de número de textos ou de palavras.

As tarefas da Gerência de Texto são o *upload* e a *exclusão* de textos e a *alteração de tipologia textual*. A tarefa de *upload* permite que os usuários enviem seus próprios textos para compor sua base textual para compilação de *córpus*. Isso é importante porque dá a liberdade do terminólogo utilizar seus próprios textos, já selecionados, sem a necessidade de encontrá-los pelo Módulo 1. Atualmente, o **e-Termos** suporta o envio das seguintes extensões de arquivos: *Texto Puro* (*.txt*),

¹⁶Índice Vocabular pode ser entendido como a riqueza vocabular de um texto ou *córpus*, e corresponde à razão entre a quantidade de palavras diferentes e o número total de palavras de um *córpus*.

Arquivo ZIP: (.zip – contendo apenas arquivos textos (.txt)), MS Word Document (.doc), Rich Text Format (.rtf), arquivos de HiperTexto (.html), Open Office Document Writer (.odt), e Portable Document Format (.pdf). Entretanto, por questões de padronização, todos os textos são automaticamente convertidos para *Texto Puro* (.txt). Por sua vez, a tarefa de exclusão permite que o usuário exclua um determinado texto, seja ele resultado da tarefa de *upload* ou da compilação automática (Módulo 1).

A *alteração de tipologia textual* permite que o usuário classifique o *gênero textual* e o *tipo de texto* de um determinado texto. Esta definição pode ser feita tanto no momento de *upload* quando na tarefa de gerência, e é importante porque possibilita a compilação de *córpus* específicos, sensíveis ao gênero e ao tipo de texto, facultando a realização de trabalhos mais especializados. A Figura 4.9 mostra a tela de Gerencia de Textos. Veja que cada texto disponível possui metadados particulares como o tamanho em Kbytes, o número de palavras, a data da última modificação, o usuário responsável pelo texto e a classificação de gênero e tipo textual, além da possibilidade de serem excluídos.

Veja abaixo os textos disponíveis neste Projeto. Para alterar a Tipologia Textual clique em **Alterar Tipologia**. Para excluir, selecione os textos desejados e clique em **Excluir Textos Selecionados**.

Textos Disponíveis:4

	Nome	Kbytes	Palavras	Modificação	Usuário	Gênero	Tipo Textual	Ações
☐	texto5.txt	362.26	61987	09/08/2009	Leandro Henrique	Científico	Artigo
☐	texto6.txt	244.88	42161	09/08/2009	Leandro Henrique	Informativo	Notícia
☐	texto7.txt	114.49	19757	09/08/2009	Leandro Henrique
☐	texto8.txt	146.20	25057	09/08/2009	Leandro Henrique

Figura 4.9: Tela de gerência de textos.

Uma vez inseridos no Ambiente os textos disponíveis podem ser compilados para o subsequente uso no trabalho terminológico. O *Gerenciador de Córpus* representa a parte de compilação de *córpus* do usuário e possui as seguintes tarefas: *compilação*, *exclusão* e *alteração de status*. A compilação de *córpus* consiste na seleção dos textos disponíveis no sistema para a formação de um *córpus*, que sempre deve ter uma identificação única. A Figura 4.10 mostra a compilação do *córpus* identificado como “cpUm” contendo os textos “texto5” e “texto7”. Conforme pode ser visto é possível especificar o idioma do *córpus* no momento da compilação.

Veja abaixo os textos disponíveis neste Projeto. Para alterar a Tipologia Textual clique em **Alterar Tipologia**. Para excluir, selecione os textos desejados e clique em **Excluir Textos Selecionados**.

Textos Disponíveis:4

	Nome	Kbytes	Palavras	Modificação	Usuário	Gênero	Tipo Textual	Ações
☒	texto5.txt	362.26	61987	09/08/2009	Leandro Henrique	Científico	Artigo
☐	texto6.txt	244.88	42161	09/08/2009	Leandro Henrique	Informativo	Notícia
☒	texto7.txt	114.49	19757	09/08/2009	Leandro Henrique
☐	texto8.txt	146.20	25057	09/08/2009	Leandro Henrique

e-Termos: novo Córpus

Informe abaixo a identificação do novo Córpus, selecione o idioma e clique em **Salvar Córpus**.

Identificação: Evite usar caracteres especiais, acentuados e numerais.

Idioma:

ATENÇÃO: campos em vermelho são obrigatórios.

Figura 4.10: Tela de compilação de *córpus*.

A tarefa de exclusão permite que o usuário elimine um determinado corpus criado anteriormente. Vale lembrar que a exclusão de corpus não implica na exclusão dos textos que o compõe. Já alteração de status permite a classificação do corpus em diferentes status que podem variar de: *Cancelado*, *Em Análise* e *Liberado*. Um corpus com status “*Cancelado*” significa que não será mais utilizado, por qualquer motivo que o usuário julgar pertinente. O status “*Em Análise*” aponta que corpus está na fase de análise de sua qualidade, na qual podem ser aplicadas todas as ferramentas linguísticas. Já o corpus “*Liberado*” define que um determinado corpus que será utilizado nas próximas etapas de trabalho.

A tela de Gerencia de Corpus está apresentada na Figura 4.11. Observe que cada corpus possui seus metadados de registro que são: a identificação única, o número de palavras, a data de modificação, o usuário responsável pela compilação, o gênero, o idioma e o status atual. Observe também que sobre cada corpus podem ser aplicadas duas ações, a *alteração de idioma*, representada pela letra maiúscula “A” e os *detalhes*, que mostra os textos pertencentes ao corpus, vista na parte amarelada na Figura 4.11.

Veja abaixo os corpus disponíveis neste Projeto. Para alterar o Idioma clique em **Alterar Idioma**. Para visualizar os textos que compõem um determinado corpus clique em **Detalhes**.

Identificação	Palavras	Modificação	Usuário	Gênero	Idioma	Status	Ações
corpus	959	27/11/2008	Dayse		Português-BR	Em Análise
informat	3128	27/11/2008	Dayse		Português-BR	Cancelado
cpUm	81744	09/08/2009	Leandro Henrique		Português-BR	Em Análise

e-Termos: Textos que compõem o Corpus.

Textos em ordem alfabética:

1 - texto5.txt

2 - texto7.txt

Total: 2

Clique e arraste para mover.

Figura 4.11: Tela de gerencia de corpus.

A última parte do Módulo 2 é dedicada a utilização das ferramentas de análise linguística com o propósito de avaliação dos corpus compilados. A disposição de tais ferramentas procura estar de acordo com os critérios externos de análise para permitir a apreciação e diagnóstico da qualidade do corpus de especialidade. As ferramentas atualmente disponíveis podem ser conferidas na Tabela 4.1.

Do exposto é possível resumir que a seqüência de trabalho deste Módulo obedece aos seguintes passos: primeiro são trazidos para o ambiente do **e-Termos** os arquivos- textos com os quais se deseja trabalhar. Depois, a partir desses, são compilados quantos corpus forem necessários, que por sua vez, podem ser submetidos ou manipulados pelas *Ferramentas* para aferir sua qualidade. Vale lembrar também que os textos desse Módulo podem ser originados do Módulo 1, após a execução da compilação automática.

Assim, paralelo a tarefa de compilação de corpus de especialidade, a função crucial do Módulo 2 é servir prioritariamente para avaliação destes corpus. Isso porque a qualidade do trabalho,

da pesquisa e consequentemente do produto terminológico final, dependem diretamente da qualidade do corpus em que está apoiado. Portanto, o Módulo 2 deve oferecer sempre que possível mais e melhores ferramentas lingüísticas que implementam os critérios de avaliação e julgamento de um determinado corpus.

4.3.4 Módulo 3 - Extração Automática de Candidatos a Termos

Uma vez que um determinado corpus de especialidade tenha sido compilado e avaliado no espaço do Módulo 2, ele está pronto para ser utilizado para a tarefa de extração de candidatos a termos, cuja responsabilidade é do Módulo 3 do **e-Termos**. Em linhas gerais, o Módulo 3 reúne e aplica vários e diferentes métodos para a extração automática de candidatos a termos, e sempre tem como entrada um corpus previamente compilado e/ou avaliado no Módulo 2. Dessa maneira, a principal função deste módulo é aplicar diversos métodos de extração automática de candidatos a termos para identificar e extrair automaticamente unidades ou conjuntos lexicais que podem constituir uma unidade terminológica (termo) bem como o armazenamento de forma estruturada desta informação.

Contudo, junto com os extratores automáticos, o Módulo 3 também implementa as funcionalidades que compreendem a gerência e o controle das *Listas de Termos* e *Stoplists*. Uma *Lista de Termos* no **e-Termos** corresponde a uma estrutura lógica que armazena um conjunto de candidatos a termos de um determinado projeto, de maneira que todos os termos envolvidos no processo de trabalho serão contemplados pelas Listas. Assim, várias *Listas de Termos* podem ser criadas e manipuladas no sistema, não havendo limite para a quantidade de Listas e termos envolvidos. No **e-Termos** a saída, ou os resultados das execuções dos extratores automáticos é uma *Lista de Termos*.

Nesse contexto, várias tarefas podem ser realizadas sobre as Listas, sendo elas: *upload*, *edição*, *consulta*, *concatenação* e *alteração de status*. A tarefa de *upload* permite que o usuário envie um conjunto qualquer de termos para o **e-Termos**, fazendo com que eles pertençam ao conjunto de termos envolvidos do projeto terminológico. Esta possibilidade é mais uma característica especial do sistema em direção à flexibilidade de tarefas e funcionalidades, já que admite que o usuário utilize outros termos, originários de outros métodos ou trabalhos, não se restringindo somente aqueles produzidos no próprio **e-Termos**. Além disso, tal fato também possibilita que um determinado projeto não utilize os extratores automáticos e, consequentemente, nenhum corpus, podendo, portanto, iniciar seu projeto já no Módulo 3.

A edição permite que o usuário edite o conjunto de termos de uma determinada Lista usando o próprio **e-Termos**, podendo neste caso, incluir, excluir ou alterar seu conteúdo. A tela da Figura 4.12 mostra uma tela de edição da Lista identificada como “listaTeste”.

Edição de Termos:

Seleção de Lista de Termos

Lista: listaTeste

Termos da Lista:

- tipos de argila
- argilas plásticas
- argila de cor branca
- ball clay
- fire clay
- flint clay
- argila de cor vermelha
- argilas refratárias
- argila refratária aluminosa
- argila refratária ou caulínica
- argila refratária sílico-luminosa
- argilas fundentes
- argilas de baixo conteúdo de carbonatos
- argilas de médio conteúdo de carbonatos
- argilas de elevado conteúdo de carbonatos
- argilas cauliniticas
- argilas ilíticas
- argilas montmoriloníticas
- argila não-plástica
- calcita
- feldspato - anortita
- sílica ou óxido de silício
- carbonatos
- talco
- água

Salvar

Figura 4.12: Tela de edição de uma Lista de Termos.

A tarefa de consulta é simples e possibilita que o conteúdo de uma determinada *Lista de Termos* seja observado. Por outro lado, a concatenação é uma tarefa especial e importante, pois permite que os conteúdos de duas ou mais Listas possam ser concatenadas, resultando em outra *Lista de Termos*. Essa funcionalidade é útil porque estabelece um bom nível de organização, já que capacita o usuário a manipular várias Listas de maneira individualizada, para em seguida juntá-las em outra. Um exemplo prático dessa funcionalidade seria, por exemplo, o usuário possuir listas de unigramas, bigramas e trigramas separadas e depois uni-las em uma “listaOficial”. A Figura 4.13 mostra um exemplo de concatenação das Listas identificadas como “listaISO” e “listaTeste” para a Lista “testeConcat”.

A alteração de status especifica a situação atual de uma determinada *Lista de Termos* que podem ser: *Cancelada*, *Em Edição* ou *Liberada*. O status “Cancelada” determina que por qualquer motivo uma determinada Lista não é mais válida não podendo ser editada ou concatenada. Em “Em Edição” indica que a Lista está em fase de alteração e correção de termos, podendo ser inclusive concatenada com outras Listas. Já o status “Liberada” significa que uma determinada Lista foi selecionada e está pronta para ser utilizada nas próximas etapas de trabalho.

Parecido com a ideia das *Listas de Termos* estão as *Stoplists*. As *Stoplists* são usadas na tarefa de extração automática de candidatos a termos e compreende em uma lista de palavras que devem ser ignoradas no processo de extração, assim como mencionado na Seção 3.2.3. Todavia,

Selecione as Lista de Termos as quais deseja concatenar e clique em **Concatenar Listas**.

The screenshot shows a web interface for managing term lists. At the top, there is a table titled 'Listas de Termos' with two columns: 'Identificação da Lista' and 'Status'. The table contains four rows: 'Nano_Tetragrama2', 'listaISO', 'listaTeste', and 'tstExtract', all with a status of 'Em edição'. Below the table is a button labeled 'Concatenar Listas'. Overlaid on the bottom right is a modal dialog titled 'e-Termos: nova Lista de Termos'. The dialog contains the instruction 'Informe abaixo a identificação da nova Lista de Termos e clique em Salvar.' and a text input field labeled 'Identificação:' with the value 'testeConcat'. At the bottom of the dialog are two buttons: 'Salvar' and 'Fechar'.

	Identificação da Lista	Status
☐	Nano_Tetragrama2	Em edição
☒	listaISO	Em edição
☒	listaTeste	Em edição
☐	tstExtract	Em edição

Concatenar Listas

e-Termos: nova Lista de Termos

Informe abaixo a identificação da nova Lista de Termos e clique em Salvar.

Identificação: testeConcat

Salvar Fechar

Figura 4.13: Tela de concatenação de Listas de Termos.

a vantagem aqui é que cada projeto do **e-Termos** pode ter suas próprias *Stoplists* de acordo com a especificidade do tema abordado pelo produto terminológico, sem limite de palavras.

Também neste caso, sobre cada *Stoplist* o usuário pode executar as seguintes tarefas: *upload*, *edição*, *consulta* e *alteração de status*. De forma semelhante, as tarefas de *upload*, *edição* e *consulta* permitem o envio, a edição/alteração e a consulta de uma determinada *Stoplist*, respectivamente. Já a alteração de status possibilita que o usuário especifique a situação em “Cancelada”, que determina que uma *Stoplist* deve ser ignorada não podendo ser utilizada ou editada, ou “Em Edição”, situação mais típica, que permite o uso e edição da *Stoplist*.

Como alternativa adicional o **e-Termos** oferece uma *Stoplist Padrão*, resultado do trabalho de Teline (2004) contendo palavras que geralmente são ignoradas no processo de extração. Entretanto, neste caso, o usuário não pode editar ou alterar seu status, podendo somente consultar ou utilizá-la em seus projetos.

A pesquisa sobre os extratores e métodos de extração automática de candidatos a termos não é recente, haja vista as contribuições de autores como Ananiadou (1994), Daille (1996), Daille (2003), Frantzy (1997), Bagot (1999), Bagot (2001) e Lino et al. (2004), e fazem parte do conjunto de trabalhos desenvolvidos no NILC desde 2002. A pesquisa de Teline (2004) é a principal representante desses trabalhos. Chamado de ExPorTer, conforme citado na Seção 3.2.3, um dos objetivos desse projeto foi a obtenção de um conjunto de informações significativas a respeito das vantagens e desvantagens de vários métodos de extração automática de candidatos a termos, contemplando as abordagens estatística, lingüística e híbrida. Nessa perspectiva, o levantamento bibliográfico realizado sobre os extratores automáticos, e como resultado, a análise e implementação dos métodos de três abordagens serviu de base para a implementação deste módulo.

Desde o início deste trabalho de doutoramento, a investigação e os experimentos sobre a extração automática de candidatos a termos têm sido freqüentes. Em virtude de tal fato, os

métodos de extração desenvolvidos no projeto ExPorTer foram avaliados em um corpus do Projeto Bloc-Eco Zavaglia et al. (2005) cujo objetivo foi a criação de uma base de conhecimento com informações ontológicas para termos em português da área de Ecologia. Nas seções a seguir apresentamos brevemente a experiência deste projeto e o estado atual de implementação do Módulo 3 do **e-Termos**.

4.3.4.1 Métodos de extração no trabalho *Bloc-Eco*

O Projeto *Bloc-Eco* avaliou especificamente métodos das três abordagens: estatística, lingüística e híbrida, estudadas no Capítulo 3, para o português. As medidas estatísticas utilizadas nesse trabalho foram quatro: Frequência Simples, *Log-likelihood*, *Informação Mútua* e *Coefficiente Dice*, implementadas no pacote para a extração de *N-grams* NSP (também visto no Capítulo 3 na Seção 3.2.3.1) com objetivo de eleger a melhor medida estatística para a extração automática de unigramas, bigramas e trigramas. O método lingüístico implementado baseia-se em expressões lingüísticas e indicadores estruturais, bem como nos padrões morfossintáticos dos termos de um dado domínio. Para isso, baseou-se tanto no trabalho de Heid et al. (1996), no sentido de realizar um pré-processamento lingüístico no corpus utilizado e posteriormente a realização de consultas sobre o mesmo, quanto no trabalho de Klavans e Muresan (2001), no sentido de realizar uma busca por expressões que introduzem definições de termos. Entretanto, o método lingüístico utilizado não se assemelha totalmente ao método proposto por Heid et. al. em razão do corpus não ter sofrido o processo de lematização. Da mesma forma, o método implementado fugiu um pouco da proposta feita por Klavans & Muresan, em razão de não terem sido realizadas buscas por padrões somente de expressões de definições, mas também de classificações, descrições e outras que concentram termos, além de não ter sido utilizado um módulo de análise gramatical, responsável por identificar definições introduzidas por fenômenos lingüísticos mais complexos, tais como anáforas e apostos. Para a abordagem híbrida foi gerado um conjunto de orações do corpus, chamado de subcorpus, que apresentassem as expressões lingüísticas definidas no método lingüístico, de maneira que cada oração é impressa no subcorpus somente uma vez, independentemente do número de expressões que pode apresentar. Este procedimento representou a “parte lingüística” do método híbrido. O subcorpus de saída, constituído pelas orações que apresentaram alguma expressão lingüística, é tomado como entrada para o pacote NSP, representando assim a parte estatística deste método. Vale ressaltar que se a abordagem de extração de termos selecionada for a lingüística ou híbrida, o corpus deve ser etiquetado morfossintaticamente antes de passar pelo extrator.

4.3.4.2 Métodos estatísticos de extração de candidatos a termos

Embora estejam previstos, atualmente somente o método estatístico de *Frequência Simples* utilizado no projeto *ExPorTer* foi implementado no Módulo 3 do **e-Termos**. A Figura 4.14 apresenta o menu de opções para a seleção deste métodos e os que estarão disponíveis. Observe que também estão previsto a integração dos métodos lingüísticos e híbridos de extração de candidatos a termos.

Figura 4.14: Menu de seleção dos métodos de extração automática de candidatos a termos.

Ao clicar na opção “Frequência Simples” na tela da Figura 4.14, surge a tela da Figura 4.15 que mostra as opções e parâmetros do extrator estatístico. Observe que após escolher o *corpus* o qual o extrator que será aplicado, deve ser selecionado: 1) o tamanho do *N-gram* desejado, que atualmente pode ser de 2 a 7, 2) se uma *Stoplist* vai ser utilizada, e se SIM, se a mesma é a Padrão (oferecida pelo sistema) ou particular criada no âmbito do projeto, 3) decidir sobre o corte na frequência¹⁷ e 4) informar a identificação da Lista de Termos resultante do processo de extração.

Extração Automática de Termos - Frequência Simples*

Defina as opções abaixo conforme desejado e clique em **Extrair Termos**.

Opções do Extrator Automático - Frequência Simples

Cópus:

Tamanho do Termo (n-gram): [Saiba mais...](#)

StopList: [Saiba mais...](#)

Valor do Corte Inferior: [Saiba mais...](#)

Identificação do Resultado: [Saiba mais...](#)

* Algoritmos estatísticos do Pacote NSP (N-Gram Statistic Package)

ATENÇÃO:

1) Dependendo das opções selecionadas esta tarefa pode demorar vários minutos.

2) Campos em vermelho são obrigatórios.

Figura 4.15: Tela que apresenta as opções de extração de candidatos a termos do **e-Termos**.

Há várias maneiras de definir o *Valor de Corte* nos processos de extração. A forma mais elaborada e de melhores resultados compreende no estudo específico do *corpus*, a partir da análise

¹⁷Corte na frequência significa o menor valor da frequência das unidades lexicais do *corpus* que será considerado.

dos histogramas gerados pelos processos de extração, na qual busca-se determinar as classes que aparecem grandes concentração de termos, para então poder especificar qual valor de frequência do último termo que aparece na lista, eliminado, dessa maneira, aquelas classes caracterizadas por um número insignificante de termos. Entretanto, tal análise é muito específica e demanda um estudo minucioso do corpus e das listas geradas, além de ser dependente do domínio do conhecimento do corpus em questão.

Como alternativa, a maneira mais fácil e rápida de definir o *Valor de Corte* é pelo tamanho do corpus. De acordo com Teline (2004) um cálculo do *Valor de Corte* pode ser definido diretamente pela fórmula:

$$\text{Valor do Corte} = (\text{tamanho do corpus} / 100.000) + 1$$

Este cálculo é baseado na premissa de que em um determinado corpus, os candidatos a termos menos frequentes não possuem valor terminológico, visto que sua frequência do corpus é baixa, e, geralmente, em textos especializados os termos ocorrem com frequências maiores. É exatamente este cálculo que o botão **[Sugerir Valor]** executa, podendo assim definir automaticamente o *Valor de Corte*.

Considerando que a principal função do Módulo 3 é a extração, os futuros extratores automáticos de candidatos termos deverão cobrir o restante das abordagens, tanto estatística quanto linguística e híbrida. Assim, a interface deste Módulo deve ser flexível aos usuários e fornecer a possibilidade de seleção dos métodos de extração a aplicar, possibilitando que o usuário selecione aquele que forneça o melhor resultado.

Ainda no contexto das atividades de extração do Módulo 3, importa salientar que mecanismos automáticos de interpretação empírica dos resultados, principalmente aqueles advindos das medidas estatísticas, também devem ser implementados. Isso é muito importante porque auxilia o usuário na avaliação dos resultados dos extratores, bem como na comparação entre medidas ou entre um método e outro, retirando dele parte da carga cognitiva inerente neste processo. A implementação desta interpretação dos resultados será baseada somente nos dados estatísticos ou numéricos, a partir do conhecimento da maneira de avaliação dos mesmos presentes na literatura.

4.3.5 Módulo 4 - Edição do Mapa Conceitual e Categorização dos Termos

Uma vez que a lista de candidatos a termos foi extraída pelas as tarefas do Módulo 3, o próximo passo no processo de criação de produtos terminológicos é a criação de um mapa conceitual correspondente à área objeto e a categorização destes termos neste mapa, para em seguida serem avaliados e validados pelos especialistas da área objeto, conforme já explicitado na Seção 3.4 do Capítulo 3.

Assim, genericamente, o objetivo do Módulo 4 é de editar um determinado mapa conceitual e categorizar os termos de uma lista de candidatos neste mapa. Dessa maneira, o resultado deste Módulo será os candidatos a termos categorizados nos campos nocionais previstos no mapa conceitual do domínio em questão.

Vale salientar que ao longo desta tese os “mapas conceituais” ou as “estruturas conceituais” do domínio tem sido chamados de **ontologia**. Conforme ressaltado em Gomez-Perez et al. (2003) a classificação de ontologias distingue entre ontologias que são principalmente uma taxonomia, denominadas de ontologias *lightweight* de ontologias que modelam o domínio de conhecimento de uma maneira mais profunda fornecendo mais restrições sobre a semântica do domínio, denominadas ontologias *heavyweight*. As ontologias *lightweight* incluem conceitos e sua hierarquia, relacionamentos entre conceitos e propriedades que descrevem os conceitos. Ontologias *heavyweight* adicionam axiomas e restrições de sentido às ontologias *lightweight*. Dessa forma, é importante destacar que as ontologias mencionadas neste trabalho se referem às ontologias *lightweight* já que as mesmas não possuem axiomas e detalhamentos semânticos.

Neste contexto uma das funcionalidades do Módulo 4 é a gerência das *Relações Conceituais* que podem ser utilizadas na especificação das ontologias para definir e nomear o relacionamento semântico entre os conceitos (nós). A *Relações Conceituais* no **e-Termos** são organizadas em duas categorias: 1) as *Relações Genéricas* e 2) as *Relações Específicas*. As *Relações Genéricas* consistem nas relações conceituais oferecidas pelo próprio sistema, podendo ser utilizadas em qualquer ontologia de qualquer projeto, e devem ser vistas como um conjunto de relações que o sistema oferece como padrão. Tais relações são fixas e não podem ser alteradas pelos usuários e são baseadas no trabalho de Lenci et al. (2000). A Tabela 4.3 mostra esse conjunto de relações.

Tabela 4.3: Relações Conceituais Genéricas do **e-Termos** Lenci et al. (2000)

Nome	Descrição	Exemplo
é_um	NO1 é um hiperônimo de NO2	é_um(<leao>,<felino>)
é_um_membro_de	NO1 é um membro ou elemento de NO2. NO1 é uma entidade contável, e NO2 é tipicamente uma entidade coletiva.	é_um_membro_de(<senador>,<senado>)
tem_como_membro	NO1, corresponde a uma entidade coletiva ou conjunto de entidades que tem NO2 como seu membro ou elemento.	tem_como_membro(<bando>,<pássaro>)
é_parte_de	NO1 é parte de NO2.	é_parte_de(<cabeca>,<corpo>)
tem_como_parte	NO1 tem NO2 como uma de suas partes.	tem_como_parte(<avião>,<asa>)
instrumento	NO1 é um evento e NO2 é um instrumento típico, veículo ou dispositivo que é usado para realizar esse evento.	instrumento(<tênis>,<raquete>)
relaciona	NO1 denota uma relação e NO2	relaciona(<amizade>,<pessoa>)
Continua na próxima página...		

	denota as entidades típicas que são relacionadas por ela.	
estado_resultante	NO1 é uma transição e NO2 é o estado resultante da transição.	estado_resultante (<morrer>,<morte>)
é_um_seguidor_de	NO1 é um indivíduo que é um seguidor, um colaborador, um adepto de uma certa religião, doutrina, escola de pensamento ou crença em NO2.	é_um_seguidor_de (<marxista>,<marxismo>)
feito_de	NO2 é tipicamente uma substância ou material do qual NO1 é feito. Alternativamente, NO2 é um elemento que entra na composição de NO1.	feito_de(<pão>,<farinha>)
está_em	NO1 é tipicamente localizado em NO2.	está_em(<oasis>,<deserto>)
vive_em	NO1 é uma entidade viva que tipicamente vive em NO2.	vive_em(<brasileiro>,<Brasil>)
tem_como_cor	NO2 tem tipicamente a cor de NO1.	tem_como_cor(<limão>,<verde>)
atividade_constitutiva	NO2 é a atividade típica de NO1, que é um tipo natural e o assunto do evento expresso por NO2.	atividade_constitutiva (<pássaro>,<voar>)
produz	NO2 é uma entidade natural que é produzida tipicamente por NO1, que é também tipo natural de entidade.	produz(<pássaro>,<ovo>)
produzido_por	NO1 é uma entidade que é produzida tipicamente por NO2 como resultado de um processo natural, intrinsecamente correlato com a natureza de NO2.	produzido_por(<abelha>,<mel>)
propriedade_de	NO2 é um adjetivo que se refere a propriedade, qualidade ou atributo expresso por NO1.	propriedade_de (<inteligência>,<inteligente>)
concerne	NO1 é fenômeno, evento ou situação que tipicamente concerne ou afeta NO2.	concerne(<hepatite>,<fígado>)
contém	NO2 é um objeto que é tipicamente contido em NO1.	contém(<livro>,<informação>)
quantifica	NO1 expressa uma quantidade de NO2.	quantifica(<garrafa>,<líquido>)
medido_por	NO1 é uma propriedade que é medida pelo NO2, uma unidade de medida.	medida_por (<temperatura>,<grau>)
relacionado_a	NO1 é relacionado em algumas maneiras não especificadas a NO2.	relacionado_a(<segundo>,<dois>)
sucessor_de	NO1 é um elemento seguinte a NO2 em uma série.	sucessor_de(<dois>,<um>)
tem_como_efeito	NO2 é um resultado, consequência ou efeito indireto de NO1.	tem_como_efeito (<tempestade>,<trovão>)
típico_de	NO1 é uma doença ou fenômeno que afeta tipicamente a entidade em NO2.	típico_de(<raiva>,<cão>)
causa	NO1 tipicamente causa NO2.	causa(<catapora>,<febre>)
é_a_atividade_de	NO2 é a atividade caracterizadora de NO1.	é_a_atividade_de (<médico>,<curar>)
é_a_habilidade_de	NO2 é tipicamente a habilidade de um indivíduo descrito por NO1.	é_a_habilidade_de (<pintor>,<pintar>)
é_o_hábito_de	NO2 é o hábito típico de um indivíduo descrito por NO1.	é_o_hábito_de (<fumante>,<fumar>)
usado_para	NO2 é a função típica de NO1. Essa relação geralmente se aplica a instrumentos ou dispositivos para conectá-los com a atividade de seu propósito.	usado_para (<empilhadeira>,<levantar>)
usado_por	NO1 é tipicamente usado por NO2.	usado_por(<bisturi>,<cirurgião>)

Continua na próxima página...

usado_contra	NO1 é usado tipicamente contra NO2.	usado_contra(<quimioterapia>,<câncer>)
usado_como	NO1 é usado tipicamente com a função que é expressa por NO2.	usado_como(<madeira>,<material>)
resultado_de	NO1 é uma entidade que é o resultado, efeito ou produto do evento expressp por NO2.	resultado_de(<perda>,<perder>)
causado_por	NO1 é um fenômeno ou evento natural que é produzido por NO2.	causado_por(<infecção>,<bactéria>)
origem	NO2 é a proveniência ou origem de NO1.	origem(<lei>,<sociedade>)
criado_por	NO1 é criada por um certo processo humano ou ação de NO2.	criado_por(<livro>,<escrever>)
derivado_de	NO1 é derivado de um outro objeto de NO2 através de um processo de alteração	derivado_de(<petróleo>,<óleo>)

Ao contrário, as *Relações Específicas* representam as relações conceituais particulares, definidas pelo próprio usuário e são válidas somente no âmbito de cada projeto o qual foram criadas. Tais relações existem justamente para individualizar um determinado projeto, principalmente quando as *Relações Genéricas* não oferecem as relações conceituais necessárias para cobrir a especificidade do trabalho. Nesse sentido, os usuários podem *criar*, *editar* e *excluir* somente as *Relações Específicas*. A Figura 4.16 mostra a tela de criação de uma *Relação Específica* chamada de “**composto_de**”. Entretanto, a tarefa de consultar pode ser executada em todas as relações, tanto as genéricas quanto nas específicas.

Preencha os campos abaixo e clique em **Salvar**

Nova Relação Específica

Nome da Relação: [Ver exemplo...](#)

Descrição:

Exemplo:

Campos em vermelho são obrigatórios.

Figura 4.16: Tela de inclusão de nova Relação Específica.

As subseções a seguir descrevem as principais funcionalidades relacionadas com as ontologias no **e-Termos**.

4.3.5.1 Ferramentas de edição e visualização de ontologias

Diversas linguagens, técnicas e ferramentas têm sido propostas para a organização do conhecimento e sua visualização na Web. Alguns padrões já foram estabelecidos, tais como XML e RDF W3C (2005). Tais propostas tentam estabelecer o uso de metáforas visuais para representar o conhecimento, bem como o estudo de mecanismos de análise e reconhecimento de suas relações. As interfaces dos sistemas de visualização também são objetos de pesquisa e procuram dar soluções viáveis para o reconhecimento e manipulação de informações.

Nesse contexto, no campo da Terminologia algumas ontologias estão sendo desenvolvidas para estabelecer um consenso sobre o significado de conceitos e termos específicos de diversos

domínios do conhecimento Venâncio et al. (2003). Uma ontologia que relaciona hierarquicamente os conceitos e objetos relativos a um determinado domínio permite localizar um objeto (deste domínio) numa base de conhecimento de forma inequívoca, evitando a ambigüidade, pois uma visão específica da ontologia ajuda o usuário a entender o significado dos conceitos e termos apresentados, bem como o contexto em que se inserem.

Atualmente, existem vários editores de ontologias que, além de sua edição, permitem sua visualização. Alguns bons exemplos desses editores são: *Protégé* Noy et al. (2002), o *OntoEdit* citado por Staab e Maedche (2001), o *Inxight StarTree* Inxight (2005), o *TreeBolic Generator*¹⁸ e o *HyperEditor*¹⁹. Uma característica comum entre eles é que são aplicações *stand-alone* que executam fora do ambiente Web, impedindo sua utilização por vários usuários ao mesmo tempo. Entretanto, apesar dessa característica, os editores *Protégé*, *Inxight StarTree*, *TreeBolic Generator* e *HyperEditor* foram construídos em Java e portanto sua migração para o ambiente Web, na forma de Applets, pode ser possível.

Assim como o *OntoEdit*, a maioria dos editores são ferramentas que implementam a visualização de ontologias na forma *folder-tree*, aqui chamada de visualização arbórea, (como o padrão para controle de árvore do Windows). Nesse tipo de visualização, quando um nó é selecionado na árvore à esquerda, seu conteúdo é apresentado à direita da seleção, e assim sucessivamente até o último nível da árvore (nó folha). Nesse caso, quando a estrutura da ontologia possui vários níveis de profundidade, a visualização fica prejudicada.

De outra forma, os editores *Protégé*, *Inxight StarTree*, *TreeBolic Generator* e *HyperEditor*, apresentam a visualização da ontologia na forma de árvore hiperbólica (Hyperbolic Tree) Lamping et al. (1995). Segundo Freitas et al. (2001), esta visualização representa hierarquias através de um layout radial disposto em um plano hiperbólico mapeado para um plano de duas dimensões (2D). Além disso, apresenta aspectos de construção – como o efeito *fish-eye* Furnas (1986) – aliados a um mecanismo simples de navegação pela indicação de um nó de interesse, que é exibido no centro da representação em detalhe, cujo contexto é mantido pela exibição do restante da estrutura da ontologia com nós diminuindo de tamanho até serem suprimidos na borda do círculo radial. A abordagem do plano hiperbólico permite o usuário navegar através dos nós e visualizar a relação da porção visível do plano com a estrutura inteira sobre uma única tela Hao et al. (1999). Com isso, amplia-se o grau de cognição humana sobre determinado assunto.

Assim, a visualização hiperbólica é uma estrutura de visualização de hierarquias baseada na técnica *focus+context* (foco+contexto). Ela destina maior espaço para o nó que está em foco e mostra o contexto (outros nós ao redor do nó focado) com tamanho progressivamente reduzido à

¹⁸Disponível para download em: <http://treebolic.sourceforge.net/en/home.htm>

¹⁹Atualmente, o *HyperEditor* é desenvolvido e distribuído gratuitamente pela Embrapa Informática Agropecuária (www.cnptia.embrapa.br).

medida que se distancia do foco. O nó focado é sempre a raiz da árvore, ou aquele que o usuário selecionar. Para implementar essa idéia, utiliza-se a geometria hiperbólica, na qual a hierarquia é traçada em um plano hiperbólico e este é mapeado em um círculo.

Desse modo, as árvores hiperbólicas são uma representação dinâmica da estrutura hierárquica de uma ontologia, e representa uma maneira eficiente de exibir árvores complexas com exatidão. De acordo com estudo realizado pelo *Xerox Palo Alto Research Center*, citado por Inxight (2005), a técnica árvore hiperbólica para navegação e visualização de coleções de informação hierárquica muito grandes mostrou ser 62% melhor para navegação que o padrão *folder-tree*.

O efeito *fisheye* (também chamado de “olho-de-peixe”, ou visão detalhada) fornece um esquema que, em geral, é suficiente para lidar com a navegação e orientação de grandes redes de informação. Na movimentação dessa interface, os nós da ontologia aumentam e diminuem de tamanho, saindo e entrando em foco, demonstrando grande flexibilidade e agilidade na tela. Inicialmente, na visualização da ontologia, o nó raiz tem o foco principal enquanto que os outros nós apresentam-se deformados ou semi-ocultos em detrimento da parte focalizada, podendo, entretanto, ser expandidos na movimentação da ontologia.

4.3.5.2 A interface Web para edição de ontologias no **e-Termos** – O **OntoEditor**

Como parte da implementação do Módulo 4 foi desenvolvido e totalmente integrado ao **e-Termos** o *OntoEditor*, uma ferramenta para edição de ontologias via Internet, que implementa, tanto a visualização arbórea (*folder-tree*) quando a visualização hiperbólica. Construído sobre várias tecnologias de desenvolvimento Web, como por exemplo, Navegador da Árvore Hiperbólica²⁰, Applets Java, JavaScript e PHP em conjunto com banco de dados MySQL, as principais contribuições deste editor são: 1) a praticidade de criar, editar e excluir quantas ontologias desejar e a qualquer tempo, sendo estas automaticamente integradas ao **e-Termos**, 2) a visualização de ontologias cuja interface combina a flexibilidade de uma árvore hiperbólica com o formato *folder-tree*, 3) a capacidade de converter a estrutura de uma determinada ontologia do formato texto em estruturas de visualização (arbórea e hiperbólica) a partir de uma operação de *upload*, 4) a facilidade de editar uma determinada ontologia, tanto pela forma de tradicional de criação e edição dos nós, como pela capacidade de ações de *arrastar-e-soltar*, 5) a possibilidade de exportar e importar ontologias no formato OWL, 6) a criação de um ambiente específico de validação de ontologias realizadas pelos usuários com perfil *Especialista*, sem a interferência de validações entre eles, e, 7) a flexibilidade de estar acessível via Internet para os usuários do **e-Termos**, possibilitando qualquer usuário a visualizar suas ontologias a qualquer tempo. A Figura 4.17 mostra o menu principal na tela do **e-Termos** que representa as funcionalidades do

²⁰Navegador da Árvore Hiperbólica utilizado nesta tese foi desenvolvido pela Embrapa Informática Agropecuária e está disponível na rede AgroLivre (<http://www.agrolivre.gov.br/>) para *download*.

Figura 4.17: Menu do **e-Termos** com as funcionalidades do *OntoEditor*.

Ainda sobre a Figura 4.17 observe que existe a funcionalidade de Alterar Status. Assim como nos casos anteriores (de *cópus*, Lista de Termos, Stoplist) os status da ontologia especificam a situação atual de uma ontologia em um dado momento durante o trabalho no **e-Termos**. Neste caso os status as ontologias podem ser:

Cancelada : define uma ontologia cancelada por qualquer motivo e que deve ser ignorada pelos membros da equipe;

Categorização dos Termos : representa uma ontologia cuja estrutura já foi validada e os termos estão sendo categorizados;

Edição da Estrutura : consiste nas ontologias que estão em fase de construção e edição, podendo sofrer alterações a qualquer momento;

Finalizada : representa uma ontologia com a tarefa de validação, tanto de termos quanto de estrutura, finalizada;

Liberada : representa uma ontologia validada, pronto e liberada para ser usadas nas próximas etapas;

Validação da Estrutura : representa uma ontologia cuja estrutura está sendo validada por um determinado usuário especialista;

Validação dos Termos : representa uma ontologia cuja categorização dos termos está sendo validada por um determinado usuário especialista.

Todas as funcionalidades do ambiente de trabalho do *OntoEditor* baseiam-se no paradigma de navegação da Web. Os elementos de formulário e de navegabilidade são utilizados para formar um ambiente de trabalho já conhecido pelos usuários da Internet.

Uma das grandes vantagens do *OntoEditor* é permitir a criação de ontologias a partir de arquivos texto (com extensão TXT) tabulados, que representam a estrutura em árvore de uma

ontologia. A título de exemplificação, utilizaremos uma “ontologia exemplo” para acompanhar as fases de criação e edição de uma ontologia. O nome dessa “ontologia exemplo” será EXEMPLO DE ONTOLOGIA e sua estrutura em formato texto está representada na Figura 4.18(a). Observe que cada linha do arquivo representa um nó da árvore que traduz uma ontologia, e que a quantidade de tabulações em cada linha representa o nível hierárquico de cada nó na estrutura. Dessa maneira, o nó de nome “ITEM1” é o nível 0, o nó “ITEM11” é o nível 1 e assim sucessivamente. É importante salientar que não há limites para a quantidade e a profundidade dos nós que podem se criados. A seguir, Figura 4.18(b) mostra esta estrutura no formato de visualização arbórea.

Figura 4.18: Estruturas da ontologia (EXEMPLO DE ONTOLOGIA)

A título de exemplo, mostramos a seguir a visualização hiperbólica gerada pelo *OntoEditor* para uma parte da ontologia do domínio de Revestimentos Cerâmicos. A visualização hiperbólica também oferece um serviço de busca textual para localização de nós, que marca o caminho em vermelho desde a raiz até os nós que contêm a expressão de busca. Observe o resultado da função de busca para o nó do termo “Mistura” mostrado na Figura 4.19.

Outra importante forma de visualização de ontologias do *OntoEditor* é a visualização em formato de grafos, possibilitada pelo uso e adaptação do pacote *TouchGraph WikiBrowser*²¹ que é construído sobre a plataforma Java e pode, dependendo da implementação, ser adaptado para o ambiente Web como *Applets*. A vantagem esta visualização está no fato de que a estrutura da ontologia neste formato é vista como um grafo e não como árvore, sendo esta uma característica de fundamental importância para representar o real comportamento dos termos

²¹<http://www.touchgraph.com/>

Figura 4.19: Visualização hiperbólica da ontologia de Revestimentos Cerâmicos e o destaque para busca do termo “Mistura”.

no âmbito da teoria da TCT, quando considera a possibilidade polissêmica do termo, contendo vários significados, conforme mostrado na Seção 3.3.1 no Capítulo 3. A Figura 4.20 mostra a mesma ontologia de Revestimento Cerâmicos em formato de grafo, novamente com destaque para o termo “Mistura”.

Figura 4.20: Visualização em grafos da ontologia de Revestimentos Cerâmicos e o destaque para o termo “Mistura”.

Dessa forma, o fato do **e-Termos** trabalhar com grafos gera vantagens para expressar o conteúdo de uma dada ontologia que uma árvore não possibilita. Para o exemplo mostrado na Figura 4.19, onde a localização do termo “Mistura” fica destacada, podemos ver que o nó “Mistura” é duplicado na estrutura (2 ramos), pois aparece tanto nos nós “Composição da Massa Cerâmica” e “Preparação da Matéria-Prima”. Por outro lado, observe na Figura 4.20 que só há

um único nó denominado “Mistura” e a busca por ele salientada em amarelo.

Isso porque a visualização hiperbólica e/ou a de grafos dos termos em uma ontologia permite que o terminólogo ou especialista identifique os termos alocados em seus respectivos campos nocionais, favorecendo a visualização das relações conceituais.

A última funcionalidade do Módulo 4 do **e-Termos** é a categorização dos termos na ontologia. Conforme visto na Seção 3.4 esta tarefa consiste em categorizar os termos de uma determinada *Lista de Termos* (com status “Liberada” no Módulo 3) na hierarquia conceitual da ontologia. Esta categorização foi automatizada de maneira a facilitar o trabalho cognitivo do usuário, conforme pode ser visto na tela apresentada na Figura 4.21.

Selecione o nó da ontologia desejado no lado esquerdo, o termo no lado direito e clique nos botões "<<" ou ">>" conforme desejado.

Figura 4.21: Tela de categorização de termos na ontologia.

Observe que a estrutura da ontologia fica do lado esquerdo e a *Lista de Termos* a categorizar do lado direito, de forma que o usuário pode mover os termos de um lado para o outro por meio das ações dos botões de movimentação localizados no centro da Figura. Para auxiliar, verifique também que o sistema muda a cor dos os termos que já foram inseridos como é o caso dos termos do nó “argila refratária” mostrados no exemplo.

Visto que a definição de ontologias é uma questão de pesquisa bastante ampla e que merece atenção especial, é importante ressaltar que o Módulo 4 do **e-Termos** apenas edita a ontologia e não a cria ou define. Portanto, este Módulo deve ser visto, de um lado, como um “editor de ontologias”, sendo necessário que a mesma já esteja previamente elaborada; e de outro, como suporte para a categorização de termos. Importa destacar ainda que para a edição das ontologias e a inclusão dos termos, este Módulo interage principalmente com os usuários *Terminólogos* e *Especialistas do Domínio*, este último, especificamente, na tarefa de avaliar a estrutura da ontologia e a categorização dos termos realizada pelos terminólogos.

4.3.6 Módulo 5 - Criação e Gerenciamento de uma base de dados terminológica

Visto de um panorama externo, o Módulo 5 do **e-Termos** é o que possui a maior carga de trabalho, pois nele se encontram as funcionalidades para o desenvolvimento, análise e edição do produto terminológico a partir da ação dos colaboradores (diferentes usuários do Ambiente) e das ferramentas computacionais de preparação dos termos. Nesse contexto, este módulo contempla duas essenciais funcionalidades do trabalho terminológico: 1) a população da *Base Definicional* e 2) a criação e o preenchimento da *Ficha Terminológica (FT)*.

Genericamente, este Módulo pode ser visto como um gerenciador de uma base de dados terminológicos, representado pela *Base Definicional* e a *Ficha Terminológica*, sendo responsável, portanto, pelo controle do conjunto de dados referentes aos termos em todo processo de preparação final do produto. Desse modo, a estrutura e a implementação deste módulo são firmadas em duas partes, igualmente importantes e interligadas: 1) **o banco de dados terminológicos** e 2) **o sistema de gerenciamento do banco de dados terminológicos**. A seguir, antes de verificar particularmente implementação da *Base Definicional* e da *Ficha Terminológica*, discutiremos cada uma delas.

4.3.6.1 Banco de Dados terminológicos

Na área da Terminologia, um banco de dados terminológico pode ser visto com um conjunto organizado de todos os dados terminológicos utilizados para construção e representação de vários produtos terminológicos. Segundo Tebé *apud* Krieger e Finatto (2004) três aspectos essenciais caracterizam a arquitetura de um banco de dados terminológico: 1) a integração, 2) a estruturação e 3) um grande volume de informação que podem ser armazenadas.

A integração consiste na incorporação, em um suporte único, de informações terminológicas variadas, procedentes de diversas fontes, de modo que possam ser acessadas e utilizadas para criação de produtos terminológicos. A estruturação do banco de dados realiza-se primordialmente pelos itens que perfazem a *Ficha Terminológica*, que por sua vez, ao registrar um determinado termo, contém seus respectivos dados colocados sistematicamente em campos definidos. Por último, o volume de informações registradas é também algo importante quando se caracteriza um banco de dados terminológico, pois um repertório considerável de informações sobre os termos se concentra em apenas um local.

Como em todas as aplicações da informática que utilizam bancos de dados, o modelo de dados corresponde à organização conceitual dos dados de um determinado domínio do conhecimento. É no modelo de dados que se define a hierarquia estrutural e o relacionamento dos dados, que por sua vez, deve refletir fielmente suas propriedades no ambiente real. Assim, a arquitetura ou modelo de dados de um banco de dados terminológico deve possuir três componentes lógicos:

1) *Um conjunto de tipos de elementos*²²; 2) *O conteúdo permitido em cada elemento ou tabela*; 3) *A relação estrutural entre os elementos ou tabelas*. O primeiro e o segundo componentes definem o “conteúdo” e a representação deste conteúdo (dados) no modelo, enquanto que o terceiro componente representa a forma como esse conteúdo se relaciona estruturalmente em cada instância Budin e Melby (2000b).

Conforme visto na Seção 3.4 no Capítulo 3, diversos esforços em direção à criação de um modelo de dados comum, aberto e padronizado que favoreça o intercâmbio de informações tem sido desenvolvidos. Apesar de alguns modelos de dados terem sido desenvolvidos usando a tecnologia SGML, como é o caso do MARTIF, atualmente as tecnologias mais utilizadas na definição de um padrão único são o XML e o Modelo Relacional de Dados, este último voltado especificamente para utilização em Sistemas Gerenciadores de Banco de Dados, relacionais ou orientados a objetos, como por exemplo, o MySQL, o PostgreSQL entre outros.

Ao contrário dos bancos de dados, o XML e o SGML podem ser considerados como linguagens de marcação. Esta marcação permite a implementação de “*templates*” (chamados de DTD ou mais recentemente de *XML Schema*) que especificam a estrutura e a organização dos dados que num dado momento pode ser instanciada. As maiores vantagens do uso do XML incluem a independência de plataforma computacional, a flexibilidade de manipulação da estrutura e seus elementos, o escopo e validação de dados. Entretanto, uma grande desvantagem do XML é a ausência de uma linguagem específica de consulta e, conseqüentemente, a incapacidade de indexação de dados.

Uma abordagem do uso de banco de dados relacional para aplicação em Terminologia foi desenvolvida no projeto *RelTef*²³. O *RelTef* é um modelo de dados relacional baseado na norma ISO 12620, e consiste num conjunto de tabelas (entidades) organizadas num diagrama de Entidade-Relacionamento (ER) que pode ser implementado por qualquer Sistema Gerenciador de Banco de Dados (SGBD) Hardman (1999).

As vantagens do uso de banco de dados relacionais nas bases terminológicas são: o uso das restrições de integridade, a indexação de dados (para melhorar os resultados das consultas), acesso aos dados através da Linguagem SQL (*Structured Query Language*), que por sua vez é amplamente aceita e utilizada, além de fornecer a possibilidade de modelagem conceitual da ontologia e permitir consultas de conjuntos de dados específicos. Outras vantagens podem incluir a grande capacidade de armazenamento, a velocidade de manipulação e recuperação de informações, a habilidade de importar e exportar dados de forma transparente, o acesso múltiplo de usuários e a proteção (operação de *lock*) de registros.

²²Os conjuntos de tipos de dados podem ser chamados de *elementos* quando são formatados na estrutura da linguagem XML ou *tabelas* quando utilizados em Bancos de Dados Relacionais.

²³Mais informações sobre o *RelTef* podem ser encontradas no endereço: <http://www.ttt.org/clsframe/reltef.html>

Em relação ao Módulo 5 e, conseqüentemente, a todo projeto do **e-Termos**, o desenvolvimento do modelo de dados seguiu o modelo relacional, que suporta todas as atividades dos módulos e permite o intercâmbio de dados e informações. Neste sentido, destaca-se a influência positiva e importância da *Diadorim*²⁴ uma base de dados lexicais disponível na Web, desenvolvida no NILC, que utiliza a abordagem de bancos de dados relacionais Greggi (2002). Portanto, vale ressaltar a contribuição desta experiência no âmbito do desenvolvimento e implementação deste Módulo, já que utilizamos a Web e o banco de dados relacionais para a elaboração da definição e armazenamento de termos.

Decorre, então, que o armazenamento estruturado de informações é fundamental para o trabalho terminográfico. Seja qual for a envergadura do banco de dados terminológico, a organização, representatividade, segurança e confiabilidade das informações figuram como elementos essenciais para o sucesso de um empreendimento terminológico.

4.3.6.2 Sistemas de Gerenciamento do Banco de Dados terminológicos

Um sistema de gerenciamento da base de dados terminológicos²⁵ consiste num conjunto de programas ou ferramentas computacionais voltadas especificamente para a gerência das atividades e processos executados sobre um banco de dados terminológicos. A principal vantagem dos sistemas de gerenciamento é a junção de das principais tarefas de elaboração e controle da base de dados terminológicos em um único sistema, com o objetivo de facilitar todo este trabalho, além de gerir de forma satisfatória e organizada toda massa de dados.

Sob esta visão, o Módulo 5 sozinho pode ser visto como um sistema desta categoria com um diferencial: a possibilidade de permitir a edição colaborativa da Base Definicional e das Fichas Terminológicas, interagindo com vários perfis de usuários e sendo capaz editar os registros do banco de dados.

Atualmente existem vários sistemas de gerenciamento de base de dados terminológicos²⁶. A Tabela 4.4 apresenta uma lista dos principais sistemas, que possuem praticamente a mesma função principal: suportar a elaboração e o gerenciamento de uma base de dados terminológicos.

²⁴Disponível para acesso em www.nilc.icmc.usp.br/nilc/tools/intermed.htm

²⁵Do inglês: TMS - Terminology Management System

²⁶Uma lista de extensa de vários sistemas de gerenciamento de base de dados terminológicos pode ser vista em: <http://www.uni-leipzig.de/xlatio/frs-soft.htm>

Tabela 4.4: Principais sistemas de gerenciamento de base de dados terminológicos

Sistema/Endereço/Principais funcionalidades	Disponibilização
Corpógrafo www.linguateca.pt/corpografo/ <ul style="list-style-type: none"> • Compilação de corpus; • Pesquisas sobre o corpus (concordâncias, estudo de N-grams); • Extração de candidatos a termos (usando uma combinação de algoritmos estatísticos e filtros lexicais); • Armazenamento de termos em bases de dados terminológicas específicas; • Pesquisa de definições no corpus de um dado termo; • Pesquisa de possíveis relações semânticas entre termos no corpus (Holonímia/Meronímia, Hiperonímia/Hiponímia, Causa/Efeito, etc.); • Exportação de dados e resultados em XML. 	Livre
System Quirk Language Engineering Workbench SystemQuirk (2003) www.computing.surrey.ac.uk/SystemQ/ <ul style="list-style-type: none"> • Gerenciamento de dados terminológico • Extração de candidatos a termos • Análise de textos • Gerenciador de corpus • Módulos adicionais: sumariizador e alinhadores de texto; 	Livre
TRADOS MultiTerm www.trados.com <ul style="list-style-type: none"> • Gerenciamento de dados terminológicos (aplicativos MultiTerm Server, MultiTerm Desktop) • Extração de candidatos a termos (aplicativo MultiTerm TermExtract) • Edição colaborativa de entradas dos dicionários (aplicativo MultiTerm Online) • Compartilhamento de dados com conexões de vários clientes (aplicativo MultiTerm Team) Proprietário 	Proprietário
TermDB - Terminological Database http://services.canoo.com/services/TermDB/ueberblick/index.html?lang=en <ul style="list-style-type: none"> • Armazenamento e consulta de termos • Publicação dos termos na Intranet e Internet Proprietário 	Proprietário
XTS - Xerox Terminology Suite XEROX (2003) www.mkms.xerox.com <ul style="list-style-type: none"> • Gerenciamento de dados terminológicos (aplicativo TermChecker e TermOrganizer) • Extração de candidatos a termos (aplicativo TermFinder) • Consulta de termos na Web (aplicativo TermOnLine) Proprietário 	Proprietário
Club Cycom Terminology Management http://club.cycom.co.uk <ul style="list-style-type: none"> • Gerenciamento de dados terminológicos • Consulta de termos • Importação e exportação de dados nos formatos Martif e TBX 	Proprietário
TermSTAR TermStar (2001) e WebTerm www.star-group.net <ul style="list-style-type: none"> • Gerenciamento de dados terminológicos (aplicativo TermStar) • Consulta de termos na Web (aplicativo WebTerm) 	Proprietário
TERMITE - ITU Telecommunication Terminology Database www.itu.int/terminology/index.htm <ul style="list-style-type: none"> • Armazenamento e consulta de termos da área de telecomunicações na Web Livre 	Livre
TERMINUS http://igraine.upf.edu/Terminus2/index.html <ul style="list-style-type: none"> • Gestão de corpus • Gerenciamento de dados terminológicos 	Proprietário
Lingo http://www.lexicool.com/soft_lingo2.asp <ul style="list-style-type: none"> • Gerenciamento de dados terminológicos • Integração com editores de texto 	Proprietário
Continua na próxima página...	

<ul style="list-style-type: none"> • Exportação de dados nos formatos TBX e Texto Puro (.txt) 	
TermWeb http://www.interverbumtech.com/Termweb.html <ul style="list-style-type: none"> • Gerenciamento de dados terminológicos • Publicação dos dicionários na Intranet e Internet • Impressão do dicionário em PDF • Compatibilidade com o formato TBX • Compatibilidade com o Trados MultiTerm 	Proprietário
IATE - InterActive Terminology for Europe http://iate.europa.eu <ul style="list-style-type: none"> • Consulta de termos na Web de várias línguas europeias 	Livre
Déjà Vu http://www.atril.com <ul style="list-style-type: none"> • Memória de tradução • Gerenciamento de dados terminológicos 	Proprietário
EDITerm http://www.editerm.com/profile.asp <ul style="list-style-type: none"> • Gerenciamento de dados terminológicos 	Proprietário
LexSyn http://www.babeling.com/accueil.html <ul style="list-style-type: none"> • Memória de tradução • Gerenciamento de dados terminológicos • Edição colaborativa dos termos no dicionário 	Proprietário

A maioria destes sistemas, na tarefa de gerenciar a base de dados terminológicos, oferecem um conjunto de funcionalidades que facilitam a tarefa de editar e preencher os campos da Ficha Terminológica. Tais funcionalidades correspondem a: 1) manter o banco de dados e a Ficha, 2) visualizar (consultar) um conjunto de termos (ou um termo específico) da base de dados, 3) editar os termos, envolvendo as operações de inclusão, alteração e exclusão dos campos de dados, 4) organizar e editar os contextos dos termos, 5) visualizar os campos de dados de um determinado termo, 6) imprimir um conjunto de termos a partir de uma consulta e 7) preencher as Fichas Terminológicas.

No contexto dos sistemas de gerenciamento terminológico, destacamos aqui a relevância do *Corpógrafo* Sarmento et al. (2004) e a sua contribuição para a pesquisa terminográfica. Desenvolvido no pólo de Porto da Linguatca²⁷ o *Corpógrafo* é uma plataforma Web destinada à análise e ao trabalho sobre *corpú*s com o objetivo essencial de produzir recursos lingüísticos computacionais, inclusive os de natureza terminológica. Em linhas gerais, o *Corpógrafo* pretende apoiar os pesquisadores da língua portuguesa num conjunto de tarefas que vão desde a compilação de *corpú*s, à extração e organização do conhecimento gerado a partir deles.

Portanto, considerando os aspectos terminológicos, contextualizamos a semelhança existente entre os objetivos do **Corpógrafo** e do **e-Termos**, principalmente em relação à utilização de *corpú*s, a extração de candidatos a termos e o gerenciamento da base de dados terminológica. Entretanto, a despeito destas semelhanças, enfatizamos que o **e-Termos** não concorre com o **Corpógrafo** e possui características próprias que os diferenciam, como a compilação automática de *corpú*s, a edição da ontologia e consequente categorização de termos, a exportação dos pro-

²⁷Mais informações sobre os trabalhos desenvolvidos da Linguatca acesse: www.linguatca.pt

duos terminológicos em formatos de intercâmbio e principalmente, a face singular do trabalho colaborativo que o **e-Termos** implementa para o exercício terminológico.

Nessa mesma linha, é muito importante destacar ainda a semelhança entre alguns dos sistemas expostos na Tabela 4.4 com o **e-Termos**. Particularmente é possível citar o **Trados MultiTerm**, o **System Quirk**, o **LexSyn** e o **Terminus**. Os três primeiros sistemas possuem componentes e funcionalidades bem parecidas com o **e-Termos**, como por exemplo: a extração automática de termos (**Trados MultiTerm** e **System Quirk**), a gerência da base terminológica (todos), a análise de texto e o armazenamento de corpus (**System Quirk**) e, principalmente, a edição colaborativa dos termos (**Trados MultiTerm** e **LexSyn**). Contudo, tais semelhanças são pontuais, e as paridades acontecem com apenas alguns módulos do **e-Termos**. Isto é, a extração automática se assemelha com o Módulo 3, a análise de texto e armazenamento de corpus ao Módulo 2, o gerenciamento da base e a edição colaborativa ao Módulo 5.

Todavia, o sistema que mais se assemelha com o **e-Termos** é o **Terminus**. Desenvolvido no *IULA (Institut Universitari De Lingüística Aplicada - Universitat Pompeu Fabra, Barcelona, Espanha)* o **Terminus** representa a metáfora de uma estação de trabalho inteiramente dedicada à terminologia e integra num único ambiente: 1) a gestão, análise e organização de textos e corpus, 2) a estruturação conceitual e categorização de dados, 3) a gestão de glossários e 4) a criação e manutenção de bases de dados e dicionários. Assim, é fácil notar que o **Terminus** também (semi-)automatiza as atividades do processo de criação dos produtos terminológicos e representa uma fundamental iniciativa de sistematização do trabalho terminológico. Entretanto, é possível citar como desvantagens deste sistema a não integração da estrutura conceitual²⁸ ao ambiente de trabalho e ao projeto, a *Ficha Terminológica* de tamanho fixo e pré-determinado, a impossibilidade de exportação de dados e o acesso proprietário.

Observa-se, portanto que embora exista sistemas concorrentes nas mais variadas funcionalidades o **e-Termos** ainda se destaca por representar um sistema que abrange e integra todas as etapas e tarefas envolvidas na gestão terminológica, não estando restrito somente as atividades pontuais e específicas de cada fase do trabalho. Além disso, sobressai desse meio o fato do **e-Termos** ser um sistema Web, gratuito e de acesso livre, diferentemente dos sistemas citados anteriormente.

4.3.6.3 A população da Base Definicional

O preenchimento da *Base Definicional* corresponde a uma das principais tarefas colaborativas do **e-Termos**. Isso porque esta funcionalidade permite que um determinado termo tenha sua base de excertos populada por vários usuários simultaneamente, em concorrência e de maneira

²⁸Neste sistema, a estrutura conceitual do domínio é criada fora do ambiente de trabalho e não é integrada ao sistema, não permitindo, portanto, a utilização dos termos hierarquizados nos campos nocionais.

totalmente transparente. A Figura 4.22 mostra a tela de inclusão de um excerto do termo “Agitador”.

Figura 4.22: Tela de inclusão de um novo excerto definicional.

Observe que nesta tela é possível consultar o corpus em busca de excertos ou trechos definitrios do termo em questo, caracterstica que mostra a integrao do sistema e a possibilidade de se basear no corpus para a execuo do trabalho.

Dessa maneira, alm do prprio texto do excerto e a sua fonte (dados inseridos pelo usurio) cada registro inserido na base tambm possui o responsvel pelo preenchimento, a data e hora de incluso, adicionados automaticamente pelo sistema. Alm do mais, todos os usurios podem ver e consultar todos os excertos de todos os termos, sem nenhuma restrio de contedo. A Figura 4.23 mostra a tela do excerto do termo “Agitador” aps incluso.

Veja abaixo o excertos definitrios para este termo. Clique em **+ Adicionar Excerto** para adicionar novos excertos. Para excluir ou editar um determinado excerto clique em **✖ Excluir** ou **✎ Editar** respectivamente.

Excertos Defin�rios				
Termo: Agitador				
Excerto	Fonte	Respons�vel	Data/Hora	A��o
Este estudo iniciou pela considera��o de diferentes alternativas, adotando-se ao final, o silicato de s�dio na propor��o de 2%. simultaneamente, conta-se com a agita��o da barbotina atrav�s de uma h�lice presa a um agitador posicionado lateralmente � caixa met�lica que comp�em a bacia de conten��o, conforme pode ser visto � esquerda da Figura 9, a qual permanece acionada durante todo o per�odo de ensaio.	C�rpus...	Leandro Henrique	2009-08-11/21:41:33	✖ ✎
As composi��es foram preparadas com a mistura dos materiais n�o-pl�sticos com o caulim que foi previamente dispersado em um agitador a disco de alta velocidade (1200 rpm).	C�rpus...	Leandro Henrique	2009-08-11/21:42:10	✖ ✎

+ Adicionar Excerto

Figura 4.23: Tela do excerto do termo “Agitador”.

Vale dizer que todos os excertos definicionais includos para um determinado termo so automaticamente recuperados no momento da redao da definio final dos termos, na funcional-

idade de preenchimento da *Ficha Terminológica*, vista na seção seguinte. Assim, os excertos e a *Base Definicional* ficam sempre disponíveis para os usuários, principalmente para fornecer mais detalhes sobre as particularidades da definição do objeto a ser definido.

Nesse contexto, a *Base Definicional* no **e-Termos** também é uma substancial contribuição do sistema, pois além de sistematizar uma tarefa do trabalho terminológico também organiza as atividades envolvidas na definição, melhorando sua qualidade.

4.3.6.4 A criação flexível da Ficha Terminológica

Um dos principais problemas encontrados na criação das *Fichas Terminológicas* refere-se à definição dos campos e os tipos de dados que serão contemplados. Isso porque, uma incorreta definição dos campos pode tornar o trabalho terminológico pobre e incompleto, e, consequentemente, prejudicar a *Edição dos Verbetes* e a divulgação final do produto terminológico (tarefas constantes no Módulo 6). Decorre deste fato que uma das principais dificuldades encontradas nos sistemas de gerenciamento terminológico é a definição flexível da Ficha, pois cada trabalho ou produto terminológico possui sua própria Ficha de acordo com seus objetivos e finalidades. A maioria destes sistemas oferece um padrão fixo de Ficha, previamente definida e que não é adaptada às necessidades de cada trabalho, de maneira que ora são muito grandes, com excesso de campos, não necessários para determinados trabalhos; ora muito pequenas, com ausência de campos, não atendendo as necessidades mínimas do registro de dados.

É também neste ponto que o Módulo 5 atua: na criação flexível e adaptada da *Ficha Terminológica* às necessidades de cada trabalho terminológico. Usando as funcionalidades de *inclusão*, *exclusão* e *alteração*, é possível criar um número indeterminado de Fichas de forma flexível, sem limite de campos, de diferentes e variados *Tipos de Dados*, de acordo com as necessidades, características e objetivos de cada trabalho ou produto terminológico. O menu com tais funcionalidades pode ser visto na Figura 4.24.

Figura 4.24: Menu das funcionalidades da Ficha Terminológica.

Na tarefa de criação da Ficha é possível adicionar ou excluir campos a qualquer momento, conforme pode ser visto na tela da Figura 4.25, onde é possível informar o nome de identificação da Ficha (neste exemplo “FichaTerminologica1”) e adicionar quantos campos desejar clicando no link “+ **Adicionar Campos**”.

Para cada campo adicionado na Ficha é possível informar: 1) o *nome* do campo, 2) o *Tipo de*

Dados do Protocolo da Ficha Terminológica

Nome:

Campos:

Nome do Campo	Tipo de Dado	Obrigatório	Ação
CodigoTermo	Alfanumérico	SIM	

[+ Adicionar Campo](#)

ATENÇÃO: O campo "CodigoTermo" é obrigatório e criado automaticamente em todo protocolo.
Campos em vermelho são obrigatórios.

Figura 4.25: Tela de criação e edição da Ficha Terminológica.

Dado e 3) a *obrigatoriedade* de preenchimento, conforme pode ser visto na Figura 4.26. O *nome* do campo representa o rótulo ou como um determinado campo será identificado. O *Tipo de Dado* revela a natureza dos dados que o campo irá assumir, e a *obrigatoriedade* diz se o campo é de preenchimento obrigatório. Uma vez adicionado, é possível excluir (ação do botão **X**) os campos incluídos ou continuar adicionando novos campos, conforme pode ser visto na tela da Figura 4.27. O botão “**Salvar Protocolo**” grava a Ficha.

e-Termos: Adição de campos no Protocolo de Ficha Terminológica.

Preencha o nome e selecione o tipo de dado do campo que deseja adicionar e clique em **Adicionar Campo**.

Nome do Campo: Evite usar caracteres especiais, acentuados e numerais.

Tipo de Dado: [Saiba Mais...](#)

Obrigatório: ☒

Figura 4.26: Tela de adição de novos campos da Ficha Terminológica.

Dados do Protocolo da Ficha Terminológica

Nome:

Campos:

Nome do Campo	Tipo de Dado	Obrigatório	Ação
CodigoTermo	Alfanumérico	SIM	
CampoUm	Alfanumérico	SIM	X

[+ Adicionar Campo](#)

ATENÇÃO: O campo "CodigoTermo" é obrigatório e criado automaticamente em todo protocolo.
Campos em vermelho são obrigatórios.

Figura 4.27: Tela de exclusão e adição de novos campos da Ficha Terminológica.

Após salvar uma determinada Ficha, a tarefa de edição permite que os campos de uma determinada Ficha sejam alterados ou excluídos, e, ainda, que novos campos sejam adicionados. A exclusão de uma determinada Ficha acontece após a confirmação realizada na tela mostrada na Figura 4.28. Vale lembrar que a exclusão elimina todos os campos e definições da Ficha.

A liberdade de criar Fichas de tamanhos variáveis com a possibilidade de configuração plural dos seus campos, somente é possível devido à implementação de uma funcionalidade que permite a criação da estrutura lógica de um banco de dados em tempo de execução. Este banco de dados pode ser entendido como um “*metabanco*” e é acoplado ao banco de dados oficial do sistema, fazendo parte do seu conjunto de dados. Dessa maneira, cada Ficha pode ser vista

Confirme o nome do Protocolo da Ficha Terminológica e clique em **Excluir Protocolo**.

Dados do Protocolo da Ficha Terminológica
Nome do Protocolo: **FichaTerminologica1**

Figura 4.28: Tela de exclusão da Ficha Terminológica.

como a imagem de um banco de dados específico, formado por uma única tabela cujos atributos representam cada campo da Ficha. Adicionalmente, para manter a integridade dos dados o sistema mantém um controle dinâmico da estrutura, do status de uso e da configuração da Ficha, de maneira que cada alteração seja atualizada corretamente. Assim, cada Ficha Terminológica pode ser entendida com uma base de dados conceitual e específica, sendo controlada e mantida de forma transparente pelo sistema.

Os Tipos de Dados

Além da capacidade de criar vários campos, o **e-Termos** oferece também diversos *Tipos de Dados* que devem ser vinculados a cada campo e representam a natureza ou a categoria de dados que um determinado campo pode assumir. Assim, por exemplo, se um determinado campo é do tipo *numérico*, apenas representações numéricas podem ser armazenadas no espaço deste campo. Na versão atual do **e-Termos** são disponibilizados nove *Tipos de Dados* diferentes, que podem ser divididos em *Primários* e *Especiais*. A Tabela 4.5 apresenta os *Tipos de Dados* disponíveis e sua classificação.

Tabela 4.5: Tipos de Dados disponíveis no **e-Termos**

Tipo de Dado	Descrição dos Valores	Classificação
Alfanumérico	Letras, números e caracteres especiais até o limite de 255 caracteres	Primário
Texto	Textos com letras, números e caracteres especiais	Primário
Definição	Representa o texto definatório do termo	Especial
Data	Números e caracteres que representam uma data	Primário
Morfologia	Representa o nome da classe morfológica de um determinado termo	Especial
Gênero	Representa o Gênero de um determinado termo	Especial
Tipo Textual	Representa o Tipo Textual de um determinado termo	Especial
Área	Representa a grande área do conhecimento a qual um determinado termo pertence	Especial
Subárea	Representa a subárea do conhecimento a qual um determinado termo pertence	Especial

Os *Primários* são valores básicos que não recebem nenhuma conotação de valor especial, ou seja, são dados numéricos, alfanuméricos, de longas cadeias de caracteres ou de data. Já os *Especiais* representam categorias de dados pré-definidos ou que possuem algum valor especial agregado, específico no âmbito do trabalho terminológico. Por exemplo, os *Tipos de Dados* **Morfologia** e **Tipo Textual** representam conjuntos de valores que representam as classes

morfológicas e os tipos textuais que um determinado *termo* pode pertencer, como *Substantivo* ou *Adjetivo* e *Notícia* ou *Decreto*, respectivamente. Outro exemplo seria o *Tipo de Dado Definição* que representa o texto definitório do *termo*. Neste caso, o **e-Termos** vincula automaticamente para este campo um *Editor de Definição*, outra ferramenta agregada ao Módulo 5 mencionada como número 10 da Tabela 4.1).

Os *Tipos de Dados Especiais* são implementados a partir de abstrações conceituais de determinados conjuntos de dados específicos, que por sua vez, são mapeados por uma *Classe* exclusiva do sistema. Assim, por exemplo, o conjunto de *Tipos Textuais* passíveis de uso no contexto do trabalho terminológico, sofre uma abstração conceitual implementada por uma determinada *Classe* de maneira que o *Objeto* de sua instância é a imagem deste conjunto, chamado apenas de “*Tipos Textuais*”. Neste caso, a carga de mapeamento e tradução do conceito para os dados primários são implementados pela *Classe*. Portanto, podemos considerar que os *Tipos de Dados Especiais* também são características diferenciais do **e-Termos** em relação à flexibilidade e liberdade de criação das Fichas, já que representa os aspectos naturais do contexto terminológico.

Do exposto, é fato que a liberdade de criação das *Fichas Terminológicas* implementada no **e-Termos** é um importante diferencial deste frente aos sistemas de gestão de terminologia, e, principalmente, no contexto da semi-automatização do processo de elaboração dos produtos terminológicos. Isso porque as diversas possibilidades de definição das Fichas pelo usuário permitem o empreendimento de múltiplos trabalhos, já que a Ficha pode concentrar todos os dados acerca dos *termos* e assim, revelar a riqueza de um repertório terminológico. Adicionalmente, tal característica elimina contratação de especialistas em informática para adequar um sistema de gestão às necessidades de um projeto.

Contudo, ainda é possível expandir as funcionalidades e características deste quesito, principalmente na implementação de novas funcionalidades de edição e preenchimento das Fichas, principalmente no desenvolvimento de ferramentas de consulta dos campos dinamicamente definidos; e na ampliação dos *Tipos de Dados* disponíveis, como por exemplo: *remissivas*, *sinônimos*, *antônimos* e dados multimídia, como *imagens*, *áudio* e *vídeo*.

Outra funcionalidade a estar presente neste módulo é a redação assistida da definição terminológica. Como esta é uma tarefa bastante complexa numa pesquisa terminográfica, mais esforço computacional será exigido para a implementação dessa ferramenta e por isso serão consideradas as abordagens de tipologia definicional utilizadas em sistemas de ensino apoiado por computador, como aquelas descritas em Aluísio (1995) e Antiqueira et al. (2003).

Tipologia Definicional

A tipologia definicional consiste em um conjunto de tipos padrões de definição previamente estabelecidos que facilitam a elaboração de definições. Podendo ser aplicada no campo da escrita

técnica (por exemplo, na escrita de artigos científicos) a tipologia definicional também pode ser aplicada à definição de termos.

Há cinco tipos de definições, sendo elas: **formal**, **semiformal**, **substituição**, **ilustração** e **estipulação**. A principal característica que as diferenciam é a precisão ou quantidade de informação. Enquanto a **formal** apresenta os elementos semânticos *termo*, *classe* e *característica(s)*, a **semiformal** não apresenta a *classe* e a **substituição** introduz uma nova informação com um significado similar ao termo que foi introduzido, isto é, apresenta uma reformulação do termo. Uma definição por **ilustração** pode ser dividida em *exemplificação* ou *particularização*. Por último, o tipo **estipulação** é encontrado unido aos outros tipos de definições e seu propósito é colocar limites de tempo, lugar, área de pesquisa ou de significado, para a definição que a acompanha Aluísio (1995).

Atualmente no **e-Termos** já existe um *Editor de Definição* acoplado ao preenchimento da Ficha Terminológica, sempre vinculado ao campo que possui o tipo de dado “definição”. Esta ferramenta representa um auxílio à redação de definições dos termos por meio do uso desses padrões pré-definidos, bem como da importação automática dos excertos definicionais dos termos, armazenados na *Base Definicional*, fundamentais para redigir as definições. Assim, o uso dessa ferramenta é bastante relevante no contexto da edição da *Ficha Terminológica*, pois permite: 1) uniformizar a redação dos textos, 2) vincular determinada tipologia definicional ao termo que será definido, e 3) facilitar a escrita das definições. A Figura 4.29 mostra uma tela do *Editor de Definição* com a definição do termo “Agitador”. Observe os campos da Ficha no lado esquerdo da tela, e os excertos definicionais do termo no lado direito, junto ao texto da definição.

Figura 4.29: Tela do *Editor de Definição*.

Particularmente, visando melhorias na sistematização do *Editor de Definição*, esta ferramenta pode incorporar os aspectos conceituais e lingüísticos relacionados com a definição terminológica, principalmente aquelas resultantes do recente trabalho de Souza (2009).

4.3.7 Módulo 6 - Edição dos Verbetes e Uso de Formatos de Intercâmbio dos Produtos Terminológicos

Em repetidas ocasiões ao longo da descrição do **e-Termos** foram feitas várias referências a um dos aspectos mais importantes da terminologia atual: a necessidade de intercâmbio de conhecimento veiculado pelos produtos terminológicos. Portanto, é de se esperar que num trabalho em que um dos objetivos é a organização e armazenamento estruturado de informações terminológicas, específicas de um determinado domínio, se fizesse um esforço para a realização de intercâmbio e difusão destes produtos.

De maneira geral, o Módulo 6 tem como função primária a difusão do conhecimento inerente aos produtos terminológicos produzidos pelo **e-Termos** a partir das tarefas de todos os Módulos anteriores. Portanto, o propósito deste Módulo é oferecer o conjunto de verbetes de um determinado produto terminológico, referente a domínio específico, em um padrão definido, seja para fins de consulta simples ou de intercâmbio com outros sistemas.

Para cumprir este propósito, três funcionalidades são oferecidas aos usuários: 1) a *edição dos verbetes*, 2) a *publicação do produto* e 3) a *exportação de dados* em um formato de intercâmbio. A *edição dos verbetes* compreende a criação dos *Modelos de Verbetes*, que por sua vez, consiste na elaboração de um modelo que especifica quais campos da *Ficha Terminológica* estarão presentes no verbete. Especificamente, esta funcionalidade disponibiliza a Ficha Terminológica e permite que os usuários selecionem os campos do verbete, sua formatação (fonte, estilo, tamanho e cor) e a ordem em que aparecem.

4.3.7.1 A criação dos Modelos de Verbetes e Verbetes

A Figura 4.30 mostra a estrutura de uma *Ficha Terminológica* e exemplos de três *Modelos de Verbetes*. Observe que o **Modelo 1** possui os campos Termo na fonte Arial 16, Negrito, Preto; Morfologia na fonte Times New Roman 14, Itálico, Preto e Definição na fonte Arial 14, Itálico, Azul claro. Já o **Modelo 2** possui os campos Termo na fonte Arial 14, Negrito, Vermelha; Definição na fonte Times New Roman 14, Itálico, Preto e Data na fonte Arial 12, Itálico, Verde. Por sua vez, o **Modelo 3** só possui dois campos, Termo na fonte Arial 16, Negrito, Azul e Informação Enciclopédica, Verdana 12, Itálico, Cinza.

A partir destes Modelos é possível criar diferentes verbetes, com aparência e formatação particularizada, conforme mostrado na Figura 4.31. Veja que os verbetes contêm os dados formatados conforme o Modelo selecionado.

Em outras palavras um *Modelo de Verbetes* representa a estrutura de formação de um *Verbetes*, que por sua vez, é a instanciação dos dados dos termos junto ao Modelo. Na implementação atual não há limite para a quantidade de Modelos e nem tamanho fixo de campos, obedecendo

Figura 4.30: Exemplos de Modelos de Verbete.

Figura 4.31: Verbetes formatados a partir dos Modelos de Verbete.

somente o número de campos da Ficha. Além disso, não existe um padrão de formatação fixa, de maneira que a aparência dos verbetes segue o desejo do usuário ou a especificação do produto. No entanto, vale salientar que a principal vantagem desta funcionalidade é a possibilidade de criar vários produtos terminológicos diferentes a partir de um mesmo conjunto de dados (a Ficha) com formatações e aparências adequadas a cada situação. A Figura 4.32 mostra a tela do **e-Termos** de criação de um *Modelo de Verbete*.

4.3.7.2 A pré-publicação e publicação dos produtos terminológicos

A *publicação* é a última tarefa do processo de criação de produtos terminológicos e consiste na divulgação formal do produto na Internet, no **Portal do e-Termos**, acessível pelo link de “Acesso aos Produtos”, conforme dito anteriormente neste mesmo capítulo. A funcionalidade de publicação é dividida em *pré-publicação* e *publicação*. Ambas têm o mesmo objetivo, com a diferença que a *pré-publicação* disponibiliza o produto apenas para o ambiente interno, no contexto do projeto, somente para os membros da equipe de trabalho. Além disso, a *pré-*

Dados Modelo de Verbetes

Identificação:

Campo	Fonte	Tamanho	Estilo	Cor
☒ Termo	Arial	10 pt.	Negrito	Preta
☐ Sinonimo	<input type="text" value="Arial"/>	<input type="text" value="1 (8pt)"/>	☐ N ☐ I ☐ S	<input type="text" value="#000000"/>
☐ Morfologia	<input type="text" value="Arial"/>	<input type="text" value="1 (8pt)"/>	☐ N ☐ I ☐ S	<input type="text" value="#000000"/>
☒ Definicao	<input type="text" value="Times New Roma"/>	<input type="text" value="3 (12)"/>	☐ N ☒ I ☐ S	<input type="text" value="#000000"/>
☒ InfEnciclopedia	<input type="text" value="Verdana"/>	<input type="text" value="3 (12)"/>	☒ N ☒ I ☐ S	<input type="text" value="#000000"/>
☒ DataInclusao	<input type="text" value="Arial"/>	<input type="text" value="3 (12)"/>	☐ N ☒ I ☒ S	<input type="text" value="#333333"/>

ATENÇÃO: O "TERMO" é inserido automaticamente em todo verbete com fonte Arial, tamanho 10pts., negrito.

Ordem:

Campos em vermelho são obrigatórios.

Figura 4.32: Tela de criação do Modelo de Verbetes.

publicação também serve como teste para a *publicação*, permitindo que os usuários ajustem detalhes ou eventuais erros antes de disponibilizar oficialmente o produto.

Tanto a *pré-publicação* quanto a *publicação* tomam de entrada: 1) o *Modelo de Verbetes* o qual deve ser aplicado dos dados dos verbetes, 2) a forma de apresentação dos campos, se na mesma linha ou cada campo em uma linha, 3) o *nome fantasia* que representa um nome de apresentação do produto, e 4) o *logo*, que consiste numa imagem ou figura que representa o produto. Adicionalmente, a tarefa de publicação final também recebe como entrada um pequeno texto de *resumo* de apresentação do produto, que é colocado abaixo do link de acesso a página. A Figura 4.33 mostra a tela de *pré-publicação*.

Para realizar a pré-publicação, selecione o Modelo de Verbetes que deseja Pré-Publicar, e, se desejar, informe o Nome Fantasia e faça o **Upload** de um logo. Clique em **Pré-Publicar** para salvar a Pré-Publicação.

Para ver um template da pré-publicação clique em [Ver Template](#).

Dados da Pré-publicação

Identificação:

Apresentação: [Ver exemplo...](#)

Nome Fantasia: [Saiba Mais...](#)

Figura do Logo: [Saiba Mais...](#)

Tamanho ideal recomendado: 180 x 90 pixels.

Campos em vermelho são obrigatórios.
ATENÇÃO: essa tarefa pode demorar alguns minutos.

Figura 4.33: Tela de entrada da pré-publicação.

Além disso, o leiaute da página de apresentação dos dados publicados segue um *template* pré-definido que pode ser visto na Figura 4.34. Observe que há um espaço reservado para o logo, o nome da fantasia, uma busca textual, uma área de apresentação dos termos, que podem ser diretamente acessados pelas letras e ordem alfabética.

Ainda na Figura 4.34 veja que também existe duas opções adicionais de navegação, a *hiper-*

Figura 4.34: *Template* de publicação de produtos do **e-Termos**.

bólica e a *grafos*. Especificamente, tais opções representam outro importante diferencial do **e-Termos**, pois fornecem variadas formas de navegação e exploração do conteúdo de um determinado produto terminológico, mais interativa, dinâmica, fácil, e principalmente, situada em outro contexto de visualização, diferente da navegação hipertexto. Para exemplificar, observe sequencialmente as Figuras 4.35, 4.36 e 4.37 onde o verbete “eletrocardiograma”²⁹ é mostrado nas três diferentes formas de visualização. Veja que além da possibilidade de movimentação da estrutura o conteúdo do verbete é mostrado dinamicamente depois que o usuário repousa o *mouse* sobre ele.

Figura 4.35: Termo “eletrocardiograma” na visualização hipertexto.

Nesse contexto, outro relevante aspecto a considerar é que as diferentes formas de navegação

²⁹Este termo faz parte da nomenclatura do Dicionário de Cardiologia Invasiva, resultado de um projeto criado e publicado no **e-Termos** e pode ser acessado no endereço: <http://www.etermos.ufscar.br/produtos/produto39.html#letrae>.

Figura 4.36: Termo “eletrocardiograma” na visualização hiperbólica.

Figura 4.37: Termo “eletrocardiograma” na visualização em grafos.

e visualização torna o **e-Termos** um portal de consulta de conhecimentos específicos, assim como acontece nos projetos IATE e TERMITE (vistos na Tabela 4.4). Vale observar que tais interfaces de consulta e navegação também permitem a busca por proximidade por meio do uso de curingas e operadores lógicos, com intuito de facilitar ainda mais a pesquisa.

4.3.7.3 A exportação de dados em formato TBX

A última funcionalidade atualmente disponível no Módulo 6 é a ferramenta de *exportação* de conteúdo de um determinado produto terminológico para o formato TBX (*TermBase eXchange*), um padrão de intercâmbio internacional amplamente aceito e recomendado pela comunidade. Assim, esta ferramenta representa a mais concreta possibilidade de difusão e intercâmbio dos produtos terminológicos desenvolvidos no **e-Termos** já que permite a permuta de dados com outros sistemas.

De forma semelhante, o usuário também fornece como entrada um *Modelo de Verbetes*, tendo como saída o download de arquivo com extensão `.tbx` com os dados dos campos constantes no Modelo.

4.4 O e-Termos como Ambiente Colaborativo Web

Depois de expor as principais funcionalidades dos módulos de trabalho é possível afirmar que o **e-Termos** se destaca como categoria de aplicação *Groupware*, já que o mesmo procura implementar os processos de apoio à colaboração de indivíduos que participam de uma equipe, possibilitando a produção em conjunto e a troca de informações acerca de uma atividade específica (a terminografia). Ou seja, um sistema que dá suporte ao trabalho realizado por um grupo de pessoas e que proporciona benefícios como resultado dessa colaboração Vieira e Fortes (2004). Isso porque as etapas do trabalho terminológico, compreendidas pela implementação dos módulos do **e-Termos**, tendem a ter mais sucesso quando não há apenas uma pessoa envolvida no processo, visto que este trabalho, além de exigir um planejamento prévio implica necessariamente a permanência de uma equipe multidisciplinar. Assim, a atividade terminográfica, por natureza e em função das necessidades de comunicação, demanda o envolvimento de um grupo de trabalho, de uma equipe que divida tarefas e responsabilidades de modo organizado Krieger e Finatto (2004).

Dessa maneira, além das questões de pesquisa cobertas, esse é um dos principais motivos para a criação e implementação do **e-Termos** sob o paradigma colaborativo, já que o trabalho terminográfico, considerando todos os perfis profissionais envolvidos, as atividades específicas de cada etapa, os pormenores e dificuldades de cada estágio, a necessidade de intercâmbio de opiniões e idéias e a essência da comunicação entre os membros da equipe, serve de insumo e, ao mesmo tempo, subsídio para o estudo da ação colaborativa sobre o trabalho terminológico.

Mais especificamente, a colaboração no **e-Termos** é um processo em que dois ou mais agentes trabalham juntos para resolver metas comuns, implicando numa atividade coordenada, resultado da tentativa constante de construir e manter uma concepção compartilhada do problema.

Desse modo, seguindo as orientações e delineamentos das aplicações CSCW para a possível automatização das etapas desse trabalho em ambiente Web, o **e-Termos** procurou essencialmente revelar a percepção do efeito e do impacto da ação e do comportamento colaborativo no trabalho terminológico.

4.4.1 Ferramentas que favorecem a colaboração no Ambiente **e-Termos**

Do exposto até aqui, percebe-se que um ambiente colaborativo envolve: 1) um grupo de profissionais de diferentes perfis, objetivos, tarefas, necessidades e permissões, e 2) um conjunto de atividades que devem ser realizadas por estes profissionais de maneira colaborativa. Dado este grupo de pessoas e o conjunto de atividades que realizam, as atuais aplicações CSCW procuram combinar diversas funcionalidades para suportar a natureza colaborativa em um único ambiente Vieira (2006).

Tais funcionalidades são representadas por diferentes ferramentas computacionais, ou componentes de *software* que implementam, suportam e favorecem a colaboração em aplicações *Groupware*. Genericamente, essas ferramentas consistem de soluções computacionais que permitem o “trabalho invisível”, principalmente por meio do fluxo de informações e a capacidade de comunicação que as mesmas oferecem.

Nesse contexto, com intuito de promover e suportar a característica colaborativa no ambiente do **e-Termos** as ferramentas colaborativas mostradas na Tabela 4.6 foram adicionadas à sua arquitetura.

Tabela 4.6: Conjunto de ferramentas colaborativas disponíveis no **e-Termos**

Número	Nome da Ferramenta
1	e-Mail Categorizado
2	Fórum de Discussão
3	Chat (bate-papo)
4	Edição Colaborativa
5	Mural de Recados

Definições das Ferramentas Colaborativas

e-Mail Categorizado : mensagem eletrônica que deve ser classificada de acordo com seu conteúdo em diversos tipos de assunto. Cada assunto representa a categoria da mensagem e podem ser, por exemplo: discussão, requisição, agendamento de reunião ou solicitação de leitura de documentos. Essas categorias, no contexto das aplicações colaborativas, servem para organizar as comunicações entre os membros de uma equipe de trabalho além direcionar os mesmos ao ler a mensagem. Também é útil para estruturar as mensagens, principalmente para classificação e recuperação na Memória Corporativa;

Fórum de Discussão : ferramenta que promove discussões sobre um determinado tema ou questão entre os usuários por meio do envio de mensagens. Os fóruns de discussões basicamente possuem duas divisões organizacionais. A primeira faz a divisão por assunto ou tema e a segunda uma divisão desses assuntos em tópicos. As mensagens ficam indentadas e ordenadas decrescentemente por data, da mesma forma que os tópicos ficam ordenados pela data da última postagem. Os fóruns são muito utilizados em grupos de usuários com temas de interesse comum;

Chat (bate-papo) : ferramenta que permite a comunicação entre os membros de uma equipe em tempo real, em um espaço que permite a interação dos mesmos e a troca de mensagens;

Edição Colaborativa : é uma ferramenta que possibilita a edição de um determinado documento por mais de uma pessoa. A edição colaborativa pode ser usada em aplicações em que um grupo de pessoas pode editar documentos, gráficos ou textos de maneira conjunta. A edição pode ser síncrona, quando um usuário pode visualizar o que outro faz em tempo real, ou assíncrona, quando os usuários editam os documentos de maneira individual, e depois visualizam as alterações dos outros usuários;

Mural de Recados : é uma ferramenta que permite o envio de recados, avisos, lembretes ou qualquer outra mensagem curta que seja direcionada aos membros da equipe de trabalho. O mural de recados não representa um espaço estático, onde os usuários devem visitar freqüentemente, e sim um componente dinâmico e interativo, que se relaciona com cada usuário à medida que o mesmo utiliza a aplicação.

Dessa maneira, a inclusão dessas ferramentas na arquitetura do **e-Termos** evidenciou sua feição colaborativa. Entretanto, a exemplo das ferramentas lingüísticas, as ferramentas colaborativas mostradas na Tabela 4.6 não são exaustivas, podendo haver necessidade de inclusão de outras ferramentas dessa categoria no futuro.

Ressalta-se que as ferramentas colaborativas são empregadas em todos Módulos do **e-Termos**, juntamente com a participação dos usuários. Ferramentas como *e-mail Categorizado*, *Fórum de Discussão* e *Mural de Recados* são utilizadas em quase todos os Módulos, pois são instrumentos que facilitam e promovem a comunicação entre os usuários. No Módulo 1, por exemplo, tais ferramentas podem ser usadas pelos terminólogos e especialistas para questionar, discutir ou avaliar sobre um determinado conjunto de palavras chaves que serão usadas na compilação automática do corpús de especialidade.

No entanto, existem outras ferramentas que embora disponíveis em todos os módulos, possuem aplicações apenas em alguns. Por exemplo, no Módulo 3 a *Edição Colaborativa* pode ser usada pelos terminólogos para a edição e limpeza das *Listas de Termos* extraídas automatica-

mente. O *Chat* nesse contexto é muito importante para permitir que os usuários discutam sobre uma determinada *Lista de Termos* extraída ou sobre um método de extração empregado, bem como dos possíveis resultados das aplicações de outros métodos.

Já no Módulo 4, as principais ferramentas colaborativas são o *Fórum de Discussão* e a *Edição Colaborativa*. No uso destas ferramentas os usuários terminólogo e especialista podem discutir sobre a organização estrutural de uma ontologia, bem como suas diferentes variações, resultantes da elaboração de outras versões, e da inserção dos candidatos a termos na mesma. No Módulo 5, na elaboração da Base Definicional é utilizada a *Edição Colaborativa*, quando dois ou mais editores estão inserindo excertos definicionais sobre um mesmo termo.

Da mesma forma, é importante ressaltar que a distribuição das ferramentas e usuários ao longo da arquitetura do **e-Termos** também não é exaustiva e pode ser alterada no futuro.

4.5 Aspectos e discussões relativas ao desenvolvimento e implementação do **e-Termos**

Considerando as diversidades das tarefas envolvidas na gestão terminológica e os diferentes tipos de usuários e suas exigências, o desenvolvimento do **e-Termos** seguiu a metodologia *Extreme Programming – XP* Teles (2004), Jeffries et al. (2001).

Extreme Programming é uma metodologia de desenvolvimento de sistemas que busca assegurar que os usuários, ou “clientes” recebam a máxima atenção da equipe de desenvolvimento. O XP é composto por um conjunto reduzido de práticas de desenvolvimento que se organizam em torno de quatro valores básicos. Essas práticas possuem fortes inter-relacionamentos e atuam de forma harmônica e coesa para assegurar que os usuários sempre recebam um satisfatório retorno do sistema a ser desenvolvido.

A metodologia XP compartilha a premissa de que os usuários aprendem sobre suas necessidades, na medida em que é capaz de manipular o sistema que está sendo produzido. Com base no *feedback* do sistema eles reavaliam as suas necessidades e prioridades, gerando mudanças que devem ser incorporadas ao sistema. O aprendizado é importante, porque permite que o usuário direcione o desenvolvimento de modo que a equipe produza sempre aquilo que tem a maior prioridade.

O XP se baseia em quatro valores fundamentais: 1) **feedback**, 2) **comunicação**, 3) **simplicidade** e 4) **coragem**. Quando os usuários aprendem com o sistema que utilizam e reavaliam suas necessidades, um **feedback** é gerado para a equipe de desenvolvimento. Isto é, os usuários realimentam a equipe com alterações nas necessidades que ainda serão implementadas e, eventualmente, naquelas que já fazem parte do sistema. O **feedback** é um mecanismo fundamental

que permite o desenvolvimento e garante que a equipe direcione as suas atenções para aquilo que é mais necessário. Para que os usuários possam compartilhar seus aprendizados com a equipe, é necessário que utilizem permanentemente o valor da **comunicação**. A comunicação entre os usuários e a equipe permite que todos os detalhes do projeto sejam tratados com a atenção e a agilidade que merecem. O XP procura assegurar que a comunicação ocorra da forma mais direta e eficaz possível. Sendo assim, ele busca aproximar todos os envolvidos no sistema de modo que todos possam se comunicar face-a-face ou da forma mais viável.

A comunicação, embora seja essencial, não é suficiente para garantir que os usuários possam aprender durante a implementação do sistema. Também é necessário que a equipe compreenda e utilize o valor da **simplicidade**, cujos ensinamentos orientam implementar apenas aquilo que é suficiente para atender a cada necessidade de cada usuário. Ou seja, ao codificar uma funcionalidade é relevante preocupar apenas com os problemas atuais e deixar os problemas do futuro para o futuro. Isso permite que o usuário explore a funcionalidade implementada, gerando valor para ela e tornando viável que o seja retornado para a equipe o quanto antes.

Dado que o sistema é desenvolvido de forma incremental, a equipe estará continuamente fazendo a manutenção do mesmo, bem como criando novas funcionalidades ao longo do tempo de desenvolvimento. Em muitos casos, devido às essas alterações o risco de gerar falhas no sistema aumenta. Por esta razão, a equipe precisa ser **corajosa** e acreditar que, utilizando as práticas e valores do XP, será capaz de fazer o software evoluir com segurança e agilidade.

Dessa forma, para ser desenvolvido com a metodologia do XP, os sistemas devem atender as seguintes características:

- Possuir requisitos e necessidades que se alteram com frequência;
- A implementação deve seguir a orientação a objeto;
- Possuir pequenas equipes de desenvolvimento;
- Poder ser desenvolvido de maneira incremental (ou iterativa), em que o sistema começa a ser implementado logo no início do projeto e vai ganhando novas funcionalidades ao longo do tempo.

Dado as quatro características acima, o **e-Termos** se encaixou nesta metodologia porque seu desenvolvimento foi realizado: 1) de forma incremental, módulo por módulo, ao longo das etapas do trabalho terminológico, 2) as necessidades dos usuários e seus requisitos alteraram ao longo de projeto, pois ao entender a automatização novas funcionalidades foram solicitadas; 3) foi implementado utilizando a orientação a objetos e 4) possui pequena equipe de desenvolvimento, sendo representado pelo autor desta pesquisa e outros alunos pesquisadores que implementaram pequenos ajustes ou classes.

4.5.1 A escolha da Web como ambiente de desenvolvimento do **e-Termos**

A decisão de utilizar a Web para a implementação do **e-Termos** veio da necessidade de sanar os problemas das reuniões do grupo de trabalho, principalmente as que envolvem a participação dos especialistas do domínio envolvidos na pesquisa terminológica, dado que estas reuniões consomem bastante tempo e são muito caras, pois na maioria das vezes é necessário o deslocamento entre cidades.

Além disso, a Web agiliza o trabalho de checagem dos especialistas, pois o mesmo pode realizá-lo no momento que mais lhe convier. Também, tal decisão foi baseada na proposta da ISO-TMS ISO 704 (2000) que indica a Web como sendo um critério básico de modelagem do Ambiente para o gerenciamento terminológico devido a sua plataforma de interligação. Outra vantagem é a idéia de “colaboração virtual” da equipe de trabalho que este Ambiente propicia, visto que o uso da Internet poupa o deslocamento e favorece a comunicação.

Sob este aspecto, destaca-se de maneira muito importante o trabalho basilar de Rice et al. (1996) que discute de maneira clara as diferenças, vantagens e desvantagens do uso de aplicações e sistemas Web em detrimento das aplicações em rede e/ou *stand-alone*, principalmente em relação à mudança de paradigma de uso da interface, o trabalho colaborativo utilizando a Web e as tecnologias utilizadas.

4.5.2 O modelo de desenvolvimento em camadas

Advinda da necessidade de independência dos módulos de trabalho, a arquitetura do **e-Termos** seguiu o modelo de desenvolvimento em camadas. A implementação em camadas é a chave para a autonomia entre os módulos, e possibilita a reutilização de código fonte e a facilidade de manutenção. Uma “camada” no contexto deste tipo de desenvolvimento de sistemas significa uma separação física e lógica das partes constituintes de um sistema, de maneira que haja a separação de responsabilidades.

Os sistemas que não seguem o desenvolvimento em camadas recebem o nome de aplicações monolíticas. Geralmente as aplicações monolíticas, ou de uma única camada, contém todas as funcionalidades em uma única parte, representada por uma grande quantidade de linhas de código e de difícil manutenção. Nos sistemas monolíticos a entrada do usuário, a verificação e validação dos dados, a lógica do negócio e o acesso a banco de dados estão presentes logicamente e fisicamente num mesmo local. Um esquema deste tipo de aplicação pode ser vista na Figura 4.38.

Figura 4.38: Representação de uma aplicação monolítica - uma única camada.

O crescimento da interatividade dos usuários por meio das redes de computadores exigiu o compartilhamento dos dados entre vários usuários simultâneos, fazendo surgir as aplicações em duas camadas. Nesta estrutura a base de dados foi colocada em uma máquina específica, separada das máquinas que executavam as aplicações. Neste modelo estrutural os aplicativos (parte de apresentação e negócios) são instalados em estações clientes contendo toda a lógica da aplicação. O maior problema neste modelo é o gerenciamento de versões do sistema em cada cliente, pois para cada alteração os aplicativos precisam ser atualizados em todas as máquinas. A Figura 4.39 apresenta o modelo em duas camadas.

Figura 4.39: Representação de uma aplicação em duas camadas, com a separação dos dados.

A aplicação com estrutura de três camadas surgiu com a chegada da Internet, que estimulou um movimento para separar a lógica de negócios (parte de negócios) da interface (parte de apresentação) do usuário. A idéia baseou-se na premissa de que os usuários da Web possam acessar as mesmas aplicações sem ter que instalar estas aplicações em suas máquinas locais. Nesse modelo estrutural, a lógica de negócios fica no lado Servidor, e não mais contida no Cliente. Dessa maneira, o aplicativo é movido para o Servidor e um navegador Web é usado como Cliente. O aplicativo é executado em servidores Web com os quais o navegador se comunica e gera a interface exibida no cliente, representando a camada de apresentação. Neste modelo a camada de dados também fica no Servidor, mas separada na lógica dos negócios. A Figura 4.40 mostra um modelo em três camadas.

Figura 4.40: Representação de uma aplicação em três camadas, com a separação das camadas de apresentação, negócios e dados.

A separação em camadas torna os sistemas mais flexíveis, permitindo que as camadas possam ser alteradas de maneira independente. As funcionalidades da camada de negócio podem ser divididas em classes, que por sua vez podem ser agrupadas em pacotes, podendo ser reutilizadas por diferentes partes do sistema ou por sistemas diferentes. Por este motivo, o modelo de três camadas tornou-se a arquitetura padrão para sistemas baseados Web.

4.5.2.1 O modelo *Model Control View* (MVC)

Conforme pôde ser visto na subseção anterior, o modelo de desenvolvimento em três camadas divide um aplicativo em três “partes”, deixando a lógica dos *negócios* no meio das duas camadas,

apresentação e dados. Neste modelo a maior parte do código escrito reside nas camadas de *apresentação* e *negócios*.

O modelo MVC (Model–Control–View)³⁰ é uma arquitetura de desenvolvimento que divide uma aplicação em três camadas, sendo elas: *Modelo*, *Controle* e *Visualização*. A camada *Modelo* representa os pacotes, as classes, os objetos e os dados de um aplicativo. A camada de *Visualização* representa a interface apropriada para a entrada e apresentação dos dados e informações aos usuários. A camada *Controle*, por sua vez, implementa a interatividade, por meio do processamento das ações tomadas pelos usuários na camada de *Visualização* e a consequente ação ou atualização na camada *Modelo*.

A idéia do modelo MVC é facilitar as atividades de desenvolvimento de aplicações, dividindo os sistemas em três camadas: *Modelo* (ou camada de negócios), *Visualização* (ou apresentação) e *Controle*. O fluxo de interação entre as camadas é intermitente. O usuário indica uma operação através da interface, ou camada de *Visualização*, que “interpreta” a operação e encaminha para a camada de *Controle*. Esta operação é processada pelo controlador que acessa os dados ou executa uma determinada ação conforme a lógica do negócio na camada *Modelo*, que por sua vez, atende o controlador e define como esses dados devem ser apresentados ao usuário. Baseado nessa interação o usuário prossegue então a sua próxima tarefa.

O controlador, ou camada de *Controle*, define o comportamento da aplicação, e é ele que interpreta as ações do usuário e as mapeia para chamadas do *Modelo*. Em uma aplicação Web as ações dos usuários são os cliques de botões, eventos, links, o preenchimento de formulários ou as seleções de menus na interface. As ações realizadas pelo *Modelo* incluem ativar processos de negócio ou alterar o estado da aplicação. Com base na ação do usuário e no resultado do processamento, o controlador seleciona uma visualização a ser exibida como parte da resposta à solicitação do usuário. Há normalmente um controlador para cada conjunto de funcionalidades relacionadas.

A separação de camadas do modelo MVC oferece vantagens significativas no desenvolvimento de aplicativos, possibilitando implementar com maior facilidade e clareza questões importantes como a integridade e fluxo de dados, controle de segurança, comunicação e visualização. Esta separação permite que aplicativos possam ser mais facilmente modificados e estendidos para atender a novas exigências, bem como possibilita que a interface com o usuário apresente várias visões de um só modelo, sem alterar a lógica do negócio. Entretanto, o modelo MVC também apresenta desvantagens, pois requer uma quantidade maior de tempo para analisar e modelar o sistema e exige pessoal especializado em programação e implementação de aplicativos. Além disso, o uso do modelo MVC não é aconselhável para pequenas aplicações, devido à relação de

³⁰Em português, o modelo MVC segue a tradução de Modelo Visualização e Controle.

custo benefício não compensar.

Para o desenvolvimento de aplicações Web, a divisão do modelo MVC é feita da seguinte forma:

- Camada de Controle: é a camada responsável pelo fluxo, qualidade e segurança das informações no sistema. Representa a ligação entre as camadas de negócio e apresentação;
- Camada de Modelo: são as atividades relativas ao próprio negócio de sistema, como as classes que se comunicam e controlam a base de dados, as regras e ações do sistema e suas relações;
- Camada de Visualização: são os formulários, relatórios, botões, menus e componentes da interface do sistema com os usuários. É usada para receber a entrada de dados e apresentar o resultado.

A Figura 4.41 apresenta a arquitetura computacional do **e-Termos** instanciada no modelo MVC. Observe que todo o sistema é dividido nas camadas de *Visualização*, *Controle* e *Modelo*, e os módulos do sistema têm suas respectivas separações em cada camada.

Figura 4.41: Arquitetura computacional modular em MVC do **e-Termos**.

No entanto, sobre esta arquitetura ainda resta mencionar que o modelo MVC foi conjugado com a técnica AJAX (Asynchronous JavaScript And XML). O AJAX pode ser visto com uma técnica de programação para o ambiente Web que reúne um conjunto de tecnologias existentes e cujo objetivo é deixar as aplicações mais interativas e dinâmicas. Está baseada na idéia da execução de requisições assíncronas entre o Cliente (browser ou navegador) e o Servidor Web,

melhorando as respostas e agilizando as atualizações da interface. As tecnologias envolvidas da utilização do AJAX são o JavaScript, o XML e o DOM (Document Object Model).

Neste contexto, as camadas de *Visualização* e *Controle* do **e-Termos** foram implementadas em AJAX, sendo utilizado as tecnologias, DOM, DHTML (Dynamic HTML), HTML, JavaScript, CSS (Cascading Style Sheets) e Applets Java para a *Visualização*. Em conjunto, tais tecnologias compreendem a orientação da W3C³¹ e formam uma estrutura estável, interligada, modular e de fácil manutenção. A camada de *Controle* foi inteiramente implementada em JavaScript e representa o controlador AJAX do sistema.

Por sua vez, a camada *Modelo* foi implementada sob o paradigma da orientação a objetos no conjunto das linguagens PHP 5, Java e Perl. Nesse contexto, vale dizer que também foi utilizado os pacotes do PEAR (PHP Extension and Application Repository), um repositório de extensões da linguagem PHP, dentre eles o que define a camada de persistência e acesso do banco de dados MDB2. Também pertencente à camada de Modelo, foi utilizado como Servidor Web Apache versão 2.2 e o banco de dados relacional MySQL versão 5.2.

É importante frisar que o **e-Termos** foi totalmente construído utilizando ferramentas, padrões, linguagens e componentes de acesso livre.

4.6 Considerações finais

Após a explanação do **e-Termos**, vale frisar que este Ambiente satisfaz várias exigências da pesquisa terminológica, que compreende o desenvolvimento de colaborativo de produtos terminológicos por meio da gestão participativa de uma equipe multidisciplinar e da disponibilização de recursos e ferramentas computacionais para a cobertura total das tarefas constantes na etapas deste trabalho.

Nessas considerações finais, gostaríamos de terminar este capítulo com a seguinte citação de Cabré (2005), sobre o futuro das aplicações terminológicas:

“Acreditamos que nesse próximo período passaremos do uso de ferramentas variadas, cada uma desenvolvida em um contexto, para a construção de plataformas integradas de trabalho que permitam criar automaticamente ou semi-automaticamente um glossário, desde sua concepção até sua edição, sem sair do mesmo posto de trabalho. (Cabré, 10)”

³¹W3C é a World Wide Web Consortium, um consórcio internacional de várias organizações internacionais que trabalham juntas para o desenvolvimento de padrões e recomendações para o desenvolvimento na Internet/Web.

Sistemas colaborativos: métodos de pesquisa e avaliação

Resumidamente, alinhado a tarefa de validação o objetivo da avaliação de um sistema é determinar se os resultados alcançados satisfazem os objetivos traçados no início do seu projeto Wainer (2007). Por meio da avaliação é possível identificar desvios, erros e problemas ocorridos durante seu desenvolvimento, e também, novas oportunidades não planejadas.

Atualmente, existem diversos métodos e técnicas para se avaliar um sistema. Embora seja necessário que a escolha do método seja focada naquilo que se pretende avaliar, é importante que não esteja centrada somente nos objetivos previamente estabelecidos e nas soluções alcançadas, mas também nos benefícios intangíveis que o sistema pode proporcionar.

Genericamente, os processos de avaliação de sistemas atuam em várias dimensões e tentam aferir diferentes aspectos tais como: a satisfação dos usuários, o desempenho do sistema, a usabilidade e acessibilidade da interface, a qualidade e aceitação do sistema e o impacto do sistema no comportamento do usuário e no grupo de trabalho. Para dar apoio ao processo de avaliação da qualidade de software, podemos usar a ISO 9126 ISO 9126 (2006) que traz seis métricas para a avaliação da qualidade de software: funcionalidade, confiabilidade, usabilidade, eficiência, manutenibilidade e portabilidade, subespecificando cada uma destas. Especificamente, os aspectos da acessibilidade nas interfaces devem levar em consideração as limitações e características da diversa população de usuários, com diferentes habilidades, necessidades e preferências, e embora não seja tratada nessa pesquisa pode ser objeto de estudo e melhoramento no futuro.

Apesar de requerer cuidados especiais, a satisfação do usuário é uma das medidas mais usadas na avaliação de sistemas de informação. Nesta dimensão, identificam-se fatores relacionados com a satisfação de características específicas do sistema, tais como: cores, funcionalidades, tempo de resposta, interface, qualidade dos resultados, confiabilidade, facilidade de uso etc., que juntas culminam na satisfação total do sistema Arouck (2001).

Contudo, a qualidade e aceitação do sistema, individual e conjuntamente, são influenciadas pela satisfação do usuário. A frequência do uso pode afetar, positiva ou negativamente, o grau de satisfação do usuário, e vice-versa. A utilização do sistema e a satisfação do usuário são antecedentes diretos do impacto individual; que por sua vez pode influenciar no impacto e no desempenho do grupo de trabalho. O impacto no grupo de trabalho pode também causar modificações no ambiente de uso.

Já o nível de impacto é a variável mais o difícil de definir e aferir de forma objetiva e não ambígua. Isso porque o impacto individual considera fatores abstratos e agrupam em um único universo a compreensão da informação, aprendizagem, eficácia das tarefas, melhora da produtividade individual e identificação e solução de problemas. Ampliando para a dimensão do impacto organizacional, esta dificuldade aumenta ainda mais, pois neste cenário é preciso valorar os ganhos de produtividade, a integração das tarefas, a união da equipe e o relacionamento dos usuários envolvidos.

Dessa maneira, fica clara a dificuldade de se avaliar um sistema principalmente se considerarmos os efeitos dos impactos causados por sua adoção. Neste sentido, um trabalho interdisciplinar se faz necessário para realizar as aproximações metodológicas que permitam identificar variáveis, modelar métodos e técnicas de avaliação, coletar e interpretar dados, bem como analisar e comparar os resultados.

Seguindo as orientações de Wainer (2007), é nesta linha metodológica que além de considerar os métodos de avaliação de sistemas, também nos apoiamos na área de pesquisa de marketing e administração aplicada a ciência da computação, principalmente nos estudos das técnicas e práticas relacionadas aos planos de pesquisa, envolvendo a definição de hipóteses, execução da investigação, coleta de dados, transformação, análise e interpretação de resultados. Como parte fundamental para este trabalho de doutoramento, a seção a seguir apresenta uma revisão bibliográfica destas técnicas e práticas.

Portanto, vale destacar que esta revisão é importante porque o processo de avaliação do **e-Termos** incluiu a administração de um *Questionário de Pesquisa* com 54 questões divididas em 6 seções que procurou cobrir cada um dos aspectos envolvidos na avaliação de um sistema. Este questionário será estudado detalhadamente na Seção 6.1.2.2.

5.1 Aspectos gerais dos métodos de pesquisa e avaliação

De acordo com Hair-Junior et al. (2006) os métodos de pesquisa e investigação em marketing e administração visam produzir resultados ou descobertas que poderão ser utilizados para o desenvolvimento de conclusões significativas sobre um dado tema. Neste sentido, um **processo de pesquisa** pode ser entendido como um conjunto de orientações que guiam uma pesquisa, e em geral é composto de três fases: a *formulação*, a *execução* e a *análise*, esboçados na Figura 5.1.

Figura 5.1: Fases de um Plano de Pesquisa.

A fase de *formulação* consiste na definição da hipótese e as instruções do processo de pesquisa. A hipótese aponta os elementos e materiais que irão produzir o resultado, e as instruções do processo de pesquisa correspondem aos passos que serão executados para conduzir aos resultados.

Uma hipótese é uma declaração formal de uma suposição incerta que procura explicar certos fatos, objetos ou fenômenos. Geralmente são descritas por eventos sistemáticos que podem ser testados com o uso de dados. Basicamente, uma hipótese afirma uma questão de pesquisa em termos mais específicos.

Conforme pode ser visto, a porção superior da Figura 5.1 compreende a observação, a descoberta e o desenvolvimento de hipóteses. Não há receita ou metodologias seguras de descoberta de hipóteses, de maneira que podem surgir a partir de idéias, suposições, palpites e intuições. Dessa maneira, uma hipótese é fundamentada a partir da organização de tais observações e idéias ou orientada por uma revisão bibliográfica Hair-Junior et al. (2006).

Depois do desenvolvimento da hipótese e da fase de formulação, inicia-se a fase de *execução*. Nesta fase ocorre a coleta de informações que representam fatos sobre variáveis da hipótese. Depois da coleta, tais informações são verificadas para detecção de erros, são codificadas e armazenadas para que a fase de análise seja feita de maneira rápida e conveniente.

A terceira fase é a *análise*, e representa o exame dos dados coletados, utilizando-se técnicas estatísticas para determinar se as descobertas sustentam ou não as hipóteses previstas na fase de formulação.

5.1.1 Planos de pesquisa científica

Os planos de pesquisa são vistos como resultado do processo de pesquisa e podem ser agrupados em três categorias: 1) *exploratório*, 2) *descritivo* e 3) *causal*. Um plano de pesquisa *exploratório* é útil quando as questões de pesquisa são vagas ou quando as hipóteses são insuficientes para orientar as previsões. É utilizada para desenvolver uma melhor compreensão sobre um fato ou objeto. A pesquisa descritiva descreve uma determinada situação ou momento. Geralmente, os fatos são descritos com a mensuração de um evento ou atividade a partir de estatísticas descritivas, o que inclui distribuição de frequências (quantidades), média e desvio padrão. Os planos de pesquisa causal são mais intrincados, e se dispõem a testar se um determinado evento causa outro.

A pesquisa exploratória é particularmente útil quando o profissional responsável dispõe de poucas informações para análise. Dessa forma, os planos exploratórios são orientados para a descoberta e não possuem a intenção de testar hipóteses específicas de pesquisa. Um exemplo de pesquisa exploratória é uma completa revisão de literatura de um tema ou assunto, pois pode ser muito vantajosa para melhor compreender uma hipótese ou questão.

Diferentemente dos estudos exploratórios, as pesquisas descritivas são confirmatórias e usadas para testar hipóteses. Em geral, são estruturadas e especificamente criadas para medir as características descritas em uma questão de pesquisa ou hipótese que, por sua vez, servem para guiar o processo e fornecer uma lista do que precisa ser mensurado. Neste sentido, os estudos descritivos envolvem todas as atividades da segunda fase do processo de pesquisa, de forma que a coleta de dados pode ser realizada por meio de entrevistas estruturadas ou questionários, contendo itens específicos que englobam aquilo que se deseja medir.

Vale destacar que as estatísticas descritivas utilizadas nesta categoria de pesquisa são baseadas em mensurações de uma amostra que representam uma população, cujas características são inferidas a partir desta amostra. Estas estatísticas incluem classificações (melhor/pior), contagem ou distribuição de frequência (quantidade), médias de grupo e correlações.

Já os estudos causais testam se um evento causa efeito em outro. Ou seja, a relação causal significa a mudança de um evento qualquer provoca alterações em outro evento. A causalidade neste caso sugere que mudança em X (a causa) faz com que ocorra uma mudança em Y (o efeito).

Os estudos causais são de grande auxílio para os pesquisadores responsáveis pela análise e apuração dos resultados, pois permite a elaboração de um mapa que expõe os possíveis acon-

tecimentos em caso de mudança, em um determinado ambiente. Em contrapartida, os planos causais exigem uma execução muito precisa, tornando-os muito complexos, de longa duração e onerosos Hair-Junior et al. (2006).

5.1.2 Experimentos de Laboratório e de Campo

Os profissionais ligados à pesquisa científica podem escolher entre dois tipos experimentais: os de **laboratório** e de **campo**. Os experimentos de laboratório manipulam as variáveis da hipótese dentro de um ambiente artificial, com o controle máximo. Estes experimentos são mais precisos cientificamente e possuem alta validade interna, o que significa dizer que qualquer diferença observada é de fato o resultado da manipulação experimental. O experimento de campo, por sua vez, é conduzido em um ambiente natural e as manipulações das variáveis são controladas em um contexto que não afeta a situação real. Isso faz com que os resultados tenham uma grande validade externa, visto que há maior chance de serem representativos da população.

Neste cenário, as vantagens e desvantagens de cada tipo de experimento se ancoram na seleção entre a alta validade interna ou externa. Embora seja suscetível à influência de outras causas e variáveis alheias ao experimento, o ponto forte dos experimentos de campo é o ambiente real. Por outro lado, tais influências não surgem nos experimentos de laboratório. Entretanto, os mesmos fatores que criam a alta validade interna tendem a reduzir a externa. Assim, é impossível recomendar o melhor ou mais adequado tipo de experimento, visto que tal seleção pode variar em cada situação de pesquisa.

5.1.3 Natureza dos dados

Independentemente do plano adotado, os projetos de pesquisa sempre envolvem a análise de dados. Os **dados** são simplesmente informações sistematicamente organizadas e registradas com a intenção de descrever e representar os fatos que sustentam uma hipótese ou questão de pesquisa. O objetivo dos pesquisadores é investigar as possíveis relações entre estes fatos, produzindo resultados específicos, que uma vez interpretados, podem gerar conhecimento.

Os dados podem ser caracterizados de várias maneiras. Uma forma importante é a distinção entre dados **objetivos** e **subjetivos**. Os dados objetivos são pré-definidos e independentes das opiniões particulares de qualquer indivíduo, representados por um conjunto de opções que se pode alterar. Os dados subjetivos são de fato as opiniões dos indivíduos. Entretanto, quando são agregados tornam-se relativamente mais objetivos, já que os números totais não representam apenas a opinião de uma pessoa.

Os dados também podem ser descritos como **primários** e **secundários**, com base em suas fontes. Os dados primários são aqueles coletados especificamente para o propósito da pesquisa.

Neste caso, os pesquisadores estão envolvidos com todos os aspectos da transformação dos dados, incluindo a coleta, organização, codificação, verificação de erros, representação, análise e interpretação. Os dados secundários, por sua vez, são aqueles coletados por outra pesquisa, com outro propósito, que são reaproveitados em um determinado projeto. Geralmente, são obtidos com mais rapidez, o que resulta em economia de tempo e dinheiro. Contudo, podem não ser adequados ao propósito de pesquisa intencionado, o que sugere uma precaução no seu uso.

Outra maneira de classificar os dados é em **quantitativos** e **qualitativos**. Os **dados quantitativos** são aqueles representados por números que consistem nas mensurações diretas das propriedades de algo que precisa ser quantificado. Como são diretamente descritos por números, os dados estão em um formato voltado para a análise estatística. Os dados qualitativos representam descrições de coisas sem a atribuição direta de números, sendo, neste caso, mais abstratos, colhidos a partir de algum tipo de entrevista e/ou questionário subjetivo.

Conforme salientado por Wainer (2007) as vantagens do uso de dados quantitativos são a estrutura, representatividade e uniformização das respostas. Entretanto, esse tipo de abordagem impede a busca ou registro em profundidade de uma determinada variável. Por outro lado, os dados qualitativos permitem a descoberta de informações mais profundas e fora do rol de respostas previamente oferecidas, dando a possibilidade de revelar questões e opções não contempladas.

5.1.4 Abordagens de Coleta de Dados

Para descreverem os fenômenos do mundo por meio das pesquisas, os pesquisadores necessitam dos dados. Os dados são coletados com a adoção das seguintes abordagens: *observação* ou *entrevistas* e/ou *questionários*.

O tipo e a quantidade de dados a serem coletados dependem da natureza do estudo e dos objetivos da pesquisa. Se exploratória, a abordagem pode ser observacional, com entrevistas pessoais, ou a observação de comportamentos ou eventos. Se descritiva ou causal, é necessário uma ampla quantidade de dados que devem ser obtidos a partir de *Levantamento de Dados*¹ de grande escala ou pelo acesso a bancos de dados. Os dados quantitativos são geralmente obtidos a partir dessa abordagem e são utilizados para pesquisas que possuem um problema ou foco bem definidos.

A informática tem revolucionado as ações de coleta de dados, pois grandes quantidades de dados, tanto qualitativos quanto quantitativos, podem ser obtidos e integrados a bancos de dados de um modo relativamente rápido e a um custo baixo, principalmente se comparado aos métodos tradicionais, como por meio de telefone, correios (mala direta), fax e entrevistas. Esses novos métodos incluem questionários computadorizados administrados pela internet ou pela captação

¹A expressão Levantamento de Dados é conhecida no inglês como *Survey*.

eletrônica de dados por meio de *logs* ou registro do fluxo de informações.

5.1.4.1 Observação

Os dados observacionais são coletados a partir do registro sistemático da observação de pessoas (ou grupo de pessoas), eventos e objetos, obtidos pela ação humana ou eletrônica.

Uma abordagem observacional resulta em: 1) dados narrativos, quando aparecem na forma de descrições de comportamento por escrito ou em registro de áudio e vídeo; ou 2) numéricos, quando são obtidos pelo registro de eventos feitos por um observador humano, por meio de um questionário ou por um instrumento eletrônico que mapeie ações específicas.

Exemplos desta abordagem podem ser as **pesquisas etnográficas** e a **análise de conteúdo**. A pesquisa etnográfica geralmente envolve a interpretação do comportamento por meio da observação de experiências de vidas reais por longos períodos e produzem narrativas que descrevem o comportamento de uma pessoa. A análise de conteúdo obtém dados por meio do exame de um conteúdo adquirido ou pelos textos escritos, produzidos a partir de relatórios, cartas, depoimentos, anúncios e questões abertas (subjetivas) em questionários. Especificamente, por meio da exploração do texto, o pesquisador examina as frequências de palavras e identifica os temas o conteúdo e as principais características presentes no texto. Este resultado é usado para quantificar os dados qualitativos.

5.1.4.2 Entrevistas e/ou Questionários

O *Levantamento de Dados* a partir de entrevistas ou questionários é um procedimento para coleta de dados primários a partir de indivíduos selecionados, participantes da pesquisa. Os dados coletados podem variar entre crenças, opiniões, atitudes e estilos de vida até informações gerais sobre a experiência do indivíduo como, por exemplo, idade, nível educacional, número de filhos, renda, etc. As entrevistas ou questionários são utilizados quando o plano de pesquisa envolve a coleta de informações de uma amostra representativa de indivíduos. Deste modo, a principal diferença do método de observação é que nestes os participantes sabem claramente que estão participando do processo de pesquisa, podendo influenciar as respostas e criar tendências.

Questionários

Um questionário é um conjunto de perguntas criadas especificamente para coletar dados dos respondentes, especificamente desenvolvido para medir características relevantes para a pesquisa em questão. A principal vantagem dos questionários é que podem ser usados para coletar dados quantitativos de um número maior de indivíduos de uma maneira relativamente rápida e conveniente.

Para a criação de questionários uma série de atividades deve ser considerada, como: o desen-

volvimento geral do questionário, sua validação através de testes de simulação e a determinação do método pelo qual o questionário será administrado. Os valores para esta última são: auto-administrado ou administrado pelo pesquisador

Os questionários auto-administrados são respondidos sem a presença do pesquisador. Nesta situação, supõe-se que o respondente tenha conhecimento e motivação suficiente para completá-lo sozinho. Isso significa que o tema, o conteúdo e a forma de apresentação das perguntas devem ser atraentes para que o questionário seja devolvido todo respondido. O principal problema dos questionários auto-administrados é a perda de controle do pesquisador, já que nesta situação é impossível saber se quem respondeu o questionário foi mesmo o respondente pretendido ou se pediram opinião de terceiros, fatos que podem incluir tendências nas respostas. Diferentemente, os questionários administrados pelo pesquisador são respondidos com a pessoa presente, em um determinado limite de tempo, e também podem ser feitos por telefone ou eletronicamente.

Entrevistas

Uma entrevista ocorre quando o pesquisador argúi e fala diretamente com o respondente, fazendo perguntas e registrando as respostas em tempo real. As entrevistas são especialmente úteis na coleta de dados quando perguntas abertas, questões complexas, delicadas e de cunho pessoal não estão elencadas. A principal vantagem de sua aplicação é a possibilidade do pesquisador obter *feedback* instantâneo, partir do relacionamento com o entrevistado e da visão de suas reações.

5.1.5 Mensuração e Escala

A mensuração é um aspecto importante dos planos de pesquisa porque devemos mensurar corretamente o conceito que estamos examinando. Caso contrário fica difícil, se não impossível, comentar um comportamento ou fenômeno de maneira que as interpretações e conclusões não serão precisas, válidas ou confiáveis.

A mensuração de coisas concretas e palpáveis, tais como altura, peso, velocidade e tempo é mais comum e fácil, pois não são muito complexas. Por outro lado, é muito mais difícil medir atitudes, opiniões, percepções e impactos, pois não há definições precisas de tais conceitos, sendo necessário definir perguntas e escalas para medi-los.

Nos planos de pesquisa atuais é muito comum a necessidade de mensurar conceitos que podem variar desde uma natureza simples e concreta até extremamente complexa e abstrata. Portanto, é fundamental definir precisamente os conceitos examinados do processo de pesquisa para evitar ambigüidades em sua interpretação. Assim, quanto mais complexo e abstrato for o conceito mais explícita a definição deve ser.

Uma vez definidos os conceitos é preciso se certificar de que as mensurações serão apropriadas.

O processo de mensuração envolve a especificação de variáveis (questões) que servem como substitutos dos conceitos. Geralmente, uma variável representa um único componente de um conceito mais amplo, de maneira que um conjunto de variáveis mede um conceito. A identificação de variáveis representa a instância prática de um conceito, possibilitando a criação de escores numéricos para medir os conceitos em termos quantitativos. Neste sentido, a mensuração envolve a atribuição de números para uma determinada variável de acordo com regras pré-definidas que, por sua vez, devem refletir as características do conceito ou fenômeno que está sendo mensurado. Por exemplo, para medir o interesse de um determinado usuário em utilizar um *software*, é possível utilizar o número **1** para representar “**totalmente desinteressado**” e **5** para “**totalmente interessado**”, de maneira que a regra define que um número baixo significa algo que é relativamente sem importância e o número alto algo mais importante.

A mensuração é realizada através do uso de *escalas*. Uma escala é um instrumento de mensuração que pode ser *distinto* ou *contínuo*². Uma escala distinta mede somente a direção da resposta, por exemplo, uma escala com respostas de SIM/NÃO mede apenas a direção SIM ou NÃO. Outros exemplos dessa escala podem ser as perguntas com respostas de DISCORDO/CONDORDO ou GOSTO/NÃO GOSTO ou ainda MASCULINO/FEMININO. Já as escalas contínuas medem a direção e a intensidade das respostas. Por exemplo, além de medir CONCORDO/DISCORDO uma escala contínua pode medir tanto se uma pessoa concorda totalmente ou concorda até certo ponto. Além disso, a escala pode variar de três pontos, representando pouca variação de intensidade, até dez pontos, permitindo mensurar grande parte da variação de sentimentos evidenciada pelos respondentes.

Existem quatro escalas de mensuração: **nominal**, **ordinal**, **intervalar** e de **razão** Wainer (2007), Hair-Junior et al. (2006). A **escala nominal** usa números ou palavras específicas como rótulos para identificar e classificar os objetos, indivíduos ou eventos. Um exemplo desta questão é mostrada na Figura 5.2.

Você está satisfeito com a interface do sistema?

() Sim () Não

Figura 5.2: Questão com respostas em **escala nominal**.

Observe que o *objeto* é a interface do sistema, a característica medida é a *satisfação* do usuário e as categorias pré-determinadas são “satisfeito” ou “insatisfeito”, refletido na escala com dois pontos distintos SIM ou NÃO. Vale salientar que as categorias das escalas nominais não se limitam a somente duas, de maneira que a única exigência é que as mesmas devem ser

²Na estatística, escalas distintas também podem ser chamadas de *categorizadas* e as contínuas de *ordinais*

mutuamente excludentes e exaustivas de todas as possibilidades. Além disso, as escalas nominais representam o mais baixo nível de mensuração, de maneira que a análise estatística dos dados restringe-se à contagem do número de respostas, a distribuição de frequências e o cálculo de porcentagens.

Uma **escala ordinal** é uma escala de ranqueamento e coloca um conjunto de objetos em uma categoria pré-determinada e ordenada de acordo com um dado critério, como preferência ou importância. Essa escala permite ao pesquisador determinar se um objeto tem mais ou menos de uma determinada característica do que um outro objeto, mas não permite determinar o quanto esse objeto tem de mais ou menos daquela característica. Por exemplo, a questão na Figura 5.3.

Considerando a adoção de um novo Sistema, classifique os seguintes atributos de 1 a 4, com 4 sendo a razão mais importante para adoção do Sistema e 1 sendo a menos importante. Atribua diferentes números para cada item.

☐ Interface

☐ Funcionalidades

☐ Meio de Distribuição

☐ Preço

Figura 5.3: Questão com respostas em **escala ordinal**.

Observe neste exemplo que os pontos em uma escala ordinal não indicam distância equivalente entre as classificações, de maneira que a diferença entre 3 e 4 não necessariamente é a mesma diferença entre 1 e 2. Dessa forma, é possível saber que 3 é melhor que 2, mas não o quanto. É importante ressaltar que os dados ordinais permitem um nível mais alto de análise estatística, já que é possível calcular a mediana e porcentagens, além das estatísticas de correlações de Spearman Hair-Junior et al. (2006).

Diferentemente da ordinal, uma **escala intervalar** utiliza números para classificar objetos ou eventos de modo que a distância entre os números seja igual, de modo que as diferenças entre os pontos da escala podem ser interpretadas e comparadas para a obtenção de algum significado. Assim, uma escala intervalar tem todas as qualidades das escalas nominais e ordinais, com o adicional de que a diferença entre as classificações 3 e 4 é a mesma que existe entre as classificações 1 e 2, possibilitando uma comparação.

As escalas intervalares são utilizadas em planos de pesquisas que medem conceitos como atitudes, percepções, sentimentos, impactos, opiniões e valores por meio de escalas de classificação que envolvem o uso de afirmações em um questionário, acompanhadas de categorias pré-codificadas que especificam o nível de concordância ou não de um determinado respondente. A Figura 5.4 mostra um exemplo de questão que utiliza esta escala.

O Sistema melhora a gestão do trabalho?				
Discorda totalmente	Discorda	Não concorda nem discorda	Concorda	Concorda totalmente
1	2	3	4	5

Figura 5.4: Questão com respostas em **escala intervalar**.

Os dados obtidos pelas escalas intervalares são passíveis de serem submetidos aos mesmos cálculos das escalas nominal e ordinal, mas também podem receber cálculos mais sofisticados como desvio padrão e coeficiente de correlação de Pearson Hair-Junior et al. (2006).

Uma **escala de razão** oferece o mais alto nível de mensuração, pois permite a fixação de uma única origem ou ponto zero na escala, possibilitando o cálculo de razões entre seus pontos. Um instrumento de medição de altura ou peso é um exemplo comum de escalas de razão porque possui pontos zeros absolutos, e permite a comparação de um determinado ponto da escala com outro. Assim, por exemplo, uma peça que mede 20 cm é duas vezes maior do que outra peça que mede 10 cm, ou que uma terceira peça que mede 5 cm, que é 4 vezes menor. As escalas de razão possuem todas as propriedades das escalas anteriores, além do ponto zero absoluto.

5.2 Aspectos Gerais dos Sistemas Colaborativos

Basicamente, o trabalho em grupo pode ser dividido em *Colaborativo* e/ou *Cooperativo*. Nesta divisão destacam-se duas vertentes principais. A primeira considera que o trabalho colaborativo está baseado na idéia de *colaboração* que, segundo Brna (1998), consiste em um conjunto de possíveis relações entre os participantes de um grupo de trabalho que envolve o empenho mútuo dos mesmos em um esforço coordenado para realizar uma determinada tarefa ou solucionar um problema. Já o trabalho *cooperativo* é realizado através da divisão do trabalho entre os participantes, como uma atividade em que cada pessoa é responsável por uma porção da solução do problema.

A segunda vertente considera o trabalho *Colaborativo* e *Cooperativo* como semelhantes, sendo que para haver trabalho cooperativo não pode haver relação de hierarquia entre os membros da equipe de trabalho. Neste cenário, a relação entre os membros deve ser horizontal, de maneira que a divisão do trabalho deve ser realizada pelo senso comum, com a participação de toda a equipe de trabalho. Entretanto, como bem assinala Vieira (2006), isso não significa que em um ambiente de trabalho cooperativo não possa existir uma hierarquia flexível.

Nesse contexto, percebe-se que os conceitos da segunda vertente são semelhantes ao conceito de trabalho *Colaborativo* da primeira, de maneira que fica clara a participação de toda equipe,

em que os membros trabalham juntos, num empenho comum, seja na divisão das tarefas ou na solução de problemas. Desse modo, as pessoas de um grupo devem trabalhar juntas de forma coletiva e colaborativa para: 1) tomar decisões, 2) compartilhar informações e 3) coordenar ações no âmbito do trabalho a ser desenvolvido Ehrlich (1999).

5.2.1 Principais características

Desde o seu surgimento, o suporte ao trabalho colaborativo em ciência da computação já passa dos 20 anos. No final dos anos 70 houve uma crescente preocupação em aumentar a produtividade das empresas e organizações, onde a maior parte do trabalho era feita em equipe. Nessa época, iniciaram-se as pesquisas sobre o comportamento dos grupos ou equipes de trabalho desempenhando atividades juntos, de maneira que os cientistas da computação aliaram-se a outros pesquisadores das áreas de humanas, tais como, sociólogos, psicólogos, antropólogos e educadores Grudin (1994), Santos (1998). Entretanto, somente em 1984 a sigla CSCW foi publicamente lançada, por ocasião da realização de um *Workshop* com o nome de *Computer-Supported Collaborative Work* organizado pelos pesquisadores Irene Greif e Paul Cashman cujos interesses individuais iam ao encontro do uso de tecnologias para apoiar o trabalho em grupo Grudin (1994), Eseryel et al. (2002), Bannon (1992).

Segundo Eseryel et al. (2002), o CSCW é um campo de pesquisa multidisciplinar que utiliza ferramentas e técnicas computacionais para prover e garantir o trabalho em grupo de várias pessoas que possuem tarefas relacionadas. Ainda, de acordo com Fluckiger *apud* Santos (1998) e Vieira e Fortes (2004), o CSCW pode ser entendido como o projeto e desenvolvimento de sistemas computacionais para suportar e aprimorar o trabalho em grupo de usuários envolvidos em tarefas ou objetivos comuns, bem como para a compreensão da utilização de tais sistemas.

De acordo com Ehrlich (1999), as aplicações que implementam o trabalho colaborativo, também conhecidas como *Groupware*, são diferentes daquelas aplicações que não suportam a colaboração, também chamadas de *single-user*. As aplicações *single-user* são desenvolvidas para apoiar as “tarefas” dos usuários, aquelas executadas no dia-a-dia, freqüentemente bem definidas, observáveis e concretas. Já as aplicações de *Groupware* são desenvolvidas para apoiar o “trabalho” dos usuários, em todo seu contexto, incluindo aí os aspectos comunicacionais e de relacionamentos com outros membros da equipe. Este “trabalho” é freqüentemente tácito, invisível e amorfo.

Nesse sentido, assim como salienta Bannon (1992), as aplicações de apoio colaborativo representam um grande esforço para se entender a natureza e as características deste “trabalho invisível”, resultado das atividades colaborativas, com o objetivo de desenvolver sistemas computacionais adequados que buscam compreender como as pessoas trabalham em grupo e o que

pode ser desenvolvido para apoiar tais atividades. Isso porque o trabalho em grupo, por meio das aplicações CSCW, permite a interação entre diferentes perfis individuais, podendo eliminar as distâncias, aproximar as pessoas, reduzir o tempo gasto no trabalho, aumentar a produtividade e diminuir os custos Vieira e Fortes (2004).

Dessa maneira, um sistema que efetivamente apóie o trabalho colaborativo deve considerar em seu projeto tanto o trabalho em si, objeto da tarefa colaborativa, quanto o aspecto da colaboração. Ou seja, a engenharia deste tipo de aplicação precisa reavaliar os processos do usuário para que os mesmos possam ser modificados ou melhorados, ao invés de simplesmente automatizá-los. É comum utilizar o termo *Groupware* como sinônimo de CSCW. Entretanto, a área de pesquisa CSCW estuda os conceitos relacionados aos trabalhos colaborativos de forma genérica e ampla. Um *Groupware* é constituído da tecnologia empregada para possibilitar o trabalho em grupo, com o foco mais restrito.

5.2.2 Aplicações e benefícios de sistemas colaborativos

De acordo com Santos (1998), as aplicações que apóiam a colaboração podem ser classificadas em três categorias:

Ferramentas : correspondem a programas básicos que são implementados para apoiar a colaboração. Podem fazer parte de um conjunto maior de outras ferramentas que compõem um *Ambiente*;

Programas Integrados : consistem em programas maiores cuja função primária é apoiar uma ou mais características do trabalho colaborativo individual;

Ambientes : coleções de ferramentas e tecnologias coordenadas para apoiar as tarefas colaborativas em um determinado âmbito de atuação. Além da coleção de ferramentas, estes Ambientes são caracterizados sob três aspectos: 1) uma função de controle, normalmente oferecida por meio da utilização de bancos de dados, 2) uma linguagem de programação ou *script* que permite a especificação dos processos de trabalho ou regras do negócio a serem implementados e 3) uma interface que permite a interação dos usuários por meio do acionamento de processos e ações, bem como da recuperação e armazenamento de informação.

Para serem consideradas completas, todas essas categorias precisam ter alguma habilidade para customizar a aplicação para um dado conjunto de necessidades dos usuários. Para isso, mesmo que em escala reduzida, tais categorias devem dar conta de três modelos essenciais: 1) o modelo de dados, que descreve conceitualmente as informações necessárias para representar a tarefa a ser executada, 2) o modelo de comunicação ou interação, que descreve como, quando

e quais informações serão trocadas e disponibilizadas aos membros do grupo, e 3) o modelo de interface, que descreve como as informações serão apresentadas a cada indivíduo.

De acordo com Mills *apud* Vieira e Fortes (2004), um dos principais desafios das aplicações CSCW é credenciar os integrantes de um grupo de trabalho a desenvolver suas tarefas, utilizando o ambiente colaborativo. Além disso, para que os aplicativos baseados em CSCW forneçam as funcionalidades necessárias para o adequado desempenho do trabalho, é preciso considerar ainda diversas características, tais como: a comunicação, coordenação, memória corporativa, cooperação e contexto Vieira e Fortes (2004). Tais características são descritas a seguir.

Comunicação : A comunicação é uma das principais características do trabalho colaborativo, pois é por meio dela que os membros da equipe se relacionam, discutem, compartilham informações e criam soluções para os problemas. Além disso, a comunicação também permite o agendamento de reuniões e encontros da equipe;

Coordenação : A coordenação consiste nos mecanismos de controle e gerência do trabalho da equipe, coordenando as atividades do grupo e de cada membro da equipe individualmente. Dessa forma, a coordenação envolve tanto as pessoas quanto seus papéis, além de ser responsável por administrar as permissões de cada recurso e grupos de usuários, proibindo ou liberando acessos;

Memória Corporativa : A memória corporativa também figura como uma característica essencial nas aplicações colaborativas. Corresponde no armazenamento de forma estruturada de tudo o que foi realizado no “ambiente de trabalho” da aplicação, possibilitando a rápida recuperação de informações acerca dos históricos das atividades, discussões, documentos e registros referentes ao trabalho diário da equipe. A memória corporativa é indispensável porque conhecer o que foi feito é essencial para poder continuar o trabalho Vieira (2006);

Cooperação : A cooperação diz respeito às atividades cooperativas que os membros da equipe de trabalho registram na Memória Corporativa. Dessa forma, a cooperação consiste no conjunto de atividades conjuntas que o grupo realiza para produzir, manipular e organizar informações e documentos. Seu principal propósito é estimular os membros do grupo a trabalharem juntos em um mesmo objetivo;

Contexto : O contexto nas aplicações colaborativas representa a capacidade de tornar os membros de um grupo e suas atividades visíveis a todos os outros participantes, com o objetivo de melhorar a interação entre os usuários e o entendimento da tarefa que cada um desempenha. Em outras palavras, o contexto representa a habilidade de perceber a presença de um usuário e suas atividades realizadas.

No cenário dessas características, desde que aplicadas de maneira planejada e utilizando os recursos computacionais adequados, as aplicações *Groupware* podem ser vistas como um valioso recurso à disposição das equipes de profissionais que realizam trabalhos em grupo.

5.3 Avaliações nos Sistemas Colaborativos

Esta seção apresenta os principais fatores relacionados com a avaliação de sistemas colaborativos e são fundamentais para o entendimento dos aspectos e características da avaliação do **e-Termos**.

5.3.1 Princípios e Dificuldades da Avaliação Colaborativa

A primeira tentativa de avaliação de aplicações colaborativas foi desenvolvida em 1997 por Baeza-Yates e Pino (1997) que sugeriram a criação de um conjunto de categorias para identificar o foco da tarefa de avaliação. Tais categorias tentaram relacionar aspectos como a qualidade do produto, o número de pessoas envolvidas e o tempo gasto para realizar o trabalho como um todo.

A partir desta tentativa, a avaliação de sistemas colaborativos tem sido um tema de pesquisa de grande interesse na área de CSCW. Desde 1988, quando Grudin salientou a importância, a necessidade e os desafios de se avaliar os sistemas desta categoria, várias pesquisas têm sido desenvolvidas Wainer e Barsottini (2004), Pinelle e Gutwin (2000a). Entretanto, apesar deste esforço, ainda não existe um consenso sobre as abordagens de avaliação de sistemas colaborativos, principalmente quando são consideradas todas as variáveis contextuais envolvidas neste processo, advindas da alteração da dinâmica de trabalho em grupo que a implantação de um sistema colaborativo causa no ambiente em que é utilizado Prates e Raposo (2006), Neale et al. (2004). Segundo Eseryel et al. (2002) e Pinelle e Gutwin (2000a), não existe uma abordagem única de avaliação e as questões a ela relacionadas ainda estão sem respostas.

De acordo com Herskovic et al. (2007), há cinco razões para a ausência de abordagens de avaliação colaborativa:

1. Os métodos de avaliação dos sistemas *single-user* ou monousuários não são aplicáveis aos sistemas colaborativos;
2. Os critérios de usabilidade utilizados nos trabalhos da interação usuário-computador não são aplicáveis aos sistemas colaborativos multiusuários;
3. A avaliação colaborativa possui custos altos e nem sempre os recursos financeiros estão disponíveis;

4. Os benefícios dos sistemas colaborativos aparecem depois de longos períodos de sua adoção, fazendo com que os métodos de avaliação também se estendam por muito tempo Neale et al. (2004);
5. Para os desenvolvedores e empreendedores dos sistemas colaborativos não está claro “o quê” se deve avaliar, qual método utilizar e quando se deve avaliar.

Do ponto de vista prático, o aspecto que mais se deseja aferir a partir da implantação de um sistema colaborativo é o impacto que ele causa nos ambientes onde é inserido Baeza-Yates e Pino (1997), Pinsonneault e Kraemer (1989). Contudo, a aferição de impactos não é clara, pois os critérios de avaliação podem ser muito amplos e variar desde o incremento da produtividade ou qualidade de uma determinada tarefa, ou a satisfação da equipe que executa o trabalho, até as mudanças organizacionais ou culturais causadas pelo uso dos sistemas colaborativos Wainer e Barsottini (2004). Dessa forma, as dificuldades de avaliação dos sistemas colaborativos e, conseqüentemente, a apuração dos impactos causados por sua adoção, é uma tarefa custosa e dependente de vários fatores externos. Tais fatores estão ligados principalmente ao contexto real de uso destes sistemas, e advém do fato que não é possível considerar o trabalho colaborativo como uma entidade isolada, justamente porque existe a essencial participação de um grupo de pessoas de diferentes origens, perfis educacionais, crenças metodológicas, aspectos sociais e culturais Pinelle (2000), Pinelle e Gutwin (2000a), Haynes et al. (2004), Herskovic et al. (2007), Stiemerling e Cremers (1998), Twidale et al. (1994), Dugan-Junior et al. (2003). Assim, fica clara a complexidade de avaliar um sistema colaborativo, considerando a influência que estes aspectos provocam na condução do trabalho.

Dada a ausência de uma abordagem certa e bem recomendada, é preciso considerar novas formas de se avaliar as aplicações CSCW, levando-se em conta os aspectos individuais e de grupo, bem como os efeitos organizacionais que estes sistemas provocam Eseryel et al. (2002). Nesse sentido, o que parece natural é seguir as orientações de Neale et al. (2004) que considera que o foco da avaliação dos sistemas colaborativos deve estar voltado para os conceitos do trabalho colaborativo, sempre alinhados à situação específica de uso e adoção dos CSCW, de maneira que a ênfase da avaliação possa estar continuamente envolvida com as questões da equipe de trabalho e sua organização. Tal visão representa o paradigma de implementação de sistemas colaborativos essencialmente centrados na visão social do trabalho com um todo, em detrimento dos sistemas *single-user*, centrados somente no usuário Neale et al. (2004), Haynes et al. (2004).

Como resultado da análise das publicações e resultados de pesquisas referentes à avaliação de sistemas colaborativos no período de 1990 a 1998, Pinelle e Gutwin (2000a) e Pinelle (2000) delinearam um modelo genérico das abordagens de avaliação. Em uma pesquisa semelhante Wainer e Barsottini (2004) analisaram as pesquisas em CSCW no período de 1998 a 2004 e

resumiram este modelo em quatro dimensões:

1 – Tipo de Avaliação : define o ambiente da avaliação e o nível de manipulação dos dados, podendo ser:

1. *Experimento em Campo:* executado no ambiente real (natural) do trabalho colaborativo com a manipulação de algumas variáveis;
2. *Estudo de Campo ou Estudo de Caso:* executado no ambiente real do trabalho colaborativo com a manipulação mínima das variáveis;
3. *Experimento de Laboratório:* executado em ambiente de laboratório com a manipulação de algumas variáveis;
4. *Exploratório:* executado em ambiente controlado sem nenhuma manipulação das variáveis.

2 – Características das Avaliações : representa o rigor e o controle da avaliação e define se a mesma será: *formal* ou *informal*, *quantitativa* ou *qualitativa* e se a manipulação de dados será rigorosa ou leve.

3 – Coleta de Dados : representa a maneira que os dados da avaliação serão coletados durante a sua realização, podendo ser: *observação do usuário, entrevistas, aplicação de questionários, discussão em grupo, gravação em vídeo e áudio, etc.*

4 – Foco da Avaliação : define o objetivo da avaliação ou “o quê” se deseja avaliar. Esta dimensão é dependente de cada trabalho ou sistema colaborativo implementado. Genericamente pode ser: *impactos organizacionais ou nas práticas de trabalho, qualidade final do produto, eficiência e desempenho da execução das tarefas, satisfação do usuário, taxa de cobertura das atividades e funcionalidades pelo sistema, usabilidade do sistema, interface, interação do usuário durante a utilização do sistema, características específicas do trabalho colaborativo, etc.*

Adicionalmente, estes modelos classificam as avaliações dos sistemas colaborativos pelas características próprias do sistema como, por exemplo, se as soluções de comunicação são síncronas ou assíncronas ou ambas, ou se o mesmo foi desenvolvido para fins de utilização em ambientes reais ou acadêmicos, bem como pelo posicionamento da avaliação durante o ciclo de desenvolvimento do sistema.

5.3.2 Classificação e Organização dos Métodos de Avaliação

A partir da organização desse modelo dimensional e sob uma perspectiva positiva da carência de abordagens para avaliação de sistemas colaborativos, surgiram diversos métodos e ou técnicas

que tentam amenizar as dificuldades e delinear soluções práticas para a tarefa de avaliação destes sistemas. Além de uma descrição detalhada de cada método especificamente desenvolvido ou adaptado para avaliação de sistemas colaborativos, o trabalho de Herskovic et al. (2007) sugere uma metodologia que auxilia os desenvolvedores a selecionar qual o melhor método aplicar em determinadas situações, a partir da especificação de características como: 1) o *status* de implementação do sistema, 2) custo de execução, 3) pessoas envolvidas na avaliação e no trabalho, 4) tempo previsto, 5) tipo e local da avaliação e 6) objetivos do sistema colaborativo e da avaliação em si. A seguir, apresentamos a lista não exaustiva destes métodos.

Groupware Heuristic Evaluation (GHE) : consiste na adaptação do método de Avaliação Heurística da usabilidade de sistemas *single-user* desenvolvido por Jakob Nielsen para os sistemas *Groupware*. É baseado em oito heurísticas bem definidas que devem ser examinadas pelos desenvolvedores e projetistas dos sistemas e que servem como características fundamentais dos mesmos Baker et al. (2001);

Performance Analysis (PAN) : consiste em um método de avaliação formal de sistemas *Groupware* na qual a aplicação a ser avaliada é modelada como uma atividade a ser executada por um conjunto de pessoas em um número de estágios pré-definidos. Geralmente, os critérios de avaliação deste método são a qualidade do resultado, o tempo gasto e a quantidade de trabalho necessário para realização de tal atividade, e devem ser analisados experimentalmente pelos avaliadores Baeza-Yates e Pino (2006);

Groupware Walktrough (GWA) : este método é uma adaptação da técnica de avaliação de usabilidade *Cognitive Walktrough* utilizada em sistemas *single-user*, que presta atenção especial ao projeto de interfaces que suportam a aprendizagem exploratória. Para a correta aplicação deste método nos sistemas de *Groupware* são elaborados: 1) uma descrição detalhada sistema que será utilizada para permitir o uso inicial, 2) uma atividade ou conjunto de tarefas representativas que o usuário pode executar, 3) um lista de ações necessárias para completar as tarefas e 4) o mapeamento do perfil dos usuários, incluindo suas experiências e expectativas de uso do sistema a ser avaliado. O(s) avaliador(es) acompanha(m) todos os usuários juntos ou individualmente, registrando os problemas encontrados durante a avaliação para posterior análise Pinelle e Gutwin (2002);

Perceived Value (PVA) : considerado com uma abordagem de baixo custo, este método tenta medir o *Valor Percebido* de sistemas de reunião eletrônica (*meetingware systems*) atribuído por seus usuários. Seu objetivo é aferir o impacto organizacional do sistema e o alinhamento entre suas capacidades e as expectativas dos desenvolvedores e usuários. Inicialmente, os desenvolvedores identificam os componentes de sistema para avaliação.

Depois, os usuários e desenvolvedores negociam juntos quais atributos serão avaliados pelos usuários, que após usarem o sistema, preenchem um mapa de avaliação observando se os componentes do sistema apóiam os atributos ou não. Usando essas classificações, uma métrica que representa o *Valor Percebido* é calculada Antunes e Costa (2004);

Collaboration Usability Analysis (CUA) : é uma técnica de avaliação focada nos aspectos de colaboração de uma equipe de trabalho que executa tarefas compartilhadas. Sua proposta é mapear cada ação colaborativa da equipe em um conjunto mecanismos de colaboração que podem ser representadas pelos componentes e elementos da interface do sistema, de maneira que possam ser criados diversos níveis de representação das situações colaborativas a serem estudadas e avaliadas. O resultado desta avaliação é um diagrama que mostra os detalhes das relações entre os componentes da interface, a noção de fluxo através deles e a distribuição de tarefas Herskovic et al. (2007);

Scenario-Base Evaluation (SBE) : é um método de avaliação que cria uma narrativa realista do ambiente de atuação de um sistema colaborativo, considerando o trabalho e os usuários envolvidos em um contexto real de uso do sistema. Pode ser visto como uma descrição detalhada de um cenário que inclui as tarefas, os atores, o contexto e os objetivos da avaliação Haynes et al. (2004);

Groupware Observation User Testing (GOT) : é uma técnica simples de avaliação de sistemas colaborativos em um ambiente de laboratório na qual os avaliadores observam os usuários executarem determinadas tarefas no sistema. A observação do avaliador pode ser monitorada para usuários que possivelmente encontram problemas na execução das tarefas ou estimulando os usuários a refletir sobre a execução de uma determinada tarefa do sistema. O objetivo desta técnica é focar nas ações de colaboração e analisar o trabalho dos usuários através de critérios pré-definidos Herskovic et al. (2007);

Cooperation Scenarios (COS) : este método busca capturar atividades específicas dos usuários no contexto de uso do sistema, criando pequenos cenários de cooperação de determinadas tarefas, que possibilitam a identificação do comportamento cooperativo, os usuários envolvidos, suas funções e a situação de uso. Para cada cenário criado, os avaliadores analisam a natureza e a execução da tarefa para verificar quais as mudanças e os benefícios alcançados com o novo sistema. Também pode ser usado para a modelagem e especificação dos requisitos do sistema, antes de sua implementação Stiemerling e Cremers (1998);

Human Performance Models (HPM) : derivado e adaptado da técnica *Keystroke Level Model (KLM)* utilizado para avaliar sistemas *single-user*, este método descreve como um

grupo de usuários interage com a interface dos sistemas colaborativos. Para avaliar o desempenho dos usuários, os avaliadores decompõem as interfaces do sistema em diversos ambientes de trabalho compartilhados, definindo cenários específicos focados nas ações colaborativas Herskovic et al. (2007);

E-Magine (EMA) : este método baseia-se em dois conceitos principais: 1) o sistema colaborativo deve se encaixar no ambiente em que será implantado e 2) a percepção do usuário é importante. O objetivo deste método de avaliação é apoiar grupos para avaliar a eficácia de aplicações groupware. O método baseia-se no compartilhamento de conhecimentos do grupo e segue uma estrutura modular. É composto de duas fases principais. A fase inicial produz um perfil inicial do grupo e fornece uma indicação dos possíveis desajustes entre o grupo e a aplicação. A formulação deste perfil inicial tem a vantagem de proporcionar uma clara orientação para futuras decisões sobre os instrumentos que devem ser aplicados na fase final do processo de avaliação Veld et al. (2003).

Visto que este trabalho de doutoramento desenvolveu e aplicou o método de *Cenário de Uso* (*Scenario-Used Evaluation - SBE*) para avaliação de um sistema colaborativo, descreveremos este método com mais detalhes na Seção 5.4.

5.4 Cenários de Uso na Avaliação Colaborativa

5.4.1 Principais características e aplicação dos Cenários de Uso

Os *Cenários de Uso* podem ser utilizados tanto na fase projeto e implementação quanto na fase de avaliação de um sistema colaborativo. Na fase de projeto, os *Cenários* são úteis para identificar os requisitos do sistema ou ajustar possíveis desvios de implementação que podem ocorrer durante o ciclo de vida de desenvolvimento. Na fase de avaliação, os *Cenários de Uso* têm sido sugeridos como um método formal e eficiente para sistemas já desenvolvidos Carroll (2000), Carroll (1995). A pesquisa de Stiernerling e Cremers (1998) é um bom exemplo do uso de *Cenários* tanto para o projeto quanto para avaliação de sistemas colaborativos.

A aplicação de *Cenários de Uso* permite que os aspectos vagos e imprecisos, tão comuns na natureza do trabalho colaborativo, sejam considerados em um ambiente único, a partir da sinergia entre os objetivos e atividades do sistema colaborativo em si e a experiência dos usuários e desenvolvedores. De acordo com Dzida e Freitag (1998) esta sinergia também contribui para superação dos possíveis desvios de entendimento que podem ocorrer no domínio da aplicação no período de tempo entre o projeto e a avaliação do sistema.

Contudo, a principal vantagem de aplicação dos *Cenários de Uso* em um processo de avaliação é a possibilidade concreta de instanciar em um mesmo panorama de execução o foco da avaliação,

evidenciando os objetivos e as intenções para os quais o sistema colaborativo foi planejado Carroll (1992). Neste contexto, um *Cenário de Uso* pode ser entendido como um conjunto de tarefas pertencentes a uma atividade maior que são sistematicamente organizadas em uma narrativa que descreve em detalhes a interação dos usuários participantes com o sistema desenvolvido. Além disso, o contexto do *Cenário* deve conter minuciosamente as intenções de avaliação, para que sejam aferidas ou coletadas durante ou após sua aplicação Carroll (1995). Dessa forma, um *Cenário de Uso* completo deve ter:

- 1 – **os atores** : são os indivíduos ou grupo de usuários que formam a equipe de trabalho do ambiente real do sistema ou os perfis de usuários que participam do *Cenário*;
- 2 – **a cena** : que descreve o contexto ou a situação específica em que o *Cenário* ocorre;
- 3 – **as tarefas** : que representam os objetivos ou a motivação de execução do *Cenário*;
- 4 – **as alegações** : que são as declarações dos participantes envolvidos no *Cenário* sobre os efeitos ou consequências de uso do sistema colaborativo objeto da avaliação. Tais alegações refletem as afirmações positivas ou negativas a respeito da experiência de uso do sistema ou do suporte tecnológico empregado em seu desenvolvimento e evidenciados na aplicação do *Cenário*.

Outra possibilidade de aplicação dos *Cenários de Uso* é em conjunção com o método **Groupware Walkthrough** conforme descreveram em Pinelle e Gutwin (2002). Neste caso, os *Cenários* foram utilizados como apoio para a criação das tarefas de avaliação da interface baseadas no contexto de uso da aplicação, fundamental para aquele método de avaliação.

De acordo com Haynes et al. (2004), o principal motivo de aplicação dos *Cenários de Uso* na avaliação dos sistemas colaborativos se justifica pela relação de custo/benefício oferecida por este método. Não que o custo de sua elaboração e aplicação seja baixo, conforme bem salienta Herskovic et al. (2007) em sua pesquisa, que classifica este método como custo médio, mas justamente pelo benefício proporcionado, advindo das possibilidades de simular contextos reais de uso, interação e colaboração entre os participantes de uma equipe de trabalho unida em busca de um objetivo comum, fundamento principal desta categoria de sistemas.

5.4.2 Criação dos Cenários de Uso

Duas das principais abordagens para a definição e elaboração dos *Cenários de Uso* são:

- 1) baseada no conhecimento dos desenvolvedores e projetistas do sistema a ser implementado a partir do estudo e análise dos requisitos e do ambiente de atuação do mesmo, 2) baseada na experiência e conhecimento dos usuários do sistema. A principal vantagem da primeira

abordagem é a praticidade e rapidez de elaboração do *Cenário*, já que é criado pelos responsáveis de implementação do sistema. Entre as desvantagens está o fato que a elaboração do *Cenário* distante dos usuários pode reduzir sua validade visto que é possível perder alguma informação importante, não mapeada pelos requisitos do projeto do sistema.

A segunda abordagem pode produzir um *Cenário* mais fiel e ajustado à realidade e o contexto de atuação do sistema, já que os usuários serão consultados. Entretanto, a desvantagem é o foco no grupo de usuários que pode produzir um grande número de informações e aspectos impossíveis de mapear no *Cenário* e, conseqüentemente, dificultar a fase de coleta de dados da avaliação. Dessa maneira, conforme assinala Haynes et al. (2004), o que nos parece mais coerente é iniciar o processo de construção do *Cenário* a partir das percepções dos projetistas e desenvolvedores, recolhendo nessa fase aquilo que foi requisitado e implementado, seguido pela consulta aos usuários reais do sistema, com intuito de expandir e validar o conteúdo do *Cenário* para garantir maior cobertura do domínio da aplicação.

Observa-se que as atividades envolvidas no processo de construção dos *Cenários de Uso* estão centradas não somente nas tarefas em si, isoladas no espaço e no tempo como uma unidade, e sim na instância de uma situação real que pode envolver um grupo de usuários e um conjunto de tarefas inseridas em um contexto de uso.

5.5 Considerações finais

Do exposto, é possível ressaltar que o uso do método de *Cenários de Uso* para a avaliação colaborativa pode aproximar os requisitos e os objetivos de um determinado sistema *Groupware* aos fatores humanos, sociais, culturais e organizacionais envolvidos em um ambiente colaborativo de trabalho e os impactos gerados por sua adoção.

Nesse contexto, considerando também a exposição dos conceitos apresentados neste capítulo referentes, aos aspectos gerais dos métodos de pesquisa e avaliação, dos sistemas colaborativos e as características que permeiam sua avaliação o Capítulo 6 mostra os resultados da avaliação empregada no **e-Termos**.

Avaliação do e-Termos

Paralelo ao processo de integração das características e exigências dos ambientes colaborativos ao **e-Termos**, considerando para isso todos os aspectos empíricos relacionados ao trabalho terminográfico, restam ainda as questões relacionadas à avaliação da qualidade do Ambiente, principalmente em relação às suas funcionalidades, a confiabilidade e a eficiência proporcionados aos profissionais que fazem terminologia.

6.1 Avaliação da qualidade e da colaboração no e-Termos

Além das questões de pesquisa cobertas neste trabalho de doutorado, o principal motivo da implementação do **e-Termos** foi instanciar o trabalho terminográfico sob o paradigma colaborativo. Isso porque o trabalho terminográfico, considerando todos os perfis de profissionais envolvidos, as atividades específicas de cada Etapa, os detalhes e as dificuldades de cada estágio, a necessidade de intercâmbio de opiniões e a essência da comunicação entre os membros da equipe serve de insumo e, ao mesmo tempo, subsídio fundamental para o estudo da ação colaborativa.

Seguindo as orientações e os delineamentos das aplicações CSCW para a (semi-)automatização das etapas desse trabalho em ambiente Web, a avaliação do **e-Termos** procurou essencialmente revelar o efeito e aferir a percepção do impacto da ação e do comportamento colaborativo sobre o trabalho terminológico, até então não explorados, pesquisados ou aferidos neste âmbito.

Nesse contexto, a avaliação do **e-Termos** e, conseqüentemente, a apuração dos resultados deste trabalho de doutoramento foi realizada a partir da simbiose dos conceitos revistos no

Capítulo 5, mais especificamente na seleção do plano de pesquisa mais adequado e nas dimensões das abordagens de avaliação de sistemas colaborativos.

6.1.1 Método de avaliação para o **e-Termos**: seleção e justificativas

Orientados por estas duas vertentes a instância da avaliação colaborativa do Ambiente **e-Termos** adotou o *Plano de Pesquisa Descritivo* juntamente com o método de *Cenários de Uso* aplicado em *Experimento de Laboratório* tendo como artifício de coleta de dados os *Questionários*. A Tabela 6.1 resume tais decisões.

Tabela 6.1: Seleções da avaliação colaborativa do Ambiente **e-Termos**

Plano de Pesquisa	Descritiva
Tipo de Avaliação	Experimento de Laboratório
Método de Avaliação Colaborativa	Cenários de Uso
Coleta de Dados	Questionários
Características da Avaliação	Formal/Quantitativa

Entretanto, no panorama dessas escolhas ainda resta o **foco da avaliação** que é aquilo que se deseja avaliar ou a hipótese e o fenômeno a ser verificado. Sob este prisma, é imperativo salientar que tais focos coincidem muito claramente com os objetivos e as contribuições esperadas desta pesquisa científica, cujo propósito envolve conceitos muito abstratos de difícil mensuração, baseados principalmente no impacto e relevância dos aspectos colaborativos na rotina e natureza do trabalho terminológico. Assim, a Tabela 6.2 mostra os focos desta avaliação.

Tabela 6.2: Focos da Avaliação Colaborativa do **e-Termos**

1	Percepção dos impactos nas práticas do trabalho terminológico
2	Melhora da Gestão Terminológica
3	Satisfação do Usuário
4	Colaboração e Interatividade do Sistema
5	Intenção de Uso e Recomendações do Sistema para Outros Usuários
6	Avaliação Geral do Usuário

Considerando que os itens da Tabela 6.2 podem ser vistos como um conjunto de hipóteses a afirmar, a seleção do *plano de pesquisa descritivo* se justifica pelo fato de que os mesmos são utilizados para testar hipóteses, na tentativa de averiguar e confirmar aquilo que se deseja medir. Além disso, a possibilidade de criar mensurações de uma amostra que retratam o comportamento de uma população é outro motivo de sua escolha.

Já a opção do *experimento em laboratório* foi taxativa neste processo, pois não havia outra possibilidade de uso do **e-Termos** senão instalá-lo especificamente para o processo de avaliação. Isso porque o **e-Termos** é um sistema novo e não havia nenhum grupo ou centro de pesquisa que já o estivesse utilizando. Dessa forma, o processo de avaliação foi o primeiro contato dos

usuários com o sistema, não havendo nenhum histórico anterior de uso que justificasse uma seleção alternativa para o tipo de avaliação.

Embora tenha o custo de preparação mais elevado e a exigência da participação de várias pessoas, a preferência pelo método dos *Cenários de Uso* se justifica pela possibilidade concreta de criar uma atmosfera muito favorável para a avaliação colaborativa do **e-Termos**, já que os *Cenários* permitem predizer a grande maioria das atividades do sistema e do trabalho terminológico envolvido, em um ambiente cômodo e natural para os avaliadores, mesmo estando em um laboratório e com relativo controle de variáveis. Neste sentido, os *Cenários de Uso* foram de fundamental importância para a formalização do processo de avaliação do **e-Termos**, principalmente pela oportunidade de estabelecer um ambiente de atividades muito parecido com o real e, conseqüentemente, para obtenção dos resultados deste trabalho.

A seleção dos *questionários* para a coleta de dados foi feita baseando-se fundamentalmente em três fatores: 1) necessidade de atingir todos os participantes da avaliação, 2) tempo de coleta e 3) custo de aplicação. Conforme visto na Seção 5.1.4.2 uma alternativa à aplicação dos *questionários* seria a *entrevista*. Todavia, este método é mais caro porque demanda considerável esforço de execução, justamente porque o tempo de administração da entrevista é bem maior que o de um questionário, tornando-se inviável o envolvimento de todos os participantes.

Outra alternativa de levantamento de dados seria a *observação*, por meio da gravação de áudio e vídeo durante a execução do *Cenário*. Contudo, além de possuir um custo muito elevado, os dados coletados nesta categoria são qualitativos, inconsistentes com opção de pesquisa descritiva, eleita para este processo de avaliação. Adicionalmente, a gravação poderia inibir os participantes, inserindo tendências na avaliação, podendo distorcer os resultados.

Quanto às características da avaliação serem *formais* e *quantitativas*, elas decorrem do fato das escolhas anteriores. Isso porque o universo formado para a execução desta avaliação, considerando o *Cenário de Uso* criado, o laboratório e a administração de um questionário sob a perspectiva de um plano de pesquisa descritivo, torna a avaliação um evento formal em que há a participação consentida de todos os indivíduos envolvidos. A característica *quantitativa* é resultado da seleção da coleta de dados ser essencialmente numérica.

Portanto, a aplicação do método do *Cenário de Uso* em um *Experimento de Laboratório* em conjunto com os *Questionários de Pesquisa* para coleta de dados em uma avaliação *Formal* foi bastante relevante, pois permitiu a criação de vários contextos de trabalho e a simulação da execução das diversas atividades terminológicas em uma esfera colaborativa, com envolvimento rico dos usuários.

6.1.2 Componentes da avaliação da colaboração no **e-Termos**

Do ponto de vista prático o processo de avaliação da colaboração no **e-Termos** é formado por três componentes: 1) um conjunto de documentos que compõem o Cenário, 2) o questionário utilizado para coleta de dados e 3) o grupo de usuários avaliadores que participaram da avaliação. As seções a seguir detalham cada um destes componentes.

6.1.2.1 O Cenário de Uso

O enredo do *Cenário de Uso* utilizado para avaliação do **e-Termos** está apoiado na idéia do trabalho colaborativo de um grupo de três indivíduos que representam o papel de cada perfil de usuário do sistema. No ato¹ único do *Cenário*, o objetivo do grupo é criar um minidicionário da área de *Revestimentos Cerâmicos* cuja nomenclatura possui seis verbetes com registro de dados completo². Vale ressaltar que o perfil *Usuário Final* (consulente) não foi incluído neste *Cenário* por não participar diretamente do processo de criação dos produtos terminológicos.

Para a criação deste *Cenário de Uso*, três cadernos de atividades foram elaborados, um para cada perfil de usuário, sendo eles: *Gerente de Projeto*, *Terminólogo/Linguista* e *Especialista do Domínio*. Cada caderno contém uma sequência de tarefas específicas ao respectivo perfil encenado, que deveria ser executada utilizando o **e-Termos**. O conjunto de todas as tarefas dos três cadernos constitui o trabalho completo necessário para a gestão e criação de um produto terminológico.

Durante a execução do *Cenário*, cada participante deveria realizar as tarefas expostas nos cadernos de atividades de seu respectivo perfil, trabalhando junto e colaborativamente com seus colegas de grupo para a elaboração do minidicionário.

Todos os cadernos de atividades são auto-explicativos, e contém nove seções. A primeira seção apresenta o **e-Termos**, contextualizando o indivíduo no universo do sistema, evidenciando seus objetivos e explicando sua estrutura e limitações. A segunda seção é dedicada exclusivamente ao conhecimento do próprio caderno, na qual foram destacados o padrão de formatação do texto e a descrição das tarefas, bem como a apresentação dos perfis dos participantes e do material a ser utilizado. A terceira seção traz as tarefas iniciais, que correspondem à Etapa Principal do sistema como, por exemplo, o cadastro de usuário, acesso ao sistema, alteração de senha, proposta e criação de um novo projeto e montagem da equipe. O restante das seções contém o conjunto de tarefas pertencentes a cada Etapa do trabalho terminológico, respectivamente, sendo que a quarta seção corresponde à primeira Etapa e a nona à sexta. Os Apêndices A, B e C apresentam na íntegra o três cadernos de atividades que compõem este *Cenário*.

¹Ato neste contexto está vinculado à idéia do ato teatral que representa as partes de uma peça de teatro.

²Conjunto completo de dados e metadados relativos aos verbetes que oferecem a maior cobertura possível, tanto de informações conceituais quanto linguísticas.

Adicionalmente, é importante salientar que os textos e o leiaute dos cadernos foram enriquecidos com diversos ícones coloridos e representativos, utilizados em diversos pontos do corpo do texto e na descrição das tarefas, principalmente para melhor orientar atividades ou descrever situações específicas. Tais ícones e suas descrições são mostrados na Figura 6.1. Além disso, objetivando deixar o conteúdo em uma linguagem mais clara, didática e de fácil entendimento, todos os cadernos foram revisados e validados por um jornalista e um *designer* gráfico.

<i>Representa a identificação da Etapa de trabalho de uma determinada tarefa.</i>
<i>Representa uma determinada tarefa a ser executada.</i>
<i>Representa uma chamada de ATENÇÃO! ou alerta.</i>
<i>Representa uma dica para execução de determinada tarefa.</i>
<i>Representa a indicação do material a ser utilizado.</i>
<i>Representa a ação de envio/leitura de um Recado.</i>
<i>Representa a ação de envio/leitura de e-Mails.</i>
<i>Representa a ação de entrada no Chat (Bate-Papo)</i>

Figura 6.1: Ícones utilizados nos cadernos de atividades.

A descrição das tarefas que deveriam ser executadas pelos participantes foram organizadas e dispostas sequencialmente em cada seção/etapa, obedecendo a um padrão de formatação previamente definido que contém os seguintes dados:

1. o número da Etapa à qual a tarefa pertence;
2. uma descrição ou orientação para a execução da tarefa;
3. uma dica ou ajuda especial quando pertinente;
4. o material exigido para realização da tarefa e o local de onde encontrá-lo;
5. uma chamada importante de atenção com a palavra “ATENÇÃO” destacada em vermelho, quando necessário;
6. comentários adicionais que podem auxiliar o entendimento da tarefa.

2	<i>Segunda Etapa</i>
<i>Fazer Upload dos arquivos de textos</i>
<i>Acesse menu “Textos”.</i>
<i>Arquivos:</i> <ul style="list-style-type: none"> • Arquivo “<i>texto1.txt</i>” e “<i>texto2.txt</i>” localizados no CD • Não selecionar nenhuma classificação
ATENÇÃO! <ul style="list-style-type: none"> • Você pode clicar no link <i>Extensões</i> para ver quais extensões são permitidas.
<i>Comentários: faça o upload de um texto por vez.</i>	

Figura 6.2: Descrição de uma tarefa no caderno de atividades.

A Figura 6.2 mostra um exemplo de descrição de uma tarefa disposta no caderno de atividades.

Também como parte do conteúdo do *Cenário*, os cadernos de atividades prevêem diversas situações e tarefas de colaboração entre os membros do grupo, na qual são orientados a usar as ferramentas comunicacionais disponíveis no sistema, tanto as síncronas (bate-papo instantâneo) quanto as assíncronas (e-mails e mural de recados). Além disso, para favorecer a colaboração, em vários momentos fora do contexto de uma determinada tarefa, o *Cenário* procura estimular constantemente o relacionamento e a troca de opiniões dos participantes do grupo. A Figura 6.3 a e b mostra dois exemplos destas situações.

a)		<i>Iniciar Chat com os usuários Terminólogo/Linguista e Especialista</i>
	<i>Clique no nome da pessoa com deseja conversar no lado direito da tela</i>
<i>Comentários: Dê boas vindas aos membros de sua equipe.</i>		
b)		<i>Enviar um Recado para o usuário Terminólogo/Linguista avisando a liberação da Segunda Etapa.</i>
	<i>Acesse menu “Recado”.</i>
<i>Comentários: Procure escrever recados com mensagens pequenas. Para mensagens maiores você pode utilizar o e-mail.</i>		

Figura 6.3: Descrição de tarefas que estimulam a comunicação da equipe de trabalho.

Como o processo de criação de produtos terminológicos é uma atividade longa e que demanda muito tempo de pesquisa, todo material necessário para a elaboração do minidicionário proposto nesta avaliação foi fornecido em um CD como parte integrante do *Cenário de Uso*. Desse modo, foram confeccionados três CDs contendo o material exigido em cada caderno, respectivo a cada perfil de usuário.

É essencial lembrar que além de fomentar a execução das tarefas no contexto do *Cenário*, o fornecimento do material faz parte do controle de variáveis, relativo à abordagem e a definição conceitual do método da avaliação. Isso porque, já que vários grupos de indivíduos fariam a avaliação, era fundamental que os dados fossem os mesmos para todos os grupos, uniformizando

os parâmetros de comparação e, consequentemente, o processo de avaliação.

6.1.2.2 Questionário de Pesquisa

Conforme dito anteriormente, para a coleta de dados desta avaliação foi aplicado um *Questionário de Pesquisa*, que devia ser respondido pelos participantes das atividades do *Cenário*, logo após a sua execução. Este questionário possui 54 questões distribuídas em 6 seções, cujo objetivo era colher respostas dos usuários que refletiam suas opiniões, sensações e percepções depois da experiência de uso do **e-Termos** no contexto do *Cenário*. Tais seções foram organizadas da seguinte maneira:

Seção 1 – Questões de Triagem e Abertura : as questões dessa seção têm o objetivo de oferecer boas vindas ao participante e identificar o seu perfil de usuário no contexto do *Cenário*, bem como conhecer seu interesse antes e depois de uso do **e-Termos**;

Seção 2 – Mensuração da usabilidade do Sistema e-Termos : abriga questões sobre a usabilidade do sistema, objetivando revelar um conjunto de características que podem ser usadas para descrever o **e-Termos**;

Seção 3 – Mensuração das contribuições do e-Termos : conjunto de questões cujo objetivo era medir um conjunto de aspectos que podem ser vistas como contribuições, tanto no aspecto técnico quanto no trabalho colaborativo, do uso do **e-Termos**;

Seção 4 – Mensuração da metodologia de trabalho : nesta seção as questões procuram aferir o nível de aceitação ou rejeição da metodologia do trabalho terminológico implementado pelo **e-Termos**, bem como sua similaridade em relação à metodologia de trabalho atualmente usada por cada participante;

Seção 5 – Mensuração do relacionamento com a informática e terminologia : aqui as questões procuram aferir e traçar a experiência dos participantes no uso de computadores, bem como identificar o nível de informatização dos trabalhos de terminologia desenvolvidos pelo participante;

Seção 6 – Questões de avaliação global : na última seção as questões procuram avaliar a opinião do participante sobre sua experiência de uso do **e-Termos** de uma maneira genérica, considerando todos aspectos positivos e negativos julgados por ele.

Do conjunto de 54 questões, apenas uma é subjetiva, pertencente à Seção 6. Esta questão permite o participante descrever livremente suas percepções e impressões quanto à utilização do **e-Termos**, bem como expressar outras opiniões e sentimentos não cobertos pelo questionário.

O restante das questões são objetivas, sendo que 46 têm respostas em escalar intervalar com cinco possibilidades, variando as opções de 1 a 5, em que 1 representa a resposta mais negativa e 5 a resposta mais positiva; e 7 questões com opções de resposta categorizadas, assim distribuídas: 4 questões com opções “SIM” ou “NÃO”; 1 questão com 3 opções de resposta; 1 com 10 opções e 1 de múltipla escolha. A metodologia de elaboração deste questionário seguiu os pressupostos da pesquisa em administração de Hair-Junior et al. (2006).

O Apêndice D apresenta o *Questionário de Pesquisa* aplicado no processo de avaliação e execução do *Cenário*.

6.1.2.3 Grupos de usuários avaliadores

Após a elaboração dos cadernos de atividades e do questionário para a coleta de dados, iniciou a tarefa de seleção dos atores que iriam participar do processo de avaliação e execução do *Cenário*. Tal seleção seguiu critérios bastante restritivos. Primeiro porque embora o **e-Termos** seja um sistema Web e possa ser utilizado por qualquer internauta, seus usuários não pertencem ao público geral da Internet e, geralmente, o perfil do utilizador é diferenciado e específico, limitando-se quase que exclusivamente ao universo da Terminologia, da Lexicografia e da Tradução. Segundo porque tendo o nicho de usuários sido definido, era essencial escolher terminólogos e/ou lexicógrafos com alguma experiência, já que estes seriam os principais avaliadores, de fundamental importância para uma avaliação comparativa, principalmente em relação a outras iniciativas de automatização da gestão terminológica. Terceiro, servindo de contraponto da anterior, também era primordial a participação de profissionais jovens, em início de carreira, justamente para se obter opiniões e sugestões de uma nova geração de terminólogos e lexicógrafos, mais acostumada com o advento da Internet e das ferramentas adaptadas para a Terminologia e Lexicologia.

Nessa perspectiva, a seleção dos participantes da avaliação do **e-Termos**, começou pelo mapeamento de diversos centros de pesquisa em Terminologia e Lexicologia no Brasil. Atualmente, o mapa da terminologia e lexicologia no Brasil conta com 22 centros de pesquisa³ que implementam diversos projetos de investigação e desenvolvem produtos terminológicos. Importa destacar que estes centros representam apenas comunidades científicas, estando tipicamente localizados em universidade e institutos de pesquisa. Portanto, neste recorte não estão consideradas as empresas ou instituições particulares que fazem terminologia em seu ambiente privado.

Uma vez selecionados, tais centros receberam uma carta convite que, em resumo, apresentava o **e-Termos** e os convidava para a avaliação. Ao todo foram selecionados seis centros de pesquisa de reconhecido valor, experiência e tradição em Terminologia, Lexicologia e Tradução, situados

³Esta contagem foi feita a partir das informações dispostas no portal do diretório de grupos de pesquisa do CNPq no endereço <http://dgp.cnpq.br> em Março de 2009.

nas cidades de São Carlos/SP, Uberlândia/MG, São José do Rio Preto/SP, Porto Alegre/RS e Brasília/DF. Dos seis centros contatados, dois não retornaram o contato (Uberlândia/MG e Brasília/DF) e quatro se interessaram em participar. Entretanto, dos quatro centros interessados, apenas três puderam realizar a avaliação, sendo que um deles (São Carlos/SP) não conseguiu agendamento. Assim, dos três centros participantes dois estavam situados em Porto Alegre/RS e um em São José do Rio Preto/SP.

A partir do mapeamento dos centros de pesquisa que participariam da avaliação, foi solicitado aos pesquisadores seniores de cada centro, também de reconhecida experiência e histórico acadêmico, a criação de um grupo de três pessoas (ele/ela inclusive) que iriam atuar como atores no *Cenário de Uso*. Neste grupo, cada participante atuaria sob um determinado perfil de usuário no **e-Termos**, sendo que o pesquisador sênior interpretava o papel de *Gerente de Projeto*, e os outros dois os papéis de *Terminólogo/Linguista* e *Especialista de Domínio*, respectivamente.

Nesta configuração foram montados 12 grupos de avaliação, envolvendo um total de 36 pessoas. Dos 12 grupos, nove eram formados por terminólogos e tradutores (27 pessoas) e três por lexicógrafos (nove pessoas). Dessa forma, a amostra de avaliação foi constituída de maneira uniforme tendo 12 usuários atores que atuariam no papel de perfil de *Gerente de Projeto*, 12 *Terminólogos/Linguistas* e 12 *Especialistas do Domínio*.

À primeira impressão, o número total de 36 participantes da avaliação parece uma amostra muito pequena e não representativa, principalmente, se considerarmos toda a população que abarca a comunidade brasileira de terminologia e lexicografia. É fato que sob a visão de uma análise estatística esse número, realmente, é muito pequeno. Entretanto, é preciso considerar que a avaliação do **e-Termos** não é somente estatística e está fundamentada principalmente em um método que considera os aspectos colaborativos correspondentes à natureza do **e-Termos**. A estatística, portanto, neste processo, é o meio pelo qual a coleta de dados foi realizada, a partir da aplicação de um questionário, e representa apenas uma fase de todo o processo de avaliação.

Importa ressaltar também que a abordagem de avaliação que utiliza os *Cenários de Uso* como método não exige uma participação exaustiva de indivíduos, primando-se mais pela representatividade dos usuários avaliadores do que pela quantidade deles. Isso porque, geralmente, o número de usuários dos sistemas colaborativos é naturalmente pequeno, restrito a um nicho de pessoas bastante contido. Além disso, para o caso específico de avaliação do **e-Termos**, os potenciais usuários avaliadores estão em diversos estados e regiões do Brasil, produzindo uma ampla capilaridade, o que elevaria consideravelmente o custo da avaliação se essa maioria fosse incluída nas atividades do processo de avaliação. Dessa forma, como o sul do Brasil abriga centros de pesquisa em terminologia de reconhecido valor, com uma quantidade de pesquisadores considerável, optou-se por convidá-los e para envidar esforços singulares para sua participação.

6.1.3 O método do *Valor Percebido* como componente da avaliação

Embora o processo de avaliação do **e-Termos** tenha se baseado e consolidado sobre os planos de pesquisa descritiva, no método de cenários de uso e nos questionários para coleta de dados, durante o pleito de sua criação, também foram utilizadas idéias e concepções pertencentes ao método de avaliação colaborativa do *Valor Percebido* (*Perceived Value*), descrito na Seção 5.3.2. Isso porque, durante todo o tempo de criação das cenas, passagens e tarefas do *Cenário*, bem como na definição e elaboração das questões que fariam parte do questionário, houve a participação ativa de um especialista em terminologia que já conhecia boa parte do **e-Termos**.

Assim, é possível afirmar que a feição dos componentes e atributos desta avaliação são um reflexo direto desse esforço conjunto que, em outras palavras, representa as expectativas do desenvolvedor (o autor deste trabalho) e do usuário especialista, já familiarizado com o sistema do **e-Termos**. Em parte, tal fato corrobora os princípios metodológicos da avaliação do *Valor Percebido*, visto que este tenta medir o impacto organizacional de um determinado sistema, a partir da identificação dos componentes relevantes para a atividade de avaliação.

Dessa forma, é correto afirmar que parte da avaliação do **e-Termos** também se inspirou nos pressupostos teóricos e práticos do método de avaliação colaborativa do *Valor Percebido*.

6.2 Aplicação do Cenário de Uso e execução da avaliação

Uma vez montados os 12 grupos que iriam participar da avaliação, as próximas etapas foram: 1) identificar os atores e os respectivos perfis de cada um dos participantes atuantes no *Cenário* e 2) agendar o período de avaliação para cada centro selecionado. Dessa maneira, os centros participantes tiveram a realização da avaliação em períodos diferentes, com cuidado de não envolver pessoas de centros distintos em um só evento de avaliação. Assim, o período total de avaliação considerando todos os grupos de todos os centros foi de 25 dias, entre os meses de Março e Abril de 2009.

A avaliação completa de cada grupo de três pessoas tomava dois dias inteiros de trabalho, perfazendo um total de 16 horas. O primeiro dia foi totalmente dedicado a um curso intensivo do **e-Termos** com a participação obrigatória de todos os membros dos grupos participantes em cada centro. A intenção deste curso era apresentar uma visão geral do **e-Termos**, destacando seu histórico, seus objetivos, suas limitações, as principais características e funcionalidades, a metodologia de trabalho empregada, as atividades de cada etapa e as ferramentas de edição e comunicação disponíveis no sistema que possibilitavam o trabalho colaborativo.

Para elucidar todos estes aspectos, o conteúdo do curso foi enriquecido com diversos exercícios e exemplos reais de atividades que são praticadas durante a elaboração de um produto

terminológico. Além disso, o intuito destes exercícios e exemplos era explorar a maioria das funcionalidades do sistema, dispostas e relacionadas com cada Etapa do trabalho e da gestão terminológica. Sob esta visão, o curso oferecia a cada usuário participante a oportunidade de conhecer uma perspectiva completa do trabalho terminológico usando o **e-Termos**⁴.

O segundo dia de avaliação foi dedicado à aplicação do *Cenário de Uso* para cada um dos grupos. As aplicações do *Cenário* aconteceram no próprio laboratório do centro de pesquisa. O propósito desta decisão foi evitar que os participantes da avaliação ficassem fora do seu ambiente rotineiro de trabalho, tentando evitar uma possível influência negativa de um “novo” ambiente.

No início da aplicação do *Cenário*, cada grupo recebeu os três cadernos de atividades e os três CDs com o material necessário para execução das tarefas previstas. Estes, por sua vez, foram distribuídos aos participantes do grupo, respectivamente, para cada ator/perfil. Assim feito, estimulados a utilizar o **e-Termos** para execução do *Cenário* e a criação do minidicionário de *Revestimentos Cerâmicos*, cada grupo iniciou os trabalhos seguindo as tarefas e orientações contidas nos cadernos de atividades. É muito importante ressaltar que durante as atividades do *Cenário* não houve nenhuma ajuda externa, e que os grupos foram fortemente incentivados a resolver colaborativamente os problemas e eventuais dificuldades encontradas utilizando as ferramentas e funcionalidades do sistema.

Finalizado o processo de avaliação, no final do segundo dia de trabalho, após a completa conclusão das atividades do *Cenário*, cada participante foi convidado a responder o *Questionário de Pesquisa*, tendo recebido uma cópia do mesmo, que era individual, intransferível e não continha a identificação do respondente. Para todos os grupos, o tempo de resposta do mesmo foi no máximo de 30 minutos.

6.3 Descrição, análise e interpretação dos resultados

Com o término do período de aplicação do *Cenário de Uso*, teve início a fase de análise dos dados coletados, seguindo a metodologia esboçada na Figura 5.1. A primeira atividade desta fase foi a transcrição das respostas do questionário para o meio digital, de maneira que os cálculos estatísticos e correlações sobre os dados pudessem ser realizados com o uso de programas específicos.

6.3.1 Conjunto de dados da avaliação do **e-Termos**

O conjunto de dados de análise da avaliação é a reunião de todas as respostas das questões do questionário. Para facilitar, agilizar e flexibilizar os cálculos, os dados foram transportados e

⁴O material deste curso pode ser acessado gratuitamente no endereço: <http://www.etermos.ufscar.br/curso/slidesCurso2009.zip>

organizados para um banco de dados relacional, armazenados em uma tabela em que as colunas representavam as questões e as linhas (tuplas) as respostas de cada avaliador.

Todos os 12 grupos participaram da avaliação. Entretanto, das 36 pessoas envolvidas, apenas um, sob o perfil de *Terminólogo/Linguista*, não respondeu o *Questionário*, mas participou do curso e do *Cenário*. Dessa forma, houve a participação efetiva de 35 pessoas, sendo 12 sob o perfil de *Gerente de Projeto*, 11 *Terminólogos/Linguistas* e 12 *Especialistas do Domínio*. Nesta organização, a tabela de dados ficou com 53 colunas (número de questões objetivas) e 35 linhas (número de participantes da avaliação).

Várias análises foram realizadas, principalmente a partir do exame detalhado das distribuições de frequência das respostas de cada questão, bem como das possíveis relações associativas entre elas. As próximas seções mostram essas análises.

6.3.2 Interpretação e análise dos resultados da avaliação

A orientação da interpretação e análise dos resultados desta pesquisa procurou encontrar respostas e explicações para as todas as hipóteses definidas como “focos de avaliação” descritos na Tabela 6.2. Tanto o *Cenário de Uso* quanto o *Questionário de Pesquisa* tentaram cobrir e explorar tais focos, de maneira que os dados obtidos durante a fase de execução fossem, teoricamente, a representação fiel das possíveis descobertas. As respostas obtidas nesta fase foram utilizadas para alcançar tais explicações.

6.3.2.1 Análise pelas Distribuições de Frequência

Os dados das distribuições de frequência de todas as questões do *Questionário de Pesquisa* são apresentadas no Apêndice E deste trabalho.

Foco de Avaliação: Melhora na Gestão Terminológica

Esta subseção inicia com a constatação do fato de que embora o tempo previsto de aplicação do *Cenário de Uso* tenha sido de um dia inteiro (segundo dia de avaliação), apenas um grupo (de lexicografia) usou o dia todo. Os outros 11 grupos terminaram as atividades do *Cenário* em média 2 horas mais cedo. É essencial destacar que este fato é de extrema importância para o contexto e os resultados deste trabalho de doutorado e corrobora um dos principais objetivos do **e-Termos** que é a melhora na gestão do tempo de desenvolvimento de produtos terminológicos.

Genericamente, a gestão pode ser entendida como uma série de ações coordenadas que procuram garantir o uso eficaz de um conjunto de recursos para cumprir metas ou alcançar resultados planejados. Especificamente, as ações de gestão são planejar, organizar, liderar, coordenar e controlar uma equipe de trabalho e as tarefas por ela realizadas, estando sempre atento ao tempo e ao tratamento dos dados e informações relevantes.

De forma semelhante, no âmbito da “gestão terminológica” é possível afirmar que este conceito recai na conjunção das atividades que envolvem: 1) a gestão do tempo de desenvolvimento de um produto terminológico, 2) a gestão de dados e informações acerca do produto, 3) na prevenção de inconsistência e integração de dados, 4) no controle da equipe de trabalho e, conseqüentemente, na 5) uniformização das tarefas e atividades da equipe.

Sob este conceito de gestão terminológica as distribuições de freqüência das questões que avaliaram tais atividades são mostradas a seguir. Neste foco, as questões possuíam respostas em uma escala intervalar de 1 (“discordo totalmente”) a 5 (“concordo totalmente”).

Perguntados se o uso do sistema *melhora a gestão do tempo de desenvolvimento do produto terminológico* (questão 8b), 82,85% (29 pessoas) responderam 5, 14,28% (5 pessoas) responderam 4 e 2,85% (1 pessoa) responderam 3. Da mesma forma, quando questionados sobre a *melhora da gestão da informação* (questão 8a), 77,14% (27 pessoas) responderam 5 e 22,85% (8 pessoas) responderam 4. As Figuras 6.4(a) e 6.4(b) mostram, respectivamente, estes resultados.

(a) Melhora na gestão do tempo de desenvolvimento de um produto terminológico.

(b) Melhora na gestão da informação.

Figura 6.4: Distribuições de Freqüências das Questões 8a e 8b

Ainda com foco na gestão de informação, na questão que indaga se o sistema *evita a inconsistência de dados* (questão 8h), 54,28% (19 pessoas) responderam 5, 31,42% (11 pessoas) responderam 4 e 14,28% (5 pessoas) responderam 3. Já na pergunta se o sistema *favorece a integração de dados e informações do produto terminológico* (questão 8i), 82,85% (29 pessoas) responderam 5, 14,28% (5 pessoas) responderam 4 e 2,85% (1 pessoa) respondeu 3. As Figuras 6.5(a) e 6.5(b) mostram respectivamente estes resultados.

Sobre a melhora na *gestão da equipe de trabalho* (questão 8c), 74,28% (26 pessoas) responderam 5 e 25,71% (9 pessoas) responderam 4. Adicionalmente, quando questionados se o uso

(a) Prevenção de inconsistência de dados.

(b) Favorecimento da integração de dados e informações do produto terminológico.

Figura 6.5: Distribuições de Frequências das Questões 8h e 8i

do **e-Termos** *uniformiza as atividades e as tarefas da equipe* (questão 8f), 65,71% (23) responderam 5, 25,71% (9) responderam 4, 5,71% (2) responderam 3 e 2,85% (1) respondeu 2. As Figuras 6.6(a) e 6.6(b) mostram, respectivamente, estes resultados.

(a) Melhora da gestão da equipe de trabalho.

(b) Uniformização das atividades e tarefas da equipe.

Figura 6.6: Distribuições de Frequências das Questões 8c e 8f

Também vinculado às ações de gestão, quando perguntados se o **e-Termos** *favorecia a tomada de decisões* (questão 9e), 88,57% (31) responderam 5, 8,57% (3) responderam 4 e 2,85% (1) respondeu 3. Finalmente, os participantes foram consultados se o uso do sistema melhorava *o controle e a gestão do trabalho terminológico* (questão 8e) como um todo, tendo as seguintes distribuições de resposta: 85,71% (30) responderam 5 e 14,28% (5) responderam 4. As Figuras 6.7(a) e 6.7(b) mostram, respectivamente, estes resultados.

A partir do exame desse conjunto de questões e suas respectivas distribuições de respostas sobre a gestão terminológica implementada pelo **e-Termos**, observa-se que a grande maioria das

Figura 6.7: Distribuições de Frequências das Questões 9e e 8e

respostas concentram-se entre as opções 4 e 5, variando entre “concordo” e “concordo totalmente”. Fora dessa maioria, as respostas concentraram-se entre as opções 2 e 3, variando entre “discordo” e “nem concordo e nem discordo”.

Assim, conclui-se por esta distribuição que o **e-Termos** melhora consideravelmente a gestão do processo de criação de produtos terminológicos, principalmente se considerarmos a questão representada pela Figura 6.7(b), na qual 100% dos participantes “concordam” ou “concordam totalmente” com tal melhora.

Foco de Avaliação: Satisfação do Usuário

Várias questões do questionário podem ser agrupadas para mensurar a satisfação do usuário. Nesse conjunto, estão abarcadas principalmente as questões relacionadas com a usabilidade do sistema, justamente porque até certo ponto a usabilidade implica diretamente na satisfação. Assim, questões como: 1) facilidade de uso, 2) aparência da interface, 3) tempo de resposta, 4) confiabilidade do sistema, 5) conjunto de funcionalidades, 6) disposição de ferramentas e 7) os textos explicativos das ajudas on-line⁵ fornecem um bom indício de avaliação deste foco. As distribuições de frequência de cada uma destas questões possuíam respostas em uma escala intervalar de 1 (péssimo) a 5 (excelente) e são mostradas a seguir.

Sobre a *facilidade de uso do sistema* (questão 7a), 57,14% (20) responderam 5, 37,14% (13) responderam 4 e 5,71% (2) responderam 3. Perguntados sobre a *aparência da interface* (questão 7b) abrangendo itens como cores, apresentação geral e fontes, 80% (28) responderam 5, 17,14% responderam 4 e apenas 2,85 (1) respondeu 3. As Figuras 6.8(a) e 6.8(b) mostram, respectivamente, estes resultados.

Em relação à questão do *tempo de resposta do sistema* (questão 7c) a ações dos usuário,

⁵As ajudas on-line correspondem às dicas que o sistema oferece em cada ferramenta, funcionalidade ou tela específica, por meio do acionamento de hiperlinks como *Saiba Mais* e *Ver Exemplo*

(a) Facilidade de uso do sistema.

(b) Aparência da interface.

Figura 6.8: Distribuições de Frequências das Questões 7a e 7b

91,42% (32) responderam 5 e 8,57 (3) responderam 4. Ainda, quando questionados sobre a *confiabilidade do sistema* (questão 7e), 77,14% (27) responderam 5 e 22,85% (8) responderam 4. As Figuras 6.9(a) e 6.9(b) mostram, respectivamente, estes resultados.

(a) Tempo de respostas das ações no sistema.

(b) Confiabilidade do sistema.

Figura 6.9: Distribuições de Frequências das Questões 7c e 7e

Na questão que avaliava o *conjunto de funcionalidades* (questão 7i), 74,28% (26) responderam a opção 5, 22,85% (8) responderam 4 e 2,85% (1) respondeu 3. Quanto à *disposição das ferramentas na interface* (questão 7g), 77,14% (27) responderam 5, 14,28% (5) responderam 4 e 5,71 (2) responderam 3. Especificamente para esta questão, somando-se as pessoas que a responderam, observa-se que 1 pessoa não preencheu nenhuma das opções, representando 2,85% (1) dos casos. As Figuras 6.10(a) e 6.10(b) mostram, respectivamente, estes resultados.

Quanto ao questionamento das *ajudas on-line* (questão 7k) espalhadas na interface do sistema em cada formulário, ferramenta ou funcionalidade, 80% (28) responderam 5, 11,42% (4)

(a) Conjunto de funcionalidades do sistema.

(b) Disposição das ferramentas.

Figura 6.10: Distribuições de Frequências das Questões 7i e 7g

responderam 4 e 5,71% (2) responderam 3. Novamente, observa-se neste caso que 1 pessoa não respondeu a questão. Já sobre a qualidade dos *textos explicativos das ajudas* (questão 7l), 74,28% (26) responderam 5 e 20% (7) responderam 4. Verifica-se especificamente nesta questão que 2 pessoas não a responderam, representando um total de 5,71%. As Figuras 6.11(a) e 6.11(b) mostram, respectivamente, estes resultados.

(a) Disposição das ajudas on-line na interface.

(b) Qualidade dos textos explicativos das ajudas.

Figura 6.11: Distribuições de Frequências das Questões 7k e 7l

Adicionalmente, para finalizar a avaliação deste foco, os participantes também foram diretamente questionados sobre seu nível de satisfação a partir da experiência de utilização do **e-Termos** durante a execução do *Cenário*. Em uma escala intervalar em que 1 significa “totalmente insatisfeito” e 5 significa “totalmente satisfeito”, 74,28% (26) responderam 5 e 25,71% (9) responderam 4. O gráfico que representa as respostas desta questão é mostrado na Figura 6.12.

Com base nestes resultados apura-se que o nível de satisfação do usuário na utilização do **e-Termos** seguiu um padrão bastante otimista, principalmente se considerarmos as questões de

Figura 6.12: Distribuição de freqüências das respostas sobre o nível de satisfação do participante na experiência de uso do **e-Termos**.

usabilidade do sistema, cuja distribuição de respostas ficou em sua maioria nas opções 4 e 5, o que representa opiniões entre “bom” e “excelente”. O restante dos respondentes escolheram a opção 3, que significa “regular” nesta escala, ou não responderam, em um total de 4 pessoas em 3 diferentes questões.

Além disso, com base nas repostas da questão da Figura 6.12, podemos afirmar que tal satisfação é positiva, já que 100% dos respondentes diziam estar “satisfeitos” ou “totalmente satisfeitos”.

Foco de Avaliação: Colaboração e Interatividade do Sistema

Nesta pesquisa de doutorado, os aspectos colaborativos e de interação do usuário, tanto com o próprio sistema, quanto com os outros usuários é um dos principais focos. Dessa forma, visto que o **e-Termos** é em sua essência um ambiente colaborativo e que o *Cenário de Uso* aplicado favoreceu em vários momentos as tarefas de colaboração, o *Questionário de Pesquisa* também abordou este foco com diversos questionamentos, tais como se o sistema: 1) promove a colaboração da equipe de trabalho, 2) melhora a interação da equipe, 3) facilita a comunicação entre os membros da equipe, 4) favorece o compartilhamento de documentos, dados e informações, e se 5) promove a integração da equipe.

Para mostrar os resultados da pesquisa para o foco de colaboração e interatividade, as respostas destes questionamentos são apresentadas a seguir. Para estas questões, as opções estavam em uma escala intervalar em que 1 representava “discordo totalmente” e 5 “concordo totalmente”.

Na questão que mede se o usuário concorda que o **e-Termos** *promove a colaboração da equipe de trabalho* (questão 8d), 80% (28) responderam 5, 14,28% (5) responderam 4 e 5,61% (2) responderam 3. Sobre a *melhora de interação da equipe* (questão 9a), 80% (28) dos participantes responderam 5 e 20% (7) responderam 4. As Figuras 6.13(a) e 6.13(b) mostram, respectivamente, estes resultados.

(a) Promoção da colaboração da equipe de trabalho.

(b) Melhora da interação da equipe.

Figura 6.13: Distribuições de Frequências das Questões 8d e 9a

Questionados se o uso das ferramentas colaborativas do sistema *facilitava a comunicação entre os membros da equipe* (questão 9b), 91,42% (32) responderam 5 e 8,57% (3) responderam 4. Ainda, quando indagados se a utilização de tais ferramentas também *favoreciam o compartilhamento de documentos e informações* (questão 9c), 88,57% (31) responderam 5 e 11,42% (4) responderam 4. As Figuras 6.14(a) e 6.14(b) mostram, respectivamente, estes resultados.

(a) Comunicação entre os membros da equipe de trabalho.

(b) Compartilhamento de documentos, dados e informações.

Figura 6.14: Distribuições de Frequências das Questões 9b e 9c

Enfim, a respeito da *promoção e integração da equipe* (questão 9d), 77,14% (27) responderam 5, 17,14% (6) responderam 4 e 5,71% (2) responderam 3. A Figura 6.15 apresenta estes resultados graficamente.

Observando as respostas dos questionamentos apresentados neste foco de avaliação, percebe-se que os aspectos de colaboração e interação do sistema foram avaliados de maneira muito positiva. Isso porque das 5 questões apresentadas apenas 1, sobre a integração da equipe, da

Figura 6.15: Distribuição de freqüências das respostas sobre a integração da equipe de trabalho no sistema **e-Termos**.

Figura 6.15, apresentou respostas com opção 3 (2 pessoas). As outras 4 questões apresentaram respostas entre as opções 4 e 5, representados por “concordo” e “concordo totalmente”, sendo que deste conjunto a maioria das respostas foi 5, conforme pode ser visto pela distribuições de freqüência.

Esta percepção remete a duas constatações que merecem destaque. A primeira é mais conceitual e vem corroborar o comportamento e a natureza do trabalho e do fazer terminológico que é essencialmente colaborativo. A sustentação desta afirmação é confiável porque tais respostas vieram de um *Questionário de Pesquisa* oferecido após a execução de um *Cenário de Uso* completo e colaborativo, que foi aplicado em um ambiente real e de contexto relevante, tendo como base um sistema cujo principal fundamento é o aspecto da colaboração deste trabalho. A segunda é mais prática, e corresponde ao fato de que por meio de todas suas características e componentes o **e-Termos** representou a instância pragmática deste trabalho reproduzida em único ambiente integrado de colaboração e interação.

Foco de Avaliação: Intenção de Uso e Recomendações do Sistema para Outros Usuários

Sob uma visão mais geral e uma perspectiva mais ampla de utilização do **e-Termos**, os participantes da avaliação foram questionados sobre três possibilidades: 1) de usá-lo em seus próximos projetos ou trabalhos de terminologia, 2) de substituir as ferramentas computacionais utilizadas atualmente pelo seu uso e 3) de recomendá-lo para outros colegas de profissão. Em um escala intervalar em que 1 representava “muito improvável” e 5 “muito provável” os resultados destes questionamentos são apresentados a seguir.

Para a questão da *probabilidade de utilizar o e-Termos em um novo projeto* (questão 11), 82,85% (29) responderam 5, 14,28% (5) responderam 4 e 2,85% (1) responderam 3. Sobre a *probabilidade substituir as ferramentas atuais pelo uso do e-Termos* (questão 14), 54,28% (19)

responderam 5 e 40% (14) responderam 4. Também neste caso, 5,71% (2 pessoas) não responderam esta questão. As Figuras 6.16(a) e 6.16(b) mostram, respectivamente, estes resultados.

Figura 6.16: Distribuições de Frequências das Questões 11 e 14

Em relação à *probabilidade de recomendar o uso do e-Termos para um colega de profissão* (questão 12), as respostas para esta questão foram unânimes e 100% (35) dos participantes responderam 5, ou seja, que o recomendariam, conforme pode ser visualizado em sua representação gráfica na Figura 6.17.

Figura 6.17: Distribuição de frequências da probabilidade de recomendação de uso do **e-Termos** para colegas de profissão.

Estes resultados também são muito satisfatórios, pois baseados nas respostas destas questões, na qual 97,13% (somas das respostas 4 e 5 na questão 11) têm a intenção de usar o **e-Termos** nos próximos projetos e 94,28% (soma das respostas 4 e 5 da questão 14) o utilizariam no lugar das ferramentas atuais, bem como 100% o recomendariam para um colega de mesma profissão, é possível afirmar que tais resultados apontam para uma provável adoção do **e-Termos** nos ambientes de criação de produtos terminológicos, principalmente nos centros de pesquisa nos quais esta avaliação foi realizada, bem como naqueles com que estes possuem alguma interface.

Entretanto, vale ressaltar que tais centros de pesquisa e, conseqüentemente, os participantes desta avaliação, representam apenas uma amostra da população de terminólogos brasileira, o que afirma que tal adoção pode não chegar a uma cobertura completa.

Foco de Avaliação: Avaliação Geral do Usuário

Para esta fase de análise e interpretação dos resultados, as questões a seguir são de cunho geral, na qual os participantes foram convidados a avaliar o **e-Termos** sob um prisma mais genérico, considerando em um mesmo panorama todos os aspectos positivos e negativos do sistema.

Nesse contexto, duas questões em sequência foram apresentadas, uma que media o *interesse do indivíduo em participar do curso e da avaliação do e-Termos* (questão 5), e outra que media o seu *interesse após a sua participação* (questão 6). Em uma escala intervalar em que 1 representava “totalmente desinteressado” e 5 “totalmente interessado”, as respostas ficaram assim distribuídas para a primeira questão: 48,57% (17) responderam 5, 22,85% (8) responderam 4, 20% (7) responderam 3 e 8,27% (3) responderam 2. Para a segunda questão 94,28% (33) responderam 5 e 5,71% (2) responderam 4. As Figuras 6.18(a) e 6.18(b) mostram, respectivamente, estes resultados.

Figura 6.18: Distribuições de Frequências das Questões 5 e 6

A comparação dos resultados desses dois questionamentos, considerando principalmente o da segunda questão, é muito importante, pois indica que o impacto inicial do **e-Termos** é bastante positivo, já que 100% dos participantes responderam as opções 4 e 5, que significam respectivamente “interessado” e “totalmente interessado”. Isso porque conforme pôde ser visto, 28,27% (soma das respostas 2 e 3 da questão 5) mudaram seu interesse pelo **e-Termos**, já que não houve ocorrências destas opções de resposta na questão 6. Da mesma forma, também houve mudança de interesse de 17,14% das pessoas que responderam 4 na questão 5 e responderam 5 na questão 6. Assim, é plausível atentar que pelo menos 45,41% das pessoas aumentaram

seu interesse após a utilização do sistema. Entretanto, o impacto inicial pode ser volátil e superdimensionado, não representando o impacto total e final da adoção de um sistema, cuja mensuração depende do tempo, utilização e aceitação da comunidade.

Ainda no quesito da avaliação geral, os participantes também foram questionados sobre a possibilidade de quantificar o conteúdo aprendido, compreendendo as novas habilidades, descobertas e conhecimentos vindos a partir da experiência de utilização do **e-Termos**, durante o processo de avaliação. A escala intervalar dessas respostas variavam de 1 “não aprendi nada” a 5 “aprendi muito” e tiveram a seguinte distribuição: 65,71% (23) responderam 5, 28,57% (10) responderam 4 e 5,71% (2) responderam 3. A Figura 6.19 apresenta estes resultados graficamente.

Figura 6.19: Distribuição de freqüências da quantidade de aprendizado adquirido durante o uso do **e-Termos**.

Terminando a análise deste foco, os participantes também foram solicitados a fazer uma avaliação geral do **e-Termos**, fornecendo uma nota de 1 a 10, devendo considerar todos os pontos positivos e negativos do mesmo. Neste quesito, 45,61% (16) dos participantes deram nota 10, 48,56% (16) deram 9, 2,85% (1) deram 8 e 2,85% (1) deu nota 7, conforme mostrado na Figura 6.20.

Figura 6.20: Percentual de respostas da notas da avaliação geral do **e-Termos**.

Com base nestes resultados, é possível avistar favoráveis e efetivos êxitos do **e-Termos** na comunidade especializada que realiza o trabalho terminológico e de quem faz terminologia no

dia-a-dia. Isso porque tais resultados permitem enxergar um adequado indício de uma ampla recepção e aprovação do sistema por parte dos representantes especialistas desta comunidade. Para corroborar este indício, basta se referir primordialmente aos percentuais de respostas da questão sobre a avaliação geral do **e-Termos**, na qual 94,28% dos avaliadores deram notas entre 9 e 10 ao sistema.

Outros focos: Viés Didático e Terminologia e Informática

Representando um dos principais objetivos definidos para esta pesquisa, o viés didático do **e-Termos** também foi abordado durante a avaliação e no questionário. Este objetivo está fundamentado em uma perspectiva na qual o **e-Termos** atua como ferramenta de ensino de terminologia e terminografia, oferecendo a possibilidade concreta de reunir em um mesmo ambiente teoria e prática. Isso porque além de implementar os aspectos colaborativos, o **e-Termos** também representa a instanciiação de uma metodologia, real e consolidada que é empregada na criação de produtos terminológicos.

Neste sentido, em uma escala intervalar em que 1 significa “muito improvável” e 5 “muito provável”, quando questionados sobre a *probabilidade de usar o e-Termos em cursos de graduação e pós-graduação para o ensino de terminologia* (questão 13), 68,57% (24) responderam 5, 11,42% (4) responderam 4, 14,28% (5) responderam 3, 2,85 (1) respondeu 1. Vale ressaltar que também nesta questão 1 pessoa não respondeu a pergunta, que representa 2,85%. A Figura 6.21 apresenta graficamente estes resultados.

Figura 6.21: Percentual de respostas da probabilidade de usar o **e-Termos** em programas de graduação e pós-graduação.

Outro importante foco de avaliação que pôde ser considerado nesta análise é a mensuração do relacionamento atual entre a informática e terminologia na visão dos avaliadores. Este quesito é de considerável relevância, pois possibilita a identificação do atual nível da informatização nas práticas terminológicas e no universo particular dos usuários, considerando os ambientes reais de trabalho. Assim, quando questionados se *usavam recursos informatizados para criação dos seus produtos terminológicos* (questão 19a), os participantes responderam em uma escala intervalar em que 1 significa “não utilizo” e 5 “utilizo muito” as seguintes distribuições: 51,42% (18)

responderam 5, 20% (7) responderam 4, 17,14% (6) responderam 3 e 11,42% (4) responderam 1. Adicionalmente, também como parte desta tarefa de identificação, logo em seguida a esta pergunta, os usuários foram perguntados se os *recursos informatizados eventualmente utilizados no dia a dia cobriam todas as etapas de trabalho* (questão 19b). Em escala intervalar em que 1 significa “não cobre” e 5 “cobre todas”, 5,71% (2) responderam 5, 20% (7) responderam 4, 28,57% (10) responderam 3, 31,42% (11) responderam 2, 8,57% (3) responderam 1 e 2 pessoas não quiseram opinar, representando 5,71% do total. As Figuras 6.22(a) e 6.22(b) mostram, respectivamente, estes resultados.

Figura 6.22: Distribuições de Frequências das Questões 19a e 19b

A análise deste conjunto de resultados origina duas conclusões, uma positiva e uma negativa. A positiva é que 71,42% dos participantes, fruto da soma dos percentuais dos participantes que responderam 4 e 5, utilizam recursos informatizados para a realização de suas atividades terminológicas. Isso mostra que embora lenta, a **Terminótica** tem avançado e os terminólogos vêm buscando soluções automatizadas para suas práticas. Entretanto, a conclusão negativa é que a despeito deste avanço, as soluções informatizadas encontradas não cobrem todas as etapas do trabalho, evidenciando ainda o problema de diferentes ferramentas não integradas para execução destas atividades. Este resultado pôde ser obtido pela soma dos percentuais das respostas 1, 2, 3 e 4 da questão 19b, dando 88,56% das respostas.

Para corroborar esta conclusão, outras duas questões podem ser analisadas. A que mensura especificamente se os *recursos informatizados são integrados entre si* (questão 19c) e a que verifica se *os mesmos são elaborados exclusivamente para os fins terminológicos* (questão 19d). Para a primeira, em uma escala em que 1 significa “não integrados” e 5 “totalmente integrados”, as resposta foram assim distribuídas: 11,42% (4) responderam 4, 22,85% (8) responderam 3, 11,42% (4) responderam 2 e 45,71% (16) responderam 1. Como pôde ser calculado, a partir da soma das opções 1, 2 e 3, 78,98% dos participantes consideram que os sistemas atuais não são

integrados. Neste caso, 3 indivíduos não responderam a questão, formando 8,57% do total. As respostas desta questão podem ser vistas na Figura 6.23(a).

Para a segunda questão (19d), os respondentes foram estimulados a responder SIM ou NÃO se os recursos informatizados que utilizados eram desenvolvidos exclusivamente para a terminologia. Conforme pode ser visto na Figura 6.23(b) 60% responderam NÃO e 34,28% responderam SIM, ficando 5,71% (2 pessoas) sem responder.

(a) Integração dos recursos informatizados utilizados no trabalho terminológico.

(b) Intenção de desenvolvimento e exclusividade dos recursos informatizados.

Figura 6.23: Distribuições de Frequências das Questões 19c e 19d

Dessa maneira, a análise destes resultados serve para evidenciar um grande diferencial do **e-Termos**, que é justamente o fato conjugado de ser um sistema integrado, desenvolvido especificamente para a terminologia e que cobre todas as etapas do trabalho de criação de produtos terminológicos.

6.3.2.2 Análise pelas Correlações

Os gráficos e resultados numéricos produzidos pelas distribuições de frequência e apresentados na subseção anterior ajudam a compreender mais facilmente os dados, permitindo uma comunicação melhor dos resultados além de deixá-los visualmente mais atraentes. Entretanto, as distribuições de frequência examinam apenas dados de uma única variável por vez e oferecem respostas para os diversos valores de resposta que uma determinada questão possui.

Embora este tipo de análise de resultados seja de vital importância, o exame de muitas hipóteses ou questões de pesquisa também se concentra na relação entre duas ou mais variáveis. Dessa forma, quando existe uma ligação coerente e sistemática entre duas questões, é possível afirmar que existe uma relação estatística entre tais questões.

A **correlação** é uma técnica de análise associativa que determina a significância estatística de uma relação, permitindo assim estabelecer a força de associação. Uma força de associação muito forte indica que há uma probabilidade alta de existir uma relação entre as variáveis. Uma

força de associação moderada indica que é provável que exista a relação, sendo que a força de associação leve significa que a relação não está presente.

Em estatística, o **coeficiente de correlação** é usado para avaliar o grau de associação entre duas variáveis, indicando se existe uma fraca ou forte relação. Assim, grandes coeficientes indicam uma forte relação e baixos coeficientes uma fraca relação. Entretanto, o coeficiente de correlação só tem validade se estiver alinhado a um nível de significância, já que o número do coeficiente em si não possui muito sentido. O nível de significância estatística pode ser entendido como a probabilidade de rejeitar uma determinada hipótese nula quando ela é verdadeira. Na literatura estatística, o valor do nível de significância mais comum é menor 0,05 (menor que 5%). Todavia, a definição do valor da significância depende de cada projeto e do risco que o administrador pode ou deseja assumir.

Uma hipótese nula (H_0) no âmbito das correlações de variáveis estabelece a ausência de associação tendo o coeficiente de correlação igual a zero. O coeficiente de correlação varia de -1,00 a +1,00, com 0 (zero) representando, portanto, nenhuma correlação. Desse modo, quanto maior for o coeficiente de correlação, mais forte é a correlação ou nível de associação entre duas variáveis, podendo ser positivo (até +1,00), ou negativo (até -1,00).

O tipo de relação entre duas questões varia de acordo com seu tipo de resposta. Isso porque não é possível utilizar o mesmo cálculo da análise associativa para questões cujas opções de resposta são categorizadas ou ordinais, já que a especificação do relacionamento das respostas é diferente. Dessa maneira, considerando tais opções de respostas, podem ser empregadas 3 tipos de correlações: 1) categorizada x categorizada, 2) categorizada x ordinal e 3) ordinal x ordinal.

Neste contexto, já que o *Questionário de Pesquisa* aplicado possui questões com opções de respostas categorizadas e ordinais (que utilizam escala intervalar) a análise de relação deste trabalho seguiu duas vertentes: 1) a análise de relação entre questões com opções de respostas categorizadas e ordinais e 2) a análise de relação entre questões com opções de respostas ordinais. Tais vertentes serão exploradas a seguir.

Análise de Relação: Categorizada x Ordinal

Conforme visto na Seção 6.1.2.3 os usuários que participaram do *Cenário de Uso* para avaliação do **e-Termos** se classificam em 3 diferentes perfis: *Gerente de Projeto*, *Terminólogo/Linguista* e *Especialista do Domínio*. A seleção dos usuários para cada perfil seguiu critérios bem definidos, de maneira que os mais experientes deveriam ser necessariamente enquadrados no perfil de *Gerente de Projeto*, seguido pelo *Terminólogo/Linguista*, e os menos experientes, ou que estivessem iniciando a carreira em terminologia, sob o perfil de *Especialista do Domínio*. Esta pré-seleção é justificada porque na definição do *Cenário* a gestão terminológica é de responsabilidade do gerente, de modo que este perfil tinha a maior carga de trabalho e o maior número

de tarefas expostas no *Cenário* e, conseqüentemente, mais contato com as várias funcionalidades do sistema. Portanto, era natural que o indivíduo mais experiente atuasse como gerente.

Esta característica possibilitou a análise de correlações de questões em que as respostas são avaliadas por perfil de atuação no *Cenário*, viabilizando um exame flexibilizado de questões específicas sobre cada perfil de usuário. Dessa maneira, já que na primeira seção do *Questionário de Pesquisa* (que continham questões de triagem e abertura) havia uma questão que solicitava a identificação do perfil do participante durante a execução do *Cenário*, foi possível relacionar tal questão com diversas outras, permitindo a análise de diferentes focos de avaliação individualizada.

A Tabela 6.3 mostra as respostas de cada perfil de usuário em relação a *melhora a gestão do tempo de desenvolvimento do produto terminológico* (questão 8b). Na parte inferior da tabela é possível verificar os valores das medidas de **Média** e **Desvio Padrão (DP)** de cada perfil, que ajudam a interpretar os aspectos desta relação. A **média** corresponde a média aritmética das opções de resposta. O **Desvio Padrão (DP)** é usado para medir a variabilidade dos valores de distribuição da amostra a partir da **média**. Em outras palavras, indica o nível de concordância entre as respostas de uma determinada questão, de maneira que se o DP é grande (maior que 3,0), as respostas estão distantes da média, e muito pequeno (menor que 1,0), muito próximas à media, significando certa coerência nas respostas.

Adicionalmente, anexados à tabela, são mostrados os valores da significância estatística da relação, baseados no cálculo da ANOVA **Análise de Variância** que é utilizada para se avaliar as diferenças estatísticas entre as médias de dois ou mais grupos. Neste caso, a hipótese nula (H_0) da relação é que as médias dos grupos são iguais, ou seja, que não existe diferença estatística significativa entre os grupos avaliados. Para as análises realizadas neste trabalho o nível de significância adotado foi $\alpha = 0,05$ (5%).

Observe que as respostas de todos os perfis variaram entre as opções 4 e 5 que significam “concordo” e “concordo totalmente”, respectivamente, estando o valor da média igual 4,833 para o perfil Gerente, 4,909 para Terminólogo/Linguista e 4,667 para o Especialista do Domínio. Adicionalmente, verificando os valores de DP, 0,389 para Gerente, 0,302 para Terminólogo/Linguista e 0,651 para Especialista do Domínio percebe-se que os dois primeiros concordam mais que a gestão do tempo é melhorada com o uso do **e-Termos**, ficando o terceiro com maior nível de variabilidade. Para corroborar esta conclusão observe que o **Valor P** é igual 0,463 do teste ANOVA, a um nível de significância de $\alpha = 0,05$ (5%), tendo o valor de **F** menor que o **F-crítico**, podendo assim aceitar a hipótese de que não existe diferença significativa entre as médias dos grupos.

Os dados sobre a relação entre perfil de usuário e a melhora da *gestão da informação do produto terminológico* (questão 8a) estão apresentados na Tabela 6.4. Observe neste caso que

Tabela 6.3: Valores da relação entre perfil do usuário e gestão do tempo

RELAÇÃO						
Perfil do Usuário x Gestão do tempo de desenvolvimento do produto terminológico						
Perfil/Medidas	Perfis					
	Gerente de Projeto	Terminólogo/Linguista	Esp. do Domínio			
Média	4,833	4,909			4,667	
Desvio Padrão	0,389	0,302			0,651	
ANOVA						
Fonte da Variação	S. Quad.	GL	M. Quad.	F	Valor P	F-crítico
Entre Grupos	0,358	2	0,179	0,790	0,463	3,295
Dentro dos Grupos	7,242	32	0,226	—		
Total	7,600	34	—			

a média das respostas do Gerente foi de 4,917 e o DP de 0,289 que significa que o nível de concordância entre os indivíduos deste perfil foi alto. Entre os perfis de Terminólogo/Linguista e Especialista de Domínio as médias foram 4,818 e 4,583, respectivamente, com valores de DP de 0,405 e 0,515, ambos sem muita variabilidade de respostas com DPs abaixo de 1,0. Examinando os valores da ANOVA, cujo **Valor P** é 0,144, mostra que não existe diferença significativa entre as médias, mostrando que todos os perfis concordam com a melhora da gestão da informação.

Tabela 6.4: Valores da relação entre perfil do usuário e gestão da informação

RELAÇÃO						
Perfil do Usuário x Gestão da informação do produto terminológico						
Perfil/Medidas	Perfis					
	Gerente de Projeto	Terminólogo/Linguista		Esp. do Domínio		
Média	4,917	4,818		4,583		
Desvio Padrão	0,289	0,405		0,515		
ANOVA						
Fonte da Variação	S. Quad.	GL	M. Quad.	F	Valor P	F-crítico
Entre Grupos	0,701	2	0,350	2,052	0,144	3,294
Dentro dos Grupos	5,469	32	0,170	—		
Total	6,17	34	—			

A relação entre a *melhora da gestão da equipe* (questão 8c) e as avaliações de cada perfil de usuário estão mostrados na Tabela 6.5. Também nesta relação a média das respostas do perfil de Gerente ficou em 4,917, com DP de 0,289 significando que a grande maioria dos Gerentes acharam que o **e-Termos** melhora a gestão da equipe. Entretanto, é importante notar os valores para os outros perfis, respectivamente, 4,545 para média e 0,522 para o DP e 4,750 para média e 0,452 para DP, já que estes perfis não fazem a gerência de equipes e sim, são gerenciados. Como podem ser examinados, os cálculos do teste da ANOVA produziram um **Valor P** igual a 0,133, sem diferença significativa das médias, mostrando que tais valores também indicam uma avaliação positiva, e expressa uma interessante observação de que a gestão foi bem avaliada de ambos os lados, de quem gerencia e de quem é gerenciado, quesito geralmente contrário nesta relação humana.

Tabela 6.5: Valores da relação entre perfil do usuário e gestão da equipe de trabalho

RELAÇÃO						
Perfil do Usuário x Gestão da equipe de trabalho						
Perfil/Medidas	Perfis					
	Gerente de Projeto	Terminólogo/Linguista	Esp. do Domínio			
Média	4,917	4,545			4,750	
Desvio Padrão	0,289	0,522			0,452	
ANOVA						
Fonte da Variação	S. Quad.	GL	M. Quad.	F	Valor P	F-crítico
Entre Grupos	0,792	2	0,396	2,149	0,133	3,295
Dentro dos Grupos	5,894	32	0,184	—		
Total	6,686	34	—			

Em uma avaliação mais genérica, a respeito da *melhora do controle e a gestão do trabalho terminológico* (questão 8e) como um todo, observe na Tabela 6.6 que a média das respostas do perfil de Gerente foi igual a 5, o que significa dizer que todos são unânimes na questão que mede se o **e-Termos** melhorou a gestão do trabalho. Verificando o valor do DP igual 0,0 (zero) significa que nenhuma variabilidade ocorreu em suas respostas. Já as respostas dos perfis de Terminólogo/Linguista e Especialista do Domínio variaram entre as opções 4 e 5, estando o valor da média igual 4,727 para o primeiro e 4,833 para o segundo. Observe também é possível aceitar a hipótese nula de que as médias dos três perfis são muito próximas, dado o **Valor P** igual a 0,178 e de **F** bem menor do eu **F-crítico**.

Tabela 6.6: Valores da relação entre perfil do usuário e gestão do trabalho terminológico

RELAÇÃO						
Perfil do Usuário x Gestão da equipe do trabalho terminológico						
Perfil/Medidas	Perfis					
	Gerente de Projeto	Terminólogo/Linguista	Esp. do Domínio			
Média	5,000	4,727	4,833			
Desvio Padrão	0,000	0,467	0,389			
ANOVA						
Fonte da Variação	S. Quad.	GL	M. Quad.	F	Valor P	F-crítico
Entre Grupos	0,437	2	0,218	1,817	0,178	3,294
Dentro dos Grupos	3,848	32	0,120	—		
Total	4,285	34	—			

Realizando uma análise conjunta da relação entre os perfis dos usuários e as 4 variáveis de gestão exploradas acima, observa-se que as médias do perfil de Gerente foram superiores em 3 ocasiões, sendo menor somente na gestão do tempo em comparação com o perfil Terminólogo/Linguista. Da mesma forma, se comparar os valores dos DPs de cada perfil, os do Gerente estarão mais próximos de 1,0 na maioria dos casos. Esta comparação é relevante porque considerando que os usuários com perfil de Gerente possuem comprovadamente mais experiência em terminologia, estes resultados corroboram muito claramente com as conclusões de que o **e-Termos** contribui efetivamente para a gestão terminológica. É claro que os bons resultados apurados dos outros perfis são bastante relevantes para também fortalecer tal conclusão, princi-

palmente baseando-se no fato de que os testes da ANOVA executados para 4 relações aceitaram a hipótese de que não há diferença significativa das médias, confirmando a tendência de que os e perfis avaliados têm a mesma opinião.

No âmbito das correlações entre questões categorizadas e ordinais, uma importante comparação é o *nível de interesse* (questão 6) de cada perfil após ter participado do processo de avaliação do **e-Termos**. A Tabela 6.7 apresenta estes resultados. Conforme pode ser visto, o interesse dos usuários com perfil de Gerente foi unânime, tendo a média de respostas igual a 5, que significa “totalmente interessado”. Em relação aos outros perfis, os resultados também foram bastante positivos, tendo média igual 4,909 para o perfil Terminólogo/Linguista e 4,917 para o Especialista de Domínio. Confirmando esta conclusão, observe que o **Valor P** igual 0,596 do teste da ANOVA foi bem maior do que o nível de significância estipulado (0,05) mostrando que os valores da médias foram bastante próximas.

Tabela 6.7: Valores da relação entre perfil do usuário e nível interesse

RELAÇÃO						
Perfil do Usuário x Nível de interesse						
Perfil/Medidas	Perfis					
	Gerente de Projeto	Terminólogo/Linguista	Esp. do Domínio			
Média	5,000	4,909	4,917			
Desvio Padrão	0,000	0,302	0,289			
ANOVA						
Fonte da Variação	S. Quad.	GL	M. Quad.	F	Valor P	F-crítico
Entre Grupos	0,060	2	0,030	0,525	0,596	3,295
Dentro dos Grupos	1,826	32	0,057	—		
Total	1,886	34	—			

Adicionalmente, um exame mais elaborado seria a comparação desta análise com os resultados da relação entre os perfis dos usuários e a *probabilidade de utilização do e-Termos no próximo projeto terminológico* (questão 11). De acordo com a Tabela 6.8, que apresenta os dados desta relação, a média do Gerente para esta questão foi de 4,917. Porém, os outros perfis tiveram médias mais baixas, sendo 4,727 para o perfil Terminólogo/Linguista e 4,750 para o Especialista do Domínio. Entretanto, mesmo sendo inferior, o teste da ANOVA mostra que não há diferença significante entre as médias já que o **Valor P** foi igual a 0,584.

O exame das duas relações supracitadas é de grande importância para as contribuições desta pesquisa. Isso porque, analisando somente os resultados do perfil de Gerente, que em muitas vezes é o responsável pelo empreendimento de novos projetos terminológicos, é plausível inferir uma segura receptividade do **e-Termos**, já que as variáveis de interesse e probabilidade de uso foram consideravelmente bem avaliadas. Tal receptividade é fundamental para a análise da percepção de impacto deste sistema, a ser realizada adiante, neste mesmo capítulo.

Outra importante comparação de respostas perfil a perfil é o *nível de satisfação do usuário*

Tabela 6.8: Valores da relação entre perfil do usuário e probabilidade de uso

RELAÇÃO						
Perfil do Usuário x Probabilidade de uso						
Perfil/Medidas	Perfis					
	Gerente de Projeto	Terminólogo/Linguista	Esp. do Domínio			
Média	4,917	4,727			4,750	
Desvio Padrão	0,289	0,647			0,452	
ANOVA						
Fonte da Variação	S. Quad.	GL	M. Quad.	F	Valor P	F-crítico
Entre Grupos	0,252	2	0,126	0,548	0,584	3,295
Dentro dos Grupos	7,348	32	0,230	—		
Total	7,600	34	—			

(questão 10), mensurado após a utilização do **e-Termos** durante o dia de aplicação do *Cenário*. O resultado desta relação está exposto na Tabela 6.9. Novamente, em quase sua totalidade o nível de satisfação do usuário Gerente foi de 5, que significa “totalmente satisfeito”. Em geral a satisfação dos perfis de Terminólogo/Linguista e Especialista do Domínio também estiveram entre as opções 4 e 5, com médias de 4,636 e 4,667, respectivamente. Analisando os valores do teste da ANOVA, tanto o **Valor P**, igual a 0,249, quanto o valor **F**, igual 1,454, observa-se que o nível de satisfação de todos os perfis é uniforme, não tendo portando diferença significativa das médias.

Tabela 6.9: Valores da relação entre perfil do usuário e o nível de satisfação

RELAÇÃO						
Perfil do Usuário x Nível de satisfação						
Perfil/Medidas	Perfis					
	Gerente de Projeto	Terminólogo/Linguista	Esp. do Domínio			
Média	4,917	4,636			4,667	
Desvio Padrão	0,289	0,505			0,492	
ANOVA						
Fonte da Variação	S. Quad.	GL	M. Quad.	F	Valor P	F-crítico
Entre Grupos	0,557	2	0,278	1,454	0,249	3,295
Dentro dos Grupos	6,129	32	0,192	—		
Total	6,686	34	—			

Conforme dito anteriormente o sistema **e-Termos** também representa a instanciación de uma metodologia prática de criação de produtos terminológicos. Esta metodologia está dividida em um conjunto de tarefas distribuídas ao longo de seis Etapas de trabalho. Entretanto, tal metodologia não é ditatória e não necessariamente todos os projetos devem seguir e executar todas as tarefas e passar por todas as Etapas, de maneira que existe uma flexibilidade controlada do sistema, justamente para atender várias metodologias de trabalho. Nesse sentido, os participantes foram questionados sobre o *nível de concordância desta metodologia aplicada* (questão 15) durante a aplicação do *Cenário*. Estes resultados, separados por perfil estão apresentados na Tabela 6.10. Observe que a média desta distribuição para o perfil do Gerente foi 4,833 com DP de 0,389, indicando que a metodologia de trabalho foi bem aceita pelos usuários mais ex-

perientes. Por outro lado, vale destacar o valor mais baixo da média dos usuários com perfil de Terminólogo/Linguista, que ficou em 4,364 com DP de 0,674, indicando um bom nível de concordância entre eles. Embora levemente mais alto, os valores de média e DP para o perfil Especialista de Domínio, 4,667 e 0,492, respectivamente, também foram mais conservadores. Contudo, mesmo tendo o **Valor P** mais baixo, igual a 0,114, ainda é possível verificar que não há diferença estatisticamente relevante entre as médias dos perfis, significando que os participantes concordam com a metodologia aplicada e implementada pelo **e-Termos**.

Tabela 6.10: Valores da relação entre perfil do usuário e o nível de concordância da metodologia de trabalho

de trabalho

RELAÇÃO						
Perfil do Usuário x Nível concordância da metodologia de trabalho						
Perfil/Medidas	Perfis					
	Gerente de Projeto	Terminólogo/Linguista	Esp. do Domínio			
Média	4,833	4,364	4,667			
Desvio Padrão	0,389	0,674	0,492			
ANOVA						
Fonte da Variação	S. Quad.	GL	M. Quad.	F	Valor P	F-crítico
Entre Grupos	1,293	2	0,646	2,329	0,114	3,295
Dentro dos Grupos	8,879	32	0,277	—		
Total	10,171	34	—			

Para corroborar a idéia de que os avaliadores com perfil de Gerente eram mais experientes, e consequentemente, os outros perfis menos, a Tabela 6.11 apresenta a relação de cada perfil com o *tempo de carreira* (questão 21). Sob uma escala intervalar que variava de 1 a 5, em que 1 significa “menos de 5 anos”, 2 “de 5 a 10 anos”, 3 “de 10 a 15 anos”, 4 “de 15 a 20 anos” e 5 “mais de 20 anos”, observe que a média das respostas do perfil de Gerente foi de 2,417, o que significa de dizer que os Gerente tinham em média 10 anos de experiência. Neste caso específico, vale destacar o valor do DP, que foi igual a 1,311, significa dizer que embora o tempo médio fosse de 10 anos, o perfil dos Gerentes foi muito heterogêneo, com respostas variando de 4 (entre 15 e 20 anos) a 1 (menos de 5 anos), com algumas respostas 2 e 3, que corresponde o tempo entre 5 e 15 anos de carreira. Por outro lado, a média das respostas dos outros perfis foram de 1,182 e 1,167, respectivamente, estabelecendo uma média de menos de 5 anos de experiência. Contudo, diferentemente do perfil Gerente, os DPs destes perfis foram baixos, sendo 0,405 e 0,389, respectivamente, revelando que o tempo de carreira foi bem mais uniforme para as pessoas deste perfil, tendo assim, pouca variabilidade.

Note que neste caso o cálculo do teste da ANOVA mostra que as médias do tempo de carreira são estatisticamente diferentes, rejeitando a hipótese nula, já que o **Valor P** foi igual 0,001, portanto, menor que o nível de significância de 0,05. Além disso, observe que o valor **F** foi muito maior que o **F-crítico**, denotando que realmente a média do tempo de carreira entre os perfis eram bastante diferentes.

Tabela 6.11: Valores da relação entre perfil do usuário e o tempo de experiência

RELAÇÃO						
Perfil do Usuário x Tempo de experiência						
Perfil/Medidas	Perfis					
	Gerente de Projeto	Terminólogo/Linguista	Esp. do Domínio			
Média	2,417	1,182	1,167			
Desvio Padrão	1,311	0,405	0,389			
ANOVA						
Fonte da Variação	S. Quad.	GL	M. Quad.	F	Valor P	F-crítico
Entre Grupos	12,180	2	6,090	8,771	0,001	3,295
Dentro dos Grupos	22,220	32	0,694	—		
Total	34,400	34	—			

A verificação da relação das notas da avaliação geral do **e-Termos** de cada perfil também é de grande relevância. Isso porque a partir desta comparação é possível conhecer qual perfil o sistema mais agradou e teve a avaliação mais positiva. A Tabela 6.12 mostra a média de cada perfil, considerando que as notas figuravam em uma escala intervalar que variava de 1 a 10. Observe que as médias das notas foram decrescentes, partindo de 9,750 para o perfil de Gerente, 9,273 para o perfil de Terminólogo/Linguista e de 9,083 para o perfil de Especialista do Domínio. Verificando os valores dos DPs, percebe-se que para o perfil de Gerente as notas tiveram pouca variabilidade, igual 0,452 estando abaixo de 1,0 e mais próximo de 0 (zero). Ao contrário o valor do DP para o perfil de Terminólogo/Linguista ficou em 0,905, bem mais próximo de 1,0, significando que embora a média tenha sido acima de 9, a variabilidade das notas foi maior. Para comprovar este fato, verifique o **Valor P** do teste da ANOVA igual a 0,046, que mostra que realmente as médias dos grupos de perfis possuem diferença significativa, também mostrado pelo valor **F**, pouco maior que **F-crítico**. Entretanto, embora exista essas diferenças, de maneira geral, as notas dadas pelos usuários foram todas acima de 9, denotando uma avaliação bastante positiva.

Tabela 6.12: Valores da relação entre perfil do usuário e as notas da avaliação geral

RELAÇÃO						
Perfil do Usuário x Avaliação geral						
Perfil/Medidas	Perfis					
	Gerente de Projeto	Terminólogo/Linguista		Esp. do Domínio		
Média	9,750	9,273		9,083		
Desvio Padrão	0,452	0,905		0,515		
ANOVA						
Fonte da Variação	S. Quad.	GL	M. Quad.	F	Valor P	F-crítico
Entre Grupos	2,823	2	1,411	3,384	0,046	3,295
Dentro dos Grupos	13,348	32	0,417	—		
Total	16,171	34	—			

Análise de Relação: Ordinal x Ordinal

Os principais os métodos de calcular o coeficiente da relação entre variáveis ordinais são a **correlação de Pearson** e a **correlação de Spearman**. A **correlação de Pearson** mensura

a associação linear entre duas variáveis que estabelecem uma “associação em linha reta” entre as duas variáveis consideradas. As principais particularidades desta correlação são a natureza dos dados, que devem ser mensurados a partir de opções métricas e a suposição que as variáveis envolvidas são de uma população que obedece a distribuição normal. Já a **correlação de Spearman** é usada para calcular o coeficiente de correlação de variáveis com escalas ordinais, ou que possuem medidas intervalares, sem necessariamente obedecer a distribuição normal. A **correlação de Spearman** é naturalmente mais conservadora do que a **correlação de Pearson**, de forma que os coeficientes possuem valores mais altos. Neste caso, a hipótese nula (H_0) da relação estabelece que não existe nenhuma associação entre as variáveis.

Nesse contexto, visto que a maioria das questões do *Questionário de Pesquisa* possui repostas em uma escala intervalar (ordinais) e não obedecem a uma distribuição normal, a **correlação de Spearman** será utilizada na análise de várias relações a seguir com o nível de significância de $\alpha = 0,05$ (5%).

Considerando a usabilidade do sistema, mensurada pela segunda seção do questionário, uma importante correlação a ser considerada é entre a *aparência da interface* (questão 7b) e o *posicionamento dos seus componentes, tais como, tabelas, botões, figuras e links* (questão 7f) e a *disposição dos diversos menus* (questão 7g). As Tabelas 6.13 e 6.14 mostram os dados das correlações de Spearman para estas questões. Observe na Tabela 6.13 que o **Valor P** (coeficiente de correlação) foi 0,0084, portanto, menor do que o nível de significância ($\alpha = 0,05$) indicando a rejeição da hipótese nula e reafirmando que esta correlação é estatisticamente significativa. Isso quer dizer que a aparência da interface está muito correlacionada com o posicionamento dos itens e componentes da tela. Esta conclusão pode ser confirmada se observadas em paralelo os valores das médias das respostas de cada questão. Veja que as médias foram, respectivamente, 4,771 e 4,686, denotando que as avaliações de ambos questionamentos foram positivos.

Tabela 6.13: Correlação de Spearman entre a Aparência da Interface e Posicionamentos dos componentes na tela

CORRELAÇÃO DE SPEARMAN		
	Aparência da Interface	Posicionamento dos componentes
N. Respostas	35	35
Média	4,771	4,686
Desvio Padrão	0,490	0,471
Valor P	0,0085	

Na correlação entre a aparência da interface e disposição dos menus, mostrado na Tabela 6.14 o coeficiente **Valor P** foi de 0,0077, ainda menor do que o anterior, indicando que as disposição dos menus está mais fortemente relacionado com a aparência da interface. Também para confirmar este relacionamento mais forte, veja que o valor da média para questão de disposição

dos menus foi de 4,800, maior do que a média da questão de posicionamento do componentes que foi de 4,686.

Tabela 6.14: Correlação de Spearman entre a Aparência da Interface e a Disposição dos menus

CORRELAÇÃO DE SPEARMAN		
	Aparência da Interface	Disposição dos menus
N. Respostas	35	35
Média	4,771	4,800
Desvio Padrão	0,490	0,406
Valor P	0,0077	

Entretanto, embora existam tais diferenças, ambas as análises anteriores mostram que a aparência da interface, que foi bem avaliada, está diretamente relacionada à disposição do menus e também ao posicionamento dos componentes da tela.

Uma importante correlação a explorar no contexto do **e-Termos** é se, afinal, o *melhoramento do controle e da gestão do trabalho terminológico* torna o *produto final mais confiável*. Este relacionamento corresponde à correlação das questões 8e e 8g, respectivamente, e estão dispostas na Tabela 6.15. Verifique que também neste caso a hipótese nula de que não há correlação pode ser recusada, já que o **Valor P** foi de 0,0218, menor que o nível de significância adotado. A conclusão desta correlação é muito valiosa para o contexto desta pesquisa, pois demonstra que uma melhor gestão do trabalho beneficia a qualidade final do produto terminológico, deixando-o mais fiável.

Tabela 6.15: Correlação de Spearman entre o controle e a gestão do trabalho terminológico e a confiabilidade final do produto

CORRELAÇÃO DE SPEARMAN		
	Melhora no controle e a gestão do trabalho terminológico	Produto final mais fiável
N. Respostas	35	35
Média	4,857	4,343
Desvio Padrão	0,355	0,725
Valor P	0,0218	

Outra relação interessante a considerar é aquela que especifica se a interação da equipe de trabalho favorece a tomada de decisões durante a execução do projeto. Este questionamento é bastante relevante e encontra uma das principais hipóteses e conceitos especificados na natureza colaborativa, que é justamente a consideração do contexto do trabalho em equipe para decisões futuras. Conforme já dito anteriormente o **e-Termos** como ambiente colaborativo explora esse conceito, tornando, portanto, esta relação objeto de análise. Este relacionamento é determinado pelas questões 9a e 9e e estão apresentados na Tabela 6.16. Verifique que o coeficiente **Valor P** para esta correlação foi de 0,0040, e representa um valor muito menor do que o nível de significância estipulado. Consequentemente, isso mostra que tais variáveis são muito relacionadas,

e corrobora com a idéia estampada no trabalho colaborativo de que a interatividade da equipe favorece a tomada de decisões.

Tabela 6.16: Correlação de Spearman entre a interatividade da equipe e a tomada de decisões

CORRELAÇÃO DE SPEARMAN		
	Interatividade da equipe	Tomada de decisões
N. Respostas	35	35
Média	4,800	4,857
Desvio Padrão	0,406	0,430
Valor P	0,0040	

Ainda nesse contexto a análise da relação entre a comunicação dos membros da equipe com a promoção de sua integração e a melhora no controle e gestão do trabalho terminológico também pode ser empregada. Observe na Tabela 6.17 que neste caso que o Valor P da correlação entre a *comunicação dos membros da equipe* (questão 9b) a *promoção de sua integração* (questão 9d) foi de 0,0003 chegando muito próximo de 0 (zero), indicando uma forte correlação entre estas questões. Desse modo, é correto afirmar que em um ambiente colaborativo a comunicação é primordial para a integração da equipe de trabalho e, conseqüentemente, para sua interação.

Tabela 6.17: Correlação de Spearman entre a comunicação dos membros da equipe e a promoção da integração

CORRELAÇÃO DE SPEARMAN		
	Comunicação da equipe	Integração da equipe
N. Respostas	35	35
Média	4,914	4,714
Desvio Padrão	0,284	0,572
Valor P	0,0003	

A Tabela 6.18 apresenta os dados da relação entre a *comunicação da equipe* (questão 9b) e a *melhora na gestão do trabalho terminológico* (questão 8e). Verifique que o **Valor P** para este relacionamento foi de 0,0056, novamente muito baixo e bem menor do que o nível de significante definido para esta análise. Este resultado também é de extrema relevância para o contexto desta pesquisa, já que uma de suas contribuições é justificar que uma boa comunicação entre os membros de uma determinada equipe, evidenciado na tarefa colaborativa, aperfeiçoa o controle e a gestão do trabalho, fato comprovado também na atividade terminológica.

Sobre o *nível de interesse do usuário após a experiência de uso do e-Termos* (questão 6), o principal relacionamento é com a *uniformização das atividades e tarefas da equipe de trabalho* (questão 8f). A Tabela 6.19 mostra os resultados desta relação. Veja que o **Valor P** foi de 0,0040, novamente muito baixo, indicando uma considerável associação entre as questões. Esta análise também é relevante porque mostra que a uniformização das tarefas e atividades da equipe envolvida no trabalho terminológico desperta um grande interesse do usuário e evidencia umas

Tabela 6.18: Correlação de Spearman entre a comunicação dos membros da equipe e o controle e gestão do trabalho terminológico

CORRELAÇÃO DE SPEARMAN		
	Comunicação da equipe	Melhora no controle e a gestão do trabalho terminológico
N. Respostas	35	35
Média	4,914	4,857
Desvio Padrão	0,284	0,355
Valor P	0,0056	

das principais carências deste trabalho que é a carência de produtos informatizados, integrados e uniformes.

Tabela 6.19: Correlação de Spearman entre o nível de interesse do usuário e a uniformização das atividades e tarefas da equipe de trabalho

CORRELAÇÃO DE SPEARMAN		
	Nível de interesse do usuário	Uniformização das atividades e tarefas da equipe de trabalho
N. Respostas	35	35
Média	4,943	4,543
Desvio Padrão	0,236	0,741
Valor P	0,0040	

Para explicar a satisfação do usuário baseado na experiência de uso do **e-Termos** três relacionamentos podem ser analisados. O primeiro deles relaciona o *nível de satisfação* (questão 10) com a *aparência da interface* (questão 7b). Observe na Tabela 6.20 que o **Valor P** para esta associação foi de 0,0023 indicando que a aparência da interface está intimamente relacionada com a satisfação dos usuários que participaram da avaliação. Isso mostra que os cuidados com a interface de recursos informatizados desenvolvidos especificamente para os fins terminológicos são fundamentais para cativar os usuários e garantir sua satisfação, já que outras ferramentas adaptadas não possuem interface específica para a terminologia.

Tabela 6.20: Correlação de Spearman entre o nível de satisfação do usuário e a aparência da interface

CORRELAÇÃO DE SPEARMAN		
	Nível de satisfação do usuário	Aparência da interface
N. Respostas	35	35
Média	4,743	4,771
Desvio Padrão	0,443	0,490
Valor P	0,0023	

O segundo relacionado com o nível de satisfação considera a *interatividade com o usuário* (questão 7h). Neste caso, conforme pode ser verificado na Tabela 6.21 o **Valor P** foi de 0,0106 e, embora mais próximo do nível de significância, denota uma influência efetiva desta questão na associação com nível de satisfação do usuário. Novamente a questão da interatividade evidencia os aspectos colaborativos da prática terminológica e a colocam em um panorama afirmativo de

consideração das contribuições desta pesquisa.

Tabela 6.21: Correlação de Spearman entre o nível de satisfação do usuário e a interatividade com o usuário

CORRELAÇÃO DE SPEARMAN		
	Nível de satisfação do usuário	Interatividade com o usuário
N. Respostas	35	35
Média	4,743	4,829
Desvio Padrão	0,443	0,382
Valor P	0,0106	

A terceira correlação com o nível de satisfação do usuário envolve a *utilização de recursos informatizados pelos usuários* (questão 19a). Neste caso, o coeficiente de correlação **Valor P** foi de 0,0063, baixo e distante do nível de significância, como pode ser confirmado na Tabela 6.22.

Tabela 6.22: Correlação de Spearman entre o nível de satisfação do usuário e a utilização de recursos informatizados

CORRELAÇÃO DE SPEARMAN		
	Nível de satisfação do usuário	Utilização de recursos informatizados
N. Respostas	35	35
Média	4,743	4,000
Desvio Padrão	0,443	1,328
Valor P	0,0063	

A análise desta correlação permite concluir que o **e-Termos** alcançou um nível de satisfação bastante otimista em relação aos usuários que utilizam recursos informatizados para a criação de produtos terminológicos. Isso porque, o fato de tais usuários já utilizarem alguns recursos informatizados (mesmo que adaptados) os tornam mais comparativos, oferecendo uma base de julgamento mais fundamentada para o exame. Se fosse ao contrário, essa relação não seria tão característica, mesmo que positiva, pois na ausência de quaisquer recursos informatizados, qualquer novo recurso poderia desviar o resultado da satisfação. Dessa forma, os recursos informatizados e atualmente utilizados pelos usuários têm um peso importante na avaliação da satisfação do **e-Termos**, visto que sua existência aumenta a capilaridade e a consciência crítica do usuário.

6.4 Considerações Finais

Este capítulo apresentou todo o processo de avaliação do **e-Termos**. As características de seleção do método e suas justificativas de escolha, bem com os componentes da avaliação foram reveladas em detalhes. Particularmente, atenção especial foi dada na descrição da elaboração do *Cenário de Uso* utilizado, na criação do *Questionário de Pesquisa* para coleta de dados e na formação dos grupos de usuários avaliadores alocados em duas regiões do Brasil. Além disso,

também foram relatados os aspectos relacionados com a aplicação e execução da avaliação nos ambientes locais dos grupos participantes.

Entretanto, o principal conteúdo deste capítulo é a descrição, análise e interpretação dos resultados obtidos após o processo avaliação. Várias análises estatísticas são realizadas, tanto pelas simples distribuições de frequência como pelo exame de várias correlações estatísticas, especificamente para avaliar os diversos focos de mensuração definidos para esta pesquisa, essenciais para as conclusões e contribuições delineadas no Capítulo 7.

Conclusão e trabalhos futuros

7.1 Conclusão

A principal questão de pesquisa deste doutorado é dupla:

Seria possível construir um ambiente colaborativo que contemple todas as etapas do trabalho terminológico de orientação descritiva e avaliar o impacto que tal sistema causa no trabalho terminológico?

Essa questão é relevante porque há dois componentes inéditos nessa iniciativa: 1) de enxergar e assumir que o trabalho terminológico é naturalmente colaborativo e 2) a partir dessa perspectiva, implementar um sistema computacional em ambiente Web que integra os conceitos do CSCW, conjugados com uma teoria terminológica nova e fundamentada.

Decorrentes desta iniciativa surgem várias questões adicionais que estão direta ou indiretamente ligadas ao “impacto” que se deseja avaliar, sendo elas:

1. O que faz do **e-Termos** um sistema diferente que convença a comunidade terminológica a usá-lo em detrimento de outros sistemas, mesmo que não integrados?
2. O **e-Termos** realmente melhora, agiliza, facilita e uniformiza o trabalho terminológico?
3. Qual a mudança ou reflexo que o **e-Termos** causa na rotina do trabalho terminológico?
4. De que modo a colaboração ou o aspecto colaborativo ajuda a prática terminológica?

Contudo, a aparição destas questões é natural e faz parte dos “impactos”, justamente porque o impacto sozinho é uma variável complexa, composta por diversos fatores que influenciam sua

mensuração. Tais fatores é que tornam a avaliação de impactos um trabalho difícil e complicado, quase sempre envolvido em contextos sociais e comportamentais do público alvo e dependente das crenças, experiências e expectativas de cada usuário.

Além disso, supondo que fosse possível mensurar o impacto diretamente, ignorando tais questões e acessórios e tratando-o como um conceito direto e isolado, não seria pertinente porque tais resultados ainda estão sujeitos ao tempo, pois não se medem os impactos imediatamente, principalmente pela necessidade de avaliar tais fatores em um contexto conjunto, possibilidade vinda somente a partir do uso constante, em um determinado intervalo e de avaliações específicas e periódicas.

Entretanto, uma maneira alternativa e satisfatória de avaliar o impacto que esta pesquisa revelou é medir a percepção do impacto, este sim, mais direto e imediato, produzido pelo exame dos diversos focos de avaliação tratados ao longo desta monografia. Desse modo, tendo em vista a dificuldade de aferir os impactos positivos e negativos causados pela adoção de um sistema colaborativo no contexto do trabalho terminológico, já que tais resultados dependem do tempo e do contexto social e prático o qual está inserido, é sensato apurar que os resultados mostrados neste trabalho sugerem uma percepção de impacto bastante positiva. Isso porque, analisando os resultados obtidos e mostrados até aqui, principalmente no que se refere à melhoria da gestão terminológica, à satisfação, interesse e receptividade dos usuários, aos aspectos de colaboração e interação das equipes e da sua provável recomendação de uso, temos o indicativo favorável dos impactos reais, vindos somente com o tempo.

Todavia, é prudente acreditar que o tempo pode mudar tais percepções, principalmente com a evolução das ferramentas de comunicação voltadas para a Internet, tão pulsante nos dias de hoje.

Adicionalmente, o empreendimento desta pesquisa de doutorado e, conseqüentemente, a implementação do **e-Termos**, exigiu do autor o estudo e o conhecimento multidisciplinar das áreas de Terminologia e Ciência da Computação. Da Terminologia foi demandado o estudo das teorias terminológicas, principalmente na diferenciação e comparação dos pilares teóricos que as sustentam e, também, das práticas terminológicas, sobretudo no mapeamento das orientações metodológicas do exercício e do fazer terminológico. Da Ciência da Computação, foram necessários os conceitos e métodos das áreas: Processamento de Língua Natural, Banco de Dados, Desenvolvimento de Sistemas Internet/Web, Interação Humano Computador, Avaliação de Sistemas (tanto *single-user* quanto *groupware*), Técnicas de Programação, Linguagens de Programação, e Sistemas Colaborativos.

Dessa maneira, a realização deste trabalho foi fundamentada e dividida em duas partes, que apesar de diferentes entre si, caminharam juntas: a parte Terminológica e a Computacional. A

parte terminológica corresponde à aplicação prática de uma metodologia de criação de produtos terminológicos conforme descrito no Capítulo 3. Esta metodologia, que tem como base teórica a Teoria Comunicativa da Terminologia (TCT) e orientação prática o trabalho de Almeida (2000), determina o cumprimento de uma sequência de etapas, compostas por um conjunto de tarefas que devem ser executadas por diferentes perfis de usuários, de maneira que os produtos terminológicos vão se construindo a cada etapa finalizada, aqui representadas por cada Módulo do **e-Termos**, conforme mostrado na Tabela 7.1.

Tabela 7.1: Etapas de trabalho terminológico e seus respectivos Módulos no **e-Termos**

Etapas	Tarefas	Módulo
01	Compilação, preparação e análise da qualidade do corpus de especialidade	0 e 1
02	Extração automática de candidatos a termos	2
03	Edição da ontologia	3
04	Categorização dos candidatos a termos na ontologia e validação dos mesmos pelos especialistas do domínio	3
05	Elaboração e organização da Base Definicional	4
	Criação e preenchimento da Ficha Terminológica	4
	Elaboração da Definição dos termos	4
06	Elaboração dos verbetes e do produto terminológico final	5
	Publicação e disponibilização do produto terminológico final para o intercâmbio e consulta	5

Resumidamente, na parte computacional, as principais tarefas executadas para o desenvolvimento desta pesquisa e implementação do **e-Termos** foram:

1. Implementação, aplicação ou adaptação das ferramentas de PLN para fins da pesquisa terminológica, no contexto específico de cada módulo do **e-Termos**;
2. Criação e avaliação ou adoção de padrões de entrada e saída específicos para cada Módulo, de maneira que permitissem o fluxo de dados transparente entre os mesmos, bem como a comunicação com as ferramentas de terceiros;
3. Adoção dos métodos estatísticos de extração automática de candidatos a termos estudados por Teline (2004);
4. Implementação e integração do Editor de Ontologias que possibilita a criação, edição e visualização de ontologias e termos alocados nas mesmas;
5. Estudo dos formatos e padrões de intercâmbio dos produtos terminológicos, bem como a implementação de classes e rotinas para exportação;
6. Implementação de ferramentas para criação flexibilizada das Fichas Terminológicas, preenchimento da Base Definicional, elaboração e edição dos Modelos de Verbetes;

7. Criação do modelo conceitual e da estrutura do Banco de Dados utilizado para atender os requisitos do sistema e os pressupostos da pesquisa terminológica, bem como para suportar as funcionalidades das ferramentas lingüísticas e colaborativas propostas por cada Módulo do **e-Termos**;
8. Estudo e avaliação dos diferentes sistemas e/ou plataformas de desenvolvimento de aplicações colaborativas;
9. Identificação dos pontos, em cada etapa do trabalho terminológico, e consequentemente em cada Módulo do **e-Termos**, de colaboração dos usuários, de maneira que a natureza do trabalho colaborativo fosse atendida;
10. Implementação dos Recados, do Chat e do suporte ao envio de e-mails que dão suporte às necessidades comunicacionais e interativas do **e-Termos**;
11. Adaptação e ajustes da ferramenta do Fórum do **e-Termos**;
12. Pesquisa, estudo, análise e seleção das abordagens e métodos de avaliação de sistemas colaborativos, bem como dos métodos de coletas de dados aplicados em pesquisas e amostragens;
13. Elaboração dos *Cenários de Uso* e do *Questionário de Pesquisa* para execução da avaliação do **e-Termos**;
14. Execução da avaliação com usuários reais e a organização, análise e interpretação dos resultados.

O processo de criação de produtos terminológicos é um trabalho complexo e longo, pois desde a compilação do corpus até edição dos verbetes finais podem ser gastos meses de esforço e dedicação. Se este trabalho for manual, esse tempo pode aumentar ainda mais. Com o uso do **e-Termos** espera-se que este tempo diminua e o trabalho seja mais estruturado e confiável.

Assim, dada a complexidade de criação destes produtos e a longa duração deste trabalho, a avaliação de sua sistematização foi por si só um grande desafio desta pesquisa, principalmente porque todo este contexto estava envolvido em uma abordagem colaborativa, natural do trabalho terminológico, mas que representava mais um ingrediente dependente do tempo. Nesse sentido, foi inevitável o desenho de uma solução de avaliação que abrangia o impacto da adoção de um sistema Web que suportava a atuação simultânea de diferentes indivíduos com perfis diferenciados sobre o extenso e complexo trabalho terminológico instanciado em uma esfera colaborativa.

Dessa forma, embora a mensuração real dos impactos não tenha sido possível, a aplicação dos *Cenários de Uso*, influenciada pela abordagem do *Valor Percebido* em conjunto com os *Questionários de Pesquisa* foi a solução mais adequada para o processo de avaliação deste trabalho.

Foi possível simular uma situação real de uso do sistema, que criava um produto terminológico completo, passando por todas as etapas de trabalho, executado no ambiente comum dos usuários, envolvendo diferentes pessoas que, por sua vez, estavam apoiadas para agir colaborativamente.

Junto ao processo de avaliação, é importante salientar que o desenvolvimento dos módulos do **e-Termos** e das ferramentas de suporte ao trabalho terminológico que foram acopladas em sua arquitetura não foi exaustiva. Ou seja, foi implementado o Ambiente completo, entretanto, sem a necessidade rigorosa de implementar ou disponibilizar todas ferramentas, todos extratores automáticos, todos padrões de intercâmbio bem como satisfazer todos os modelos e formatos de dados. Entretanto, houve sim um compromisso de criar todo Ambiente e adotar ou adaptar as principais ferramentas linguísticas e colaborativas, os padrões internacionais mais utilizados e promover a interação com outras ferramentas, possibilitando uma adequada gestão terminológica.

Nesse contexto, embora esta visão pareça minimalista, de Abril a Agosto de 2009 (data de finalização da escrita desta tese), o **e-Termos** conta com mais de 130 usuários, alocados em 68 diferentes projetos terminológicos que, em tese, representa um futuro produto. Estes dados revelam uma promissora aceitação do sistema na comunidade terminológica brasileira, principalmente se considerarmos que não houve nenhuma ação de divulgação ou disseminação do mesmo até este momento, e que se trata do resultado de um trabalho tipicamente acadêmico.

7.2 Contribuições

Abaixo, avaliamos a importância e contribuições do **e-Termos**, tanto para as pesquisas terminológicas em língua portuguesa quanto para o domínio da ciência da computação.

Contribuições para Terminologia

Para a Terminologia, sua contribuição não se dá somente pelo fato de sistematizar o trabalho terminológico, tornando-o mais ágil e fiável, mas fundamentalmente pelo fato do **e-Termos** ser o resultado prático de reflexões cuidadosas sobre a proposição de um método em Terminologia, que procura explicitar os postulados de uma Terminologia de orientação descritiva, fundamentada em princípios da Linguística, e que tem como novidade a ação colaborativa, inédita neste âmbito de aplicações da computação. Tal contribuição é relevante porque naturalmente a sistematização de uma metodologia implica na especificação dos seus componentes e atividades em um nível mais detalhado e, conseqüentemente, diferente daquela elaborada conceitualmente.

Dessa maneira, é necessário dizer que este trabalho é de grande valia para o cenário da pesquisa terminológica no Brasil, pois, instancia todas as etapas do processo de gestão terminológica, tornando possível o empreendimento de projetos de terminologia de maneira sistemática, organizada e mais fiel aos pressupostos científicos.

Decorrente disso, outra contribuição a ressaltar é a junção e uniformização das tarefas e atividades constantes nas etapas da criação de produtos e da gestão terminológica em um único ambiente computacional, evitando a utilização de diferentes ferramentas e recursos informatizados não integrados e adaptados para este fim que, por sua vez, representa uma dificuldade e um gargalo reconhecido na terminografia.

Contribuições para Ciência da Computação

Também nesse contexto, e já listando as contribuições da informática, importa sublinhar que o **e-Termos** atua no propósito fundamental da ciência da computação, que é a automatização das tarefas e processos manuais da atividade humana, auxiliando os homens em suas atividades, com o propósito de agilizar e torná-las mais rápidas, confiáveis e estruturadas. Diante disso, é essencial destacar que indiretamente este projeto tem a virtude de tornar público o conhecimento específico de uma determinada área, pois produz bases de conhecimentos de um domínio especializado, utilizadas fundamentalmente como instrumentos para uma comunicação eficaz e para a difusão do conhecimento.

Para a ciência da computação, a primeira contribuição a destacar é o acoplamento em uma única plataforma computacional de diferentes ferramentas de PLN voltadas especificamente para atender as carências e exigências da pesquisa terminológica, com o propósito de incrementar a qualidade dos produtos gerados. Além disso, vale salientar a implementação e sistematização das regras e requisitos que atuam no controle dos trabalhos em Terminologia, possibilitando uma melhor gestão do tempo, dos dados e da equipe.

Adicionalmente, também como substancial contribuição, este fato permite que o **e-Termos** seja vinculado à prática pedagógica aplicada aos estudos da terminologia em conjunto com os exercícios e aplicações de PLN, podendo ser usado em cursos ou projetos de ensino e aprendizagem da terminologia em nível de graduação e pós-graduação.

Outra relevante contribuição é o emprego da abordagem colaborativa para todas as etapas que compõem o trabalho e a gestão terminológica. Até então, alguns sistemas de gerenciamento terminológico utilizam essa abordagem somente em etapas específicas do processo de criação de produtos terminológicos, prioritariamente naquelas para as quais são desenvolvidos, sendo esta então uma notória contribuição desta pesquisa de doutoramento. Entretanto, embora inicialmente tenha sido projetado para uso exclusivo da terminologia, é essencial assinalar que o **e-Termos** também pode ser usado para os trabalhos e iniciativas da lexicologia, para a elaboração de dicionários da língua geral, e da tradução, como ferramenta de organização de termos e palavras multilíngües empregadas no ofício de traduzir. Novamente, tal possibilidade é bastante valiosa, pois expande as fronteiras de uso do **e-Termos** e alcança outros perfis de profissionais.

No contexto de uso plural do **e-Termos**, tanto para terminologia quanto para lexicologia

e tradução, surge mais uma contribuição desta pesquisa, que é justamente o fato de torná-la como uma ação de inovação tecnológica, pois atua diretamente no aumento da eficiência de um processo produtivo por meio da sistematização de um conjunto de tarefas específicas que aprimoram a qualidade do produto final. Contudo, é possível perceber que a inovação tecnológica está ligada à melhoria e implementação de um processo novo e substancialmente aperfeiçoado, mas não necessariamente inédito, para a comunidade ou área de atuação.

No campo das aplicações *Groupware* ou CSCW esta pesquisa também forneceu uma fundamental contribuição, pois desenvolveu e aplicou um método original e formal de avaliação colaborativa, a partir da conjunção de duas abordagens de avaliação (*Cenários de Uso* e *Valor Percebido*) e do emprego de um *Questionário de Pesquisa* para coleta de dados, utilizado para avaliar as percepções de impactos produzidos pela implantação do **e-Termos** no âmbito da gestão terminológica.

Além disso, é importante salientar que o **e-Termos** ultrapassa os requisitos iniciais de ser “somente” um ambiente de gestão terminológica para também funcionar como um *workbench* (mesa de trabalho) para aplicação e estudo dos conceitos e práticas da colaboração entre todos os usuários e pesquisadores envolvidos em um projeto de criação e elaboração de um produto terminológico. Assim, esta é a principal contribuição desta pesquisa, pois atua na identificação e aplicação dos aspectos do ofício colaborativo no âmbito da terminologia e lexicologia, bem como na identificação da percepção dos efeitos, impactos e influências que a colaboração tem sobre estas atividades.

Nesse sentido, esperamos ter desenvolvido uma solução inovadora que considera a abordagem colaborativa para a gestão de terminologias que segue os pressupostos da Teoria Comunicativa da Terminologia, juntamente com o emprego de ferramentas linguísticas específicas, podendo assim contribuir para as pesquisas terminológicas, principalmente no Brasil.

7.3 Limitações

Paralelas aos resultados positivos desta pesquisa várias dificuldades ou limitações foram encontradas durante seu percurso, tanto aquelas relacionadas à parte operacional ou prática quanto aquelas relacionadas às atividades de pesquisa em si.

No contexto dos aspectos práticos, a primeira limitação desta pesquisa foi a necessidade de recorte e seleção das funcionalidades e/ou funções a implementar no **e-Termos**. Devido às questões de pesquisa envolvidas neste trabalho, foi necessário limitar o desenvolvimento ou adaptação de várias ferramentas por três motivos diferentes: 1) por estas consumirem muito tempo de implementação, 2) por não possuírem subsídios de pesquisa que justificassem o esforço para seu empreendimento ou 3) por serem naturalmente questões longas, que poderiam ser

desenvolvidas em outras pesquisas, como por exemplo, a compilação automática de *cópus* (tarefa do Módulo 1), a extração de candidatos a termos (tarefa do Módulo 3), a criação e proposição automática de ontologias (tarefas de Módulo 4) e o Editor de Definição (tarefa do Módulo 5). Todas essas ferramentas envolvem questões de pesquisa que merecem atenção especial e fugiram do escopo deste trabalho.

Também vinculado à parte operacional está a dificuldade de empreender iniciativas de divulgação do sistema, bem como na adoção das práticas ou metodologias de projetos específicos de interfaces que adota as soluções da Web Design. Embora tal dificuldade pareça ser distante das questões de pesquisa deste trabalho, pois a primeira vista uma interface elegante e funcional não responde as hipóteses definidas. Esta característica é fundamental para a obtenção dos resultados, pois a funcionalidade, e os aspectos de usabilidade podem interferir na aceitação do sistema pelo usuário, e conseqüentemente, influenciar os fatores da colaboração.

Em relação às atividades de pesquisa, o custo da avaliação do **e-Termos** foi uma grande limitação. Isso porque estatisticamente falando, o objetivo era reunir o número máximo de potenciais usuários para ter uma amostra maior da população de avaliadores. Entretanto, envolver todos os grupos de pesquisa em terminologia no Brasil que poderiam participar do processo de avaliação foi inviável, pois o método de avaliação de *Cenário de Uso* possui naturalmente um custo elevado, e tais grupos estão espalhados pelo país, tendo representantes em quase todas as regiões. Cada grupo envolvido gerava despesas de viagem, alimentação, transporte, hotel e material do curso e do *Cenário*. Além disso, devido ao número de pessoas e tempo gasto no processo de avaliação, também foi difícil encontrar agenda para todos os grupos, configurando-se como outra limitação.

Contudo, a principal limitação desta pesquisa foi a dependência do tempo exigido para aferição concreta e real dos possíveis impactos e efeitos da adoção do **e-Termos** na (semi-)automatização da gestão terminológica e principalmente, do comportamento colaborativo sobre trabalho terminológico. Tal limitação exigiu do autor uma readequação das atividades do cronograma e uma alteração da análise e interpretação dos resultados obtidos durante o processo de avaliação.

7.4 Trabalhos futuros

Os diversos trabalhos futuros oriundos desta pesquisa de doutoramento podem ser agrupados em duas categorias: 1) de ordem técnica ou operacional e 2) de pesquisa e investigação. Os de ordem técnica correspondem às atividades mais práticas relacionadas com a implementação de novas funcionalidades do **e-Termos**, principalmente aquelas relacionadas com a evolução e melhoria do sistema, bem como aquelas provenientes das solicitações ou novos requisitos dos

usuários ou correções de erros. Atualmente, existe uma série de tarefas relacionadas a esta categoria, surgidas, sobretudo após a fase de avaliação e aplicação do *Cenário de Uso*.

Os de ordem operacional estão relacionados primordialmente àqueles que estabelecem a formação de parcerias com outros projetos ou na fixação da relação inter-sistemas tendo o **e-Termos** de um lado. Exemplos destas relações são, por exemplo, os sistemas de busca e *download* de *cópus*, o intercâmbio (importação e exportação) de ontologias, bem como a comunicação via *Serviços Web (Web Services)* com outras bases terminológicas. Também nesse campo está prevista a melhoria da infra-estrutura computacional do **e-Termos**, particularmente no aumento do poder de processamento dos servidores.

Além disso, ações de divulgação e demonstração do **e-Termos** devem ser empreendidas por meio da publicação de artigos científicos, cursos de curta e longa duração, ou apresentação em feiras e exposições científicas relacionados às áreas de terminologia, lexicografia e tradução ou onde as comunidades terminológicas ou lexicográficas se fizerem presentes. Nesse sentido, vale dizer que tal divulgação já vem sendo realizada pelo envio de mensagens eletrônicas em listas de discussão especializadas.

Todavia, os trabalhos futuros de maior relevância são aqueles relacionados à pesquisa, nos quais podem ser empreendidos novos trabalhos de mestrado, doutorado ou pós-doutorado, objetivando sempre o avanço das fronteiras do conhecimento e a progressão da ciência terminológica. Nesse contexto, vários temas de pesquisa podem ser elencados tais como:

1. O estudo e a melhoria da compilação automática de *cópus* no sentido de aprimorar a eficiência dos algoritmos de busca de textos para principalmente diminuir o tempo avaliação da qualidade dos resultados e aumentar sua relevância, reduzindo os resultados *falsos-positivos*;
2. A criação de novos métodos de extração automática de termos, ou a combinação dos métodos existentes para obtenção de melhores resultados. Tais iniciativas podem, por exemplo, envolver a anotação semântica dos textos, ou associar esta com as anotações sintáticas, bem como estudar maneiras semi-automáticas para descobrir o padrão de formação sintática dos termos de uma determinada área de conhecimento;
3. A evolução dos algoritmos e soluções para a proposição automática de mapas conceituais a partir de um *cópus*, principalmente na harmonização das técnicas estatísticas e o conhecimento lingüístico;
4. O estudo e investigação das técnicas e métodos envolvidos com a *definição terminológica*, principalmente na identificação dos padrões definitórios dos termos e nos componentes da definição, no sentido de sistematizar esta tarefa para a criação e implementação de *Editores*

de Definição. Vale ressaltar que o resultado desta proposta permite o desenvolvimento de ferramentas que identificam automaticamente os trechos de textos definitórios presentes em um corpus e possibilita a população automática da *Base Definicional*;

5. A análise da interação e colaboração que acontece especificamente durante a tarefa de *definição terminológica* ou de refinamento das informações dos termos, principalmente para melhorar seu suporte computacional;
6. A avaliação da interface de acordo com as técnicas e métodos da área de Interação Humano Computador (IHC) para, se necessário, a proposição de um novo leiaute que atenda às requisições dos usuários e facilite ainda mais sua interação com o sistema.

Entretanto, o trabalho futuro de mais valia para esta pesquisa será realmente a avaliação dos impactos causados pela implantação do **e-Termos** e, conseqüentemente, o exame criterioso do efeito e dos resultados do trabalho colaborativo na rotina do trabalho terminológico surgido somente a partir de sua adoção. Isso porque, conforme já dito anteriormente, tais resultados são produzidos com o tempo e constitui-se essencialmente em descobertas vindouras.

Referências Bibliográficas

- Alder, A. D. (1998). An experiment in blind terminology interchange: developing and testing conversion algorithms for externally supplied data. Dissertação de Mestrado, Brigham Young University, Department of Linguistics. Citado na página 91.
- Almeida, G. M. B. (2000). *Teoria Comunicativa da Terminologia (TCT): Uma Aplicação*. Tese de Doutorado, FCLAR-UNESP, Araraquara-SP. Vols. 1 e 2. Citado nas páginas 36, 68, 83, 86, e 223.
- Almeida, G. M. B. (2003a). O percurso da terminologia: de atividade prática à consolidação de uma disciplina autônoma. *TradTerm Revista do Centro Interdepartamental de Tradução e Terminologia FFLCH-USP*, 9(1). Citado nas páginas 79 e 94.
- Almeida, M. B. (2003b). Roteiro para construção de uma ontologia bibliográfica através de ferramenta automatizada. *Perspectivas da Ciência da Informação*, 08(02):164-179. Citado nas páginas 71, 73, e 75.
- Aluísio, S. M. (1995). *Ferramentas de Auxílio à Escrita de Artigos Científicos em Inglês como Língua Estrangeira*. Tese de Doutorado, IFSC-USP, São Carlos-SP. 228 p. Citado nas páginas 101, 141, e 142.
- Aluísio, S. M. e Muniz, F. A. M. (2007). Instalação e administração do portal de corpus do projeto pln-br. Relatório Técnico NILC-TR-07-12, Universidade de São Paulo (USP). Relatórios NILC. Citado na página 110.
- Aluísio, S. M., Muniz, F. A. M., e Infante, K. (2007). Projeto pln-br: O cabeçalho em xml para os textos do corpus e o editor web de cabeçalhos. Relatório Técnico NILC-TR-07-05, Universidade de São Paulo (USP). Relatórios NILC. Citado na página 110.
- Aluísio, S. M., Pinheiro, G. M., Finger, M., Nunes, M. G. V., e Tagnin, S. E. (2003). The lácio-web project: Overview and issues in brazilian portuguese corpora creation. *Corpus Linguistics 2003*, 16(1). Special Issue. Citado nas páginas 42 e 101.
- Aluísio, S. M., Pinheiro, G. M., Manfrim, A., Oliveira, L. H. M., Junior, L. C. G., e Tagnin, S. E. (2004). The lácio-web: Corpora and tools to advance brazilian portuguese language investigations and computational linguistic tools. Em *Proceedings of IV International Conference on Language Resources and Evaluation (LREC)*. Citado na página 42.
- Alves, I. M. (1999). A delimitação da unidade lexical nas línguas de especialidade. Em *Revista Palavra*, páginas 69-80. Citado nas páginas 4, 22, 57, 65, e 66.
- Ananiadou, S. (1994). A methodology for automatic term recognition. Em *Proceedings of 15th International Conference on Computational Linguistics*, páginas 1034-1038. Citado nas páginas 64, 65, e 118.
- Antiqueira, L., Feltrim, V. D., e Nunes, M. G. V. (2003). Projeto e implementação do sistema scipo. Relatório Técnico 26, Instituto de Ciências Matemáticas e de Computação - ICMC-USP. Citado nas páginas 103 e 141.
- Antunes, P. e Costa, C. J. (2004). Perceived value: A low-cost approach to evaluate meetingware. Em *Lecture notes in computer science*, páginas 109-125. Springer Berlin / Heidelberg. Citado na página 177.
- Arouck, O. (2001). Avaliação de sistemas de informação: revisão da literatura. *Revista Transinformação*, 13(1):7-21. Citado na página 160.
- Auer, S. (2005). powl - a web based platform for collaborative semantic web development. Citado na página 75.

- Avancini, H., Lavelli, A., Magnini, B., e Sebastiani, F. (2002). Expanding domain-specific lexicons by term categorization. Citado na página 79.
- Baeza-Yates, R. e Pino, J. A. (1997). A first step to formally evaluate collaborative work. Em *GROUP '97: Proceedings of the international ACM SIGGROUP conference on Supporting group work*, páginas 56–60, New York, NY, USA. ACM. Citado nas páginas 173 e 174.
- Baeza-Yates, R. e Pino, J. A. (2006). Towards formal evaluation of collaborative work. Em *Information Research*, volume 11(4). Citado na página 176.
- Bagot, R. E. (1999). *Extracció de terminologia: elements per a la construcció d'un SEACUSE*. Tese de Doutorado, Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, Barcelona. Citado nas páginas 60, 61, 63, 64, e 118.
- Bagot, R. E. (2001). Extracción de terminología: elements para la construcción de um extractor. *TradTerm Revista do Centro Interdepartamental de Tradução e Terminologia FFLCH-USP*, 7(1). Citado nas páginas 59, 62, 64, e 118.
- Baker, K., Greenberg, S., e Gutwin, C. (2001). Heuristic evaluation of groupware based on the mechanics of collaboration. Em *Proceedings of the 8th IFIP Working Conference on Engineering for Human-Computer Interaction (EHCI01)*. Citado na página 176.
- Bannon, L. J. (1992). Discovering CSCW. Em *15th Information Systems Research in Scandinavia (IRIS)*, páginas 507–520. Citado nas páginas 7 e 170.
- Baroni, M. e Bernardini, S. (2004). Bootcat: Bootstrapping corpora and terms from the web. Em *Proceedings of LREC 2004, Fourth International Conference on Language Resources and Evaluation*, páginas 1313–1316, Lisboa, 26–28 May 2004. Citado na página 51.
- Biber, D. (1993). Representativeness in corpus design. *Literary and Linguistic Computing*, 8(4):243–257. Citado na página 38.
- Blandford, A. E. (1994). Teaching through collaborative problem solving. *Journal of Artificial Intelligence in Education*, 5(1):51–84. Citado na página 9.
- Bourigault, D. e Jacquemin, C. (1999). Term extraction + term clustering: An integrated platform for computer-aided terminology. Em *Proceedings of EACL 1999*, páginas 15–22. Citado na página 73.
- Brna, P. (1998). Modelos de colaboração. Em *Anais do IV WIE-SBC 1998*. Citado na página 169.
- Bruckschen, M., Saccol, D., Muniz, F., Souza, J. G. C., Fuchs, J. T., Infante, K., Muniz, M., Gonçalves, P. N., Vieira, R., e Aluísio, S. (2008). Anotação linguística em xml do corpus pln-br. Relatório Técnico NILC-TR-09-08, Universidade de São Paulo (USP). Relatórios NILC. Citado na página 109.
- Budin, G. e Melby, A. K. (2000a). Accessibility of multilingual terminological resources: Current problems and prospects for the future. Em *Second International Conference on Language Resources and Evaluation*, volume II, páginas 837–844, Athens, Greece, 31 May–2 June 2000. Citado nas páginas 88, 89, e 90.
- Budin, G. e Melby, A. K. (2000b). Accessibility of multilingual terminological resources: current problems and prospects for the future. Em *Second International Conference on Language Resources and Evaluation*, volume II. Citado na página 132.
- Cabré, M. T. (1993). *La terminología: teoría, metodología, aplicaciones*. Antártida/Empúries, Barcelona. Citado nas páginas 1, 14, e 40.
- Cabré, M. T. (2005). La terminología, una disciplina en evolución: pasado, presente y algunos elementos de futuro. *Revista Debate Terminológico*, 03(01). Citado nas páginas 2, 5, 15, 16, 18, 19, 25, 93, e 157.
- Cano, W. M. (2001). *Teoria e prática de um dicionário escolar de ciências*. Tese de Doutorado, FCLAR-UNESP, Araraquara-SP. Citado na página 1.
- Carroll, J. M. (1992). Getting around the task-artifact cycle: How to make claims and design by scenario. *ACM Transactions on Information Systems*, 10(2):181–212. Citado na página 179.
- Carroll, J. M. (1995). *Scenario-Based Design: Envisioning Work and Technology in System Development*. John Wiley & Sons, New York–EUA. Citado nas páginas 178 e 179.
- Carroll, J. M. (2000). *Making Use: Scenario-Based Design of Human-Computer Interactions*. The MIT Press, Massachusetts–EUA. Citado na página 178.

- Caseli, H. M. (2003). Alinhamento sentencial de textos paralelos português-inglês. Dissertação de Mestrado, ICMC-USP. Citado na página 101.
- Caseli, H. M., Nunes, M. G. V., e Forcada, M. L. (2005). Evaluating the lihl lexical aligner on spanish, brazilian portuguese and basque parallel texts. Em *Proceedings of the XXI Congreso de la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN)*, páginas 1–8, Granada, Spain. Citado na página 101.
- Correia, M. e Andrade, A. M. R. (2002). Desenho de um corpus de especialidade: a propósito do projecto termináutica. Em *VIII Simpósio Ibero-Americano de Terminologia*, Cartagena de Índias, Colômbia. Citado na página 43.
- Daille, B. (1996). Study and implementation of combined techniques for automatic extraction of terminology. Em Klavans, J. e Resnik, P., editores, *The Balancing Act: Combining Symbolic and Statistical Aroaches to Language*, páginas 49–66. The MIT Press, Cambridge, Massachusetts. Citado nas páginas 59, 61, 64, e 118.
- Daille, B. (2003). Conceptual structuring through term variations. Em *Proceedings of ACL 2003*. Citado na página 118.
- Dias, C. A. (2000). Terminologia: conceitos e aplicações. *Ciência da Informação*, 29(1):90–92. Citado nas páginas 2, 14, e 64.
- Dugan-Junior, R. F., Glinert, E. P., e Rogers, E. H. (2003). Camelot: Technology focused testing of cscw applications. *Hawaii International Conference on System Sciences*, 9:328c. Citado na página 174.
- Dzida, W. e Freitag, R. (1998). Making use of scenarios for validating analysis and design. *IEEE Transactions on Software Engineering*, 24(12):1182–1196. Citado na página 178.
- Ehrlich, K. (1999). Designing groupware applications: A work-centered design approach. Em Beaudouin-Lafon, M., editor, *Computer Supported Co-operative Work*, volume 7 of *Trends in Software*, páginas 1–28. John Wiley & Sons. Citado nas páginas 7 e 170.
- Eseryel, D., Ganesan, R., e Edmonds, G. S. (2002). Review of computer-supported collaborative work systems. *Educational Technology & Society*, 5(2). Citado nas páginas 6, 170, 173, e 174.
- EURODICAUTOM (2004). *Eurodicautom: Users' Guide*. European Terminology Database of EUROPA. Gateway to the European Union. Disponível em: <http://europa.eu.int/eurodicautom/Controller>. Citado nas páginas 5, 17, e 24.
- Farquhar, A., Fikes, R., Pratt, W., e Rice, J. (1995). Collaborative ontology construction for information integration. Citado na página 75.
- Farquhar, A., Fikes, R., e Rice, J. (1996). The ontolingua server: A tool for collaborative ontology construction. Citado na página 70.
- Farquhar, A., Fikes, R., e Rice, J. (1997). Tools for assembling modular ontologies in ontolingua. Citado na página 70.
- Feliu, J. (2004). *Relacions conceptuais i terminologia: anàlisi i proposta de detecció semi-automàtica*. Tese de Doutorado, Institut Universitari de Lingüística Aplicada, Universitat Pompeu Fabra, Barcelona. Citado na página 49.
- Feliu, J., Vivaldi, J., e Cabré, M. T. (2004). Ontologies: A review. Em *Série Informes*, volume 34. Citado nas páginas 67, 70, e 74.
- Finatto, M. J. B. (1998). Elementos lexicográficos e enciclopédicos na definição terminológica: questões de partida. *Organon*, 12(26):19–31. Citado nas páginas 29, 30, e 83.
- Frantzy, K. T. (1997). Incorporating context information for the extraction of terms. Em *Proceedings of ACL/EACL 1997*, páginas 501–503. Citado nas páginas 64 e 118.
- Frantzy, K. T. e Ananiadou, S. (1997). Automatic term recognition using contextual cues. Em *THIRD DELOS WORKSHOP Cross-Language Information Retrieval*. Citado na página 64.
- Freitas, C. M. D. S., Chubachi, O. M., Luzzardi, P. R. G., e Cava, R. A. (2001). Introdução à visualização de informações. *RITA*, 8(2):1–16. Citado na página 125.
- Furnas, G. W. (1986). Generalized fisheye views. Em *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, páginas 16–23. Citado na página 125.

- Genovês, L. C. e Aluísio, S. M. (2005). Avaliação de ambientes de suporte à montagem automática de corpus a partir de textos da web e extração automática de termos. Relatório técnico, ICMC-USP. NILC-TR-05-15 - Relatório Técnico do ICMC No 266. Citado na página 51.
- Ghani, R., Jones, R., e Mladenic (2001). Using the web to create minority language corpora. Em *Tenth International Conference on Information and Knowledge Management (CIKM)*, páginas 279–286. Citado na página 51.
- Ghani, R., Jones, R., e Mladenic (2005). Building minority language corpora by learning to generate web search queries. *Knowledge and Information Systems*, 7:56–83. Citado na página 51.
- Gillam, L. e Tariq, M. (2004). Ontology via terminology. Em *Proceedings of Workshop on Terminology, Ontology and Knowledge Representation*. Citado na página 73.
- Gillam, L., Tariq, M., e Ahmad, K. (2005). Terminology and the construction of ontology. *Terminology*, 11(1):55–81. Citado nas páginas 67 e 73.
- Gomez-Perez, A., Corcho-Garcia, O., e Fernandez-Lopez, M. (2003). *Ontological Engineering*. Springer-Verlag New York, Inc., Secaucus, NJ, USA. Citado na página 122.
- Gregghi, J. G. (2002). Projeto e desenvolvimento de uma base de dados lexicais do português. Dissertação de Mestrado, ICMC-USP. Citado na página 133.
- Grudin, J. (1994). CSCW: History and focus. *IEEE Computer*, 27(5):19–26. Citado na página 170.
- Ha, L. A. (2004). A computer aided terminology processing system prototype. Em *Terminology, ontology & Knowledge Representation*, Université Jean-Moulin Lyon. Citado na página 58.
- Hair-Junior, J. F., Babin, B., Money, A. H., e Samouel, P. (2006). *Fundamentos e Métodos de Pesquisa em Administração*. Bookman. Citado nas páginas 161, 163, 167, 168, 169, e 188.
- Hao, M., Hsu, M., Dayal, U., e Krug, A. (1999). Visual mining large web-based hyperbolic space using hidden links. Relatório técnico, Palo Alto-HP Laboratories-Software Technology Laboratory. Citado na página 125.
- Hardman, D. (1999). The cls-framework: Reltef especificaton. Citado nas páginas 90, 91, e 132.
- Haynes, S. R., Purao, S., e Skattebo, A. L. (2004). Situating evaluation in scenarios of use. Em *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*, páginas 92–101. Citado nas páginas 174, 177, 179, e 180.
- Heid, U., Jaub, S., Kruger, K., e Hohmann, A. (1996). Term extraction with standard tools for corpus exploration. Em *4th International Congress on Terminology and Knowledge Engineering*. Citado na página 119.
- Hernández, C. P. (2002). *Explotación de los corpórea textuales informatizados para la creación de bases de datos terminológicas basadas en el conocimiento*, volume 18. Iris. Citado nas páginas xii, 2, 3, 14, 19, 23, 24, 37, 40, 41, 42, 43, 46, 47, 52, 54, 60, 62, 68, 69, 74, 75, 83, 89, 90, 91, e 93.
- Herskovic, V., Pino, J. A., Ochoa, S. F., e Antunes, P. (2007). Evaluation methods for groupware systems. Em *CRIWG 2007*, páginas 328–336. Citado nas páginas 173, 174, 176, 177, 178, e 179.
- Inxight (2005). *Inxight Star Tree*. INXIGHT SOFTWARE INCORPORATED. <http://www.inxight.com/products/oem/star-tree/>. Citado nas páginas 125 e 126.
- ISO 12620 (1999). *Terminologia: Aplicações de Computador e Categoria de Dados*. International Organization Standardization (ISO). ISO DIS 12620. Citado nas páginas 55, 74, e 90.
- ISO 704 (2000). *Terminologia: Princípios e Métodos*. International Organization Standardization (ISO). ISO WD 704.1. Citado nas páginas 34 e 153.
- ISO 9126 (2006). *The ISO 9126 Standard*. UNIVERSITÉ DE GENÈVE. Citado na página 159.
- Jeffries, R., Anderson, A., e Hendrickson, C. (2001). *Extreme Programming Installed*. Addison-Wesley. Citado na página 151.
- Junior, L. C. R. (2008). Ontolp: construção semi-automática de ontologias a partir de textos da língua portuguesa. Dissertação de Mestrado, UNISINOS-USP. Citado nas páginas 61 e 74.

- Kavanagh, J. (1995). *The Text Analyser: A Tool for Extracting Knowledge from Text*. Tese de Doutorado, University of Ottawa, Ottawa. Citado na página 47.
- Kifer, M., Lausen, G., e Wu, J. (1995). Logical foundations of object oriented and frame based languages. Em *Journal of the Association for Computing Machinery*. Citado na página 72.
- Kilgariff, A. e Grefenstette, G. (2003). Introduction to the special issue on the web as a corpus. *Computational Linguistics*, 2(3):333–347. Citado na página 108.
- Klavans, J. L. e Muresan, S. (2000). Definder: Rule-based methods for the extraction of medical terminology and their associated definitions from online text. Em *Proceedings of AMIA 2000*. Citado na página 63.
- Klavans, J. L. e Muresan, S. (2001). Evaluation of definder: A system to mine definitions from consumer-oriented medical text. Em *Proceedings of JCDL 2001*. Citado nas páginas 63 e 119.
- Krieger, M. G. (1998). Terminologia em contextos de integração: funcionalidade e fundamentos. *Organon*, 12(26):19–31. Citado nas páginas 2, 16, 17, e 87.
- Krieger, M. G. (2000). Terminologia revisitada. *D.E.L.T.A.*, 16(2):209–228. Citado nas páginas 2, 3, 18, 19, e 27.
- Krieger, M. G. (2001). O termo: questionamentos e configurações. Em *Temas de Terminologia*, páginas 62–81. Universidade/UFRGS/Humanitas/USP. Citado nas páginas 27 e 28.
- Krieger, M. G. e Bevilacqua, C. R. (2005). A pesquisa terminológica no brasil: Uma contribuição para a consolidação da área. *Revista Debate Terminológico*, 03(01). Citado nas páginas 3, 5, 20, 21, e 29.
- Krieger, M. G. e Finatto, M. J. B. (2004). *Introdução à Terminologia: teoria e prática*. Editora Contexto, São Paulo–SP. Citado nas páginas 2, 4, 14, 15, 16, 17, 18, 19, 21, 23, 26, 28, 29, 30, 31, 32, 33, 34, 35, 36, 57, 67, 80, 81, 82, 112, 131, e 148.
- Lamping, J., Rao, R., e Pirolli, P. (1995). A focus+context technique based on hyperbolic geometry for visualizing large hierarchies. Em *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*. Citado na página 125.
- Lenci, A., Busa, F., Ruimy, N., Gola, E., Monachini, M., Calzolari, N., e Zampolli, A. (2000). Linguistic specifications. simple deliverable d2. Relatório técnico, University of Pisa and Institute of Computational Linguistics of CNR. Citado nas páginas xiv e 122.
- Levy, A. Y. e Rousset, M.-C. (1996). Carin – a representation language combining horn rules and description logics. Em *European Conference on Artificial Intelligence*, páginas 323–327. Citado na página 72.
- Lino, M. T., Costa, R., e Silva, R. (2004). Extract-term: na extractor for portuguese language. Em *Proceedings of IV International Conference on Language Resources and Evaluation (LREC)*. Citado na página 118.
- Maciel, A. M. B. (1993). Termisul e terminótica. *Cadernos do IL–UFRGS*, 10:133–139. Citado nas páginas 5, 17, 21, 24, e 25.
- Maciel, A. M. B. (1998). Terminologia, linguagem de especialidade e dicionários. Em *VI Congresso Brasileiro de Linguística Aplicada*. Citado nas páginas 27, 33, e 87.
- Maciel, A. M. B. (2000). Processamento semi-automatizado da linguagem jurídica: problemas na seleção de termos. Em *Temas de Terminologia*, páginas 372–382. Universidade/UFRGS/Humanitas/USP. Citado na página 57.
- Maciel, A. M. B. (2001). O verbo: fator determinante da especificidade do termo no texto especializado. Em *II Congresso Internacional da Associação Brasileira de Linguística (ABRALIN)*. Citado na página 57.
- Maciel, A. M. B., Parmeggiani, A. A., Mallmann, F. S., e Pereira, G. A. (1994). Banco de dados de termos jurídicos-ambientais. Em *Simpósio Iberoamericano de Terminologia*, volume 2, páginas 133–136. RITERM. Citado nas páginas 5 e 25.
- Maia, B. e Sarmento, L. (2005). The corpógrafo - an experiment in designing a research and study environment for comparable corpora compilation and terminology extraction. Em *Proceedings of eCoLoRe / MeLLANGE Workshop, Resources and Tools for e-Learning in Translation and Localisation*, páginas 45–48. Citado na página 48.
- Melby, A. K. (1999). Computer applications in terminology: Method for sharing terminological data. Citado nas páginas 88 e 89.

- Melby, A. K. (2001). SALT: Standards-based access service to multilingual lexicons and terminologies. Em *Translation, theory & technology*. Citado na página 89.
- Meyer, I., Kavanagh, D. S. J., e Davidson, L. (1997). Integrating linguistic and conceptual analysis in a www-based tool for terminography. Em *ASSOCIATION FOR COMPUTERS AND THE HUMANITIES (ALLC)*. Citado nas páginas 4, 23, 24, e 48.
- Mineiro, A. (2003). What kind of multi-word units in the nautical discourse? a study based in a linguistic corpus. Em *3a. Conferência Internacional de Terminologia Marítima*. ILTEC. Citado na página 43.
- Morales, D. F. G. (2001). Termdic: Um dicionário eletrônico. Em *Temas de Terminologia*. Universidade/UFRGS/Humanitas/USP. Citado na página 17.
- Moura, M. F., Marcacini, R. M., Nogueira, B. M., Conrado, M. S., e Rezende, S. O. (2008). A proposal for building domain topic taxonomies. Em *WTI 08: Proceedings of I Workshop on Web and Text Intelligence - SBIA 08: XIX Simpósio Brasileiro de Inteligência Artificial*, volume 1, páginas 83–84. São Carlos: ICMC-USP. Citado na página 74.
- Muniz, M., Paulovich, F. V., Minghim, R., Infante, K., Muniz, F., Vieira, R., e Aluísio, S. (2007). Taming the tiger topic: an xces compliant corpus portal to generate subcorpus based on automatic text topic identification. Em *Corpus Linguistics 2007 Conference*. Citado na página 110.
- Muniz, M. C. M., Nunes, M. G. V., e Laporte, E. (2005). Unitex-pb, a set of flexible language resources for brazilian portuguese. Em *Sociedade Brasileira de Computação – SBC2005*. Citado na página 47.
- Nascimento, M. F. B. (2003). O papel dos corpora especializados na criação de bases terminológicas. Em *Razão e Emoção: Miscelânea de estudos em Homenagem a Maria Helena Mira Mateus*, volume 2, páginas 167–180. Citado nas páginas 39, 42, 45, e 58.
- Neale, D. C., Carroll, J. M., e Rosson, M. B. (2004). Evaluating computer-supported cooperative work: Models and frameworks. Em *Proceedings of the 2004 ACM conference on Computer Supported Cooperative Work 2004*, páginas 112–121. Citado nas páginas 7, 173, e 174.
- Novello, T. C. (2002). Ontologias, sistemas baseados em conhecimento e modelos de banco de dados. Citado nas páginas 67 e 73.
- Noy, N. F., Sintek, M., Decker, S., Crubézy, M., Ferguson, R. W., e Musen, M. A. (2002). Creating semantic web contents with protégé-2000. *IEEE Intelligent Systems*, 16(2):60–71. Citado na página 125.
- Nunes, M. G. V. e Junior, O. N. O. (2000). O processo de desenvolvimento do revisor gramatical regra. Em *XXVII SEMISH (XX Congresso Nacional da Sociedade Brasileira de Computação)*, volume 1, página 6. Citado nas páginas 101 e 103.
- Pardo, R. M. (2004). Critérios de construção e organização de um corpus de especialidade: O corpus técnico-científico de ortodontia. Dissertação de Mestrado, FFLCH-USP. Citado na página 104.
- Pavel, S. e Nolet, D. (2002). *Manual de Terminologia*. Translation Bureau, 165 Hôtel-de-Ville Street, 5th Floor, Suite 500 East, Hull, Canadá. Traduzido para o português por Enilde Faulstich. Citado nas páginas 6, 17, 36, 67, e 81.
- Pereira, T. S. M. e Baptista, A. A. (2004). Navegação semântica sobre uma base de metadados rdf implementada no projeto omnipaper. Em *Conferência ELPUB 2004*. Universidade do Minho. Citado na página 71.
- Pinelle, D. (2000). A survey of groupware evaluations in cscw proceedings. Em *CSCW Proceedings, 2000*. Citado na página 174.
- Pinelle, D. e Gutwin, C. (2000a). A review of groupware evaluations. Citado nas páginas 173 e 174.
- Pinelle, D. e Gutwin, C. (2000b). A survey of groupware evaluations. Em *9th IEEE Workshop on Enabling Technologies – Infrastructure for Collaborative Enterprises*, páginas 110–115. Citado na página 7.
- Pinelle, D. e Gutwin, C. (2002). Groupware walkthrough: adding context to groupware usability evaluation. Em *CHI 02: Proceedings of the SIGCHI conference on Human factors in computing systems*, páginas 455–462, New York, NY, USA. ACM Press. Citado nas páginas 176 e 179.

- Pinsonneault, A. e Kraemer, K. L. (1989). The impact of technological support on groups: An assessment of the empirical research. *Decision Support Systems*, 5(2):197–216. Citado na página 174.
- Pinto, A. S. e Oliveira, D. (2004). Extração de definições no corpógrafo. Relatório técnico, Linguateca Polo FLUP. Citado nas páginas 48, 82, e 101.
- Prates, R. O. e Raposo, A. B. (2006). Desafios para testes de usuários em sistemas colaborativos – lições de um estudo de caso. Em *VII Simpósio Sobre Fatores Humanos em Sistemas Computacionais. IHC 2006*, páginas 9–12. Citado nas páginas 7 e 173.
- PROJECT, I. (1998). Interlinguistics terminology validation. Citado na página 89.
- Rice, J., Farquhar, A., Piernot, P., e Gruber, T. (1996). Using the web instead of a window system. Em *CHI*, páginas 103–110. citeseer.ist.psu.edu/rice96using.html. Citado na página 153.
- Santos, A. D. (1998). Autoria colaborativa: Tendências. Disponível em: <http://www.dca.fee.unicam.br/courses/IA368F/1s1998/>. Citado nas páginas 7, 170, e 171.
- Sardinha, T. B. (2000). O que é um corpus representativo? DIRECT Papers 44. Disponível em: <http://www.direct.f2s.com>. Citado na página 38.
- Sardinha, T. B. (2004). *Linguística de Corpus*. Editora Manole, Barueri-SP. Citado nas páginas 37, 38, 47, e 102.
- Sarmiento, L. e Maia, B. (2003). Gestor de corpora: Um ambiente web integrado para lingüística baseada em corpora. Em *Corpora Paralelos, Aplicações e Algoritmos Associados*, páginas 25–30, Braga. Citado na página 41.
- Sarmiento, L., Maia, B., e Santos, D. (2004). The corpógrafo: a web-based environment for corpora research. Em *Proceedings of LREC 2004, Fourth International Conference on Language Resources and Evaluation*, páginas 449–452, Lisboa, 26–28 May 2004. Citado nas páginas 5, 42, 51, e 135.
- Silva, A. M. P. (2004). Alinhamento lexical de textos paralelos português-inglês. Dissertação de Mestrado, ICMC-USP. Citado na página 101.
- Sinclair, J. M. (1997). Corpus evidence in language description. Em *Wichmann, A., Fligelstone, S., McEnery, T. and Knowles G. Teaching and language corpora*, páginas 27–39. Citado na página 37.
- Skuce, D. (1996). Ikarus intelligent knowledge acquisition and retrieval universal system. Citado na página 75.
- Skuce, D. e Lethbridge, T. C. (1995). Code4 a unified system for managing conceptual knowledge. Em *International Journal of Human-Computer Studies*, volume 42, páginas 413–451. Citado nas páginas 48 e 75.
- Souza, D. S. L. (2009). Aspectos conceituais e linguísticos da definição terminológica (dt): descrição e proposta de sistematização. Dissertação de Mestrado, Universidade Federal de São Carlos - UFSCar. Citado nas páginas 32 e 142.
- Staab, S. e Maedche, A. (2001). Knowledge portals ontologies at work. Em *AI Magazine*, páginas 63–75. Citado na página 125.
- Stiemerling, O. e Cremers, A. B. (1998). The use of cooperation scenarios in the design and evaluation of a cscw system. *IEEE Transactions on Software Engineering*, 24(12):1171–1181. Citado nas páginas 174, 177, e 178.
- SystemQuirk (2003). *System Quirk: Language Engineering Workbench*. University of Surrey. Users Manual. Citado na página 134.
- Teles, V. M. (2004). *Extreme Programming*. Novatec Editora. Citado na página 151.
- Teline, M. F. (2004). Avaliação de métodos para extração automática de terminologia de textos em português. Dissertação de Mestrado, ICMC-USP. Citado nas páginas 60, 61, 62, 64, 101, 118, 121, e 223.
- Terminometro, editor (1994). *Terminometro*, França. União Latina e TermNet. Especial Programas Terminológicos. Citado nas páginas 17, 63, e 64.
- TermStar (2001). *TermStar Terminology System*. Multilingual Information Management Group. Citado na página 134.

- Translation, T. e Technology (2000). *XML representations of Lexicons and Terminologies (XLT)*. SALT Project. Citado na página 90.
- Transterm (1996). *Creation, reuse, normalisation and integration of terminologies in natural language processing systems*. Artificial Intelligence Group (AI). Citado na página 89.
- Twidale, M., Randall, D., e Bentley, R. (1994). Situated evaluation for cooperative systems. Em *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, páginas 441–452. Citado na página 174.
- Veld, M. A. A. H., Andriessen, J. H. E., e Verburg, R. M. (2003). E-magine: The development of an evaluation method to assess groupware applications. *Enabling Technologies, IEEE International Workshops on*, 0:153. Citado na página 178.
- Venâncio, L. R., Fileto, R., e Medeiros, C. B. (2003). Aplicando ontologias de objetos geográficos para facilitar navegação em gis. Em *V Simpósio Brasileiro de GeoInformática*. Citado na página 125.
- Vieira, V. H. (2006). web services de apoio a aplicações voltadas ao trabalho em grupo. Dissertação de Mestrado, ICMC-USP. Citado nas páginas 149, 169, e 172.
- Vieira, V. H. e Fortes, R. P. M. (2004). Documentação do software ws4csw (web services for csw). Relatório Técnico 263, Instituto de Ciências Matemáticas e de Computação - ICMC-USP. Citado nas páginas 7, 148, 170, 171, e 172.
- W3C (2005). *Extensible Markup Language (XML)*. The World Wide Web Consortium (W3C). <http://www.w3.org/XML>. Citado nas páginas 70 e 124.
- Wainer, J. (2007). Métodos de pesquisa quantitativa e qualitativa para a ciência computação. Em *Atualizações em Informática 2007. Sociedade Brasileira de Computação*, páginas 221–262. Tomasz Kowaltowski and Karin Breitman, Editora PUC-Rio. Citado nas páginas 159, 160, 164, e 167.
- Wainer, J. e Barsottini, C. G. N. (2004). Empirical research in csw - a review of the acm/csw conferences from 1998 to 2004. *Journal of the Brazilian Computer Society*, 13:27–36. Citado nas páginas 7, 173, e 174.
- XAIRA (2005). Using xaira under windows – reference guide to the windows client. Citado na página 42.
- XEROX (2003). *Create and Capitalize on your corporate terminology with XTS*. Temis Group. Citado na página 134.
- Zarri, G. P. (1997). Nkrl, a knowledge representation language for narrative natural language processing. Citado na página 72.
- Zavaglia, C., Oliveira, L. H. M., Nunes, M. G. V., Teline, M. F., e Aluísio, S. M. (2004). Avaliação de métodos de extração automática de termos para a construção de ontologias. Em *RITERM*. Citado na página 66.
- Zavaglia, C., Oliveira, L. H. M., Nunes, M. G. V., Teline, M. F., e Aluísio, S. M. (2005). Avaliação de métodos de extração automática de termos para a construção de ontologias. Relatório técnico, ICMC-USP. NILC-TR-05-01. Citado na página 119.

Caderno de Atividades

Perfil Gerente de Projeto

Caderno de Atividades

Perfil Terminólogo/Linguista

Caderno de Atividades

Perfil Especialista do Domínio

Questionário de Pesquisa

Distribuição de frequência das questões do Questionário de Pesquisa