

ANDRÉ COUTINHO CASTILLA

**Instrumento de Investigação Clínico-Epidemiológica em
Cardiologia Fundamentado no
Processamento de Linguagem Natural**

Tese de Doutorado apresentada à Faculdade de
Medicina da Universidade de São Paulo para
obtenção do Título de Doutor em Ciências

Área de Concentração: Cardiologia

Orientador: Prof. Dr. Sérgio Shiguemi Furuie

São Paulo

2007

ANDRÉ COUTINHO CASTILLA

**Instrumento de Investigação Clínico-Epidemiológica em
Cardiologia Fundamentado no
Processamento de Linguagem Natural**

Tese de Doutorado apresentada à Faculdade de
Medicina da Universidade de São Paulo para
obtenção do Título de Doutor em Ciências

Área de Concentração: Cardiologia

Orientador: Prof. Dr. Sérgio Shiguemi Furuie

São Paulo

2007

Dados Internacionais de Catalogação na Publicação (CIP)

Preparada pela Biblioteca da Faculdade de Medicina da Universidade de São Paulo

©reprodução autorizada pelo autor

Castilla, André Coutinho

Instrumento de investigação clínico-epidemiológica em cardiologia fundamentado no processamento de linguagem natural / André Coutinho Castilla. -- São Paulo, 2007.

Tese(doutorado)--Faculdade de Medicina da Universidade de São Paulo.
Departamento de Cardio-Pneumologia.

Área de concentração: Cardiologia.

Orientador: Sérgio Shiguemi Furuie.

Descritores: 1.Sistemas de informação 2.Processamento automatizado de dados
3.Sistemas computadorizados de registros médicos 4.Radiologia 5.Radiografia torácica
6.Processamento de linguagem natural 7.Armazenamento e recuperação da informação

USP/FM/SBD-177/07

DEDICATÓRIA

Dedico esta tese aos meus avós José e Maisa que me apontaram o caminho da Medicina.

Aos meus pais Carlos e Maria Stella pela minha educação.

À Fernanda por todos estes anos de compreensão e companheirismo.

AGRADECIMENTOS

Agradeço a todos os amigos e colegas pelo suporte e companheirismo durante o desenvolvimento deste trabalho. Em especial agradeço ao meu orientador Prof. Dr. Sérgio Shiguemi Furuie e a Alice Shimada Bacic, parceira na linha de pesquisa.

A todos meus colegas da informática do InCor principalmente Marina Rebelo, Ramon Moreno, Marcelo Santos e Nivaldo Bertozzo, que me auxiliaram em diversos momentos desta tese.

Agradeço aos professores Lincoln de Assis Moura Jr e Ana Maria Malik por indicarem o caminho certo.

A Dra. Carol Friedman e Dra. Eneida Mendonça da Universidade de Columbia por permitirem o uso do MedLee e me orientarem a respeito de seu uso e como elaborar uma pesquisa clínica com esta ferramenta.

Agradecimentos especiais aos colegas radiologistas que se dispuseram a realizar as avaliações indispensáveis a esta tese: Fernando Sardinha, Gustavo Alcoforado, José Carlos Piccina, José Pedro Zorowich e Meire Izaki.

A todos demais que contribuíram direta ou indiretamente para a conclusão deste trabalho.

SUMÁRIO

Lista de Figuras

Lista de Tabelas

Lista de Abreviaturas e Siglas

Resumo

Abstract

1	Introdução	1
1.1	Objetivos	4
1.2	Delineamento da Tese	5
2	Revisão Teórica	6
2.1	Terminologia Médica	6
2.1.1	UMLS	7
2.1.2	Terminologias em Radiologia	9
2.1.3	Glossário Fleischner	9
2.1.4	RadLex	10
2.2	Processamento da Linguagem Natural	11
2.2.1	Avaliação de PLN em Medicina	13
2.3	Tradução Automatizada	14
2.3.1	Avaliação da Tradução Automatizada	15
2.4	Ontologias	16

2.4.1	Princípios de Ontologias em Radiologia	18
2.4.2	Protégé	24
2.5	Agentes Lógicos	25
2.5.1	Inferência em Algernon	28
2.6	Representação das Informações	30
2.7	Recuperação e Extração de Informações	32
2.7.1	Acurácia e Reprodutibilidade de Testes Diagnósticos	34
2.7.2	Medidas da Concordância	36
2.8	Processamento de Linguagem Natural em Medicina	36
2.8.1	SAPHIRE	38
2.8.2	SimText	38
2.8.3	MEDLEE	39
3	Relatórios de Radiologia Torácica	42
3.1	Metodologia	42
3.2	Resultados	44
3.3	Discussão	44
4	Dicionário para Tradução Automatizada	47
4.1	Dicionários do SYSTRAN	47
4.2	Metodologia	49
4.3	Resultados	49
4.4	Discussão	51
5	Tradução Automatizada de Textos Médicos	52
5.1	Avaliação Inicial	52
5.1.1	Metodologia	53
5.1.2	Resultados	53

5.1.3	Discussão	54
5.2	Avaliação Automática	55
5.2.1	Metodologia	55
5.2.2	Resultados	56
5.2.3	Discussão	57
5.3	Avaliação Humana	59
5.3.1	Metodologia	60
5.3.2	Resultados	60
5.3.3	Discussão	62
6	Implementação de Inferência Baseada em Ontologia	64
6.1	Construção da Ontologia	64
6.1.1	Estrutura de entrada	65
6.1.2	Estrutura do RADLEX	66
6.2	Raciocínio Lógico	67
6.2.1	Declaração de Instâncias	67
6.2.2	Declaração de Regras	68
6.2.3	Consultas	69
6.2.4	Funcionamento da Aplicação	70
6.3	Discussão	72
7	Resultados da Extração e Recuperação de Informações	75
7.1	Extração de Informações de Sentenças Médicas	75
7.1.1	Metodologia	75
7.1.2	Resultados	76
7.2	Recuperação de Informações de Textos Médicos	77
7.2.1	Metodologia	77

7.2.2	Resultados	80
8	Discussão	84
9	Conclusões	88
	Anexo - Produção Bibliográfica	92
	Referências	92

LISTA DE FIGURAS

1	Projeto	3
2	Continuantes e Ocorrentes	18
3	Derrame pleural e Cardiomegalia	20
4	Relações	22
5	Protégé	24
6	Mundo real e representação	25
7	Árvore Genealógica do Adam	27
8	Árvore Genealógica da Família	29
9	Relatório	31
10	XML	31
11	MEDLEE XML	32
12	XSLT	33
13	Instruções <i>Algernon</i>	33
14	MEDLEE	39
15	Gerenciador de dicionários	48
16	Sentença Compreensível	53
17	Sentença Incompreensível	54
18	Avaliação automática	57
19	RADLEX	66
20	Adicionar instância	68
21	REA	68
22	RER	69

23	Assertivas	70
24	Termos simples	70
25	Termos compostos	71
26	Instanciamento genérico	71
27	Instanciamento normal	71
28	Consulta à KB	72
29	Sintaxe Arden	73
30	Consulta Algernon	73
31	Relacionamento através de regras	74
32	Escore 3	76
33	Escore 0	77
34	Interface	80

LISTA DE TABELAS

1	Conceito e termos do UMLS	8
2	Tipos Semânticos	8
3	Superclasses RADLEX	10
4	RADLEX	11
5	<i>Tokens</i>	12
6	<i>Tags</i>	12
7	Relações	23
8	Tabela de contingência	34
9	PLN em Medicina	41
10	Rótulos	43
11	Classes treino	44
12	Classes teste	45
13	Memória de tradução	50
14	Grupos de termos	58
15	Escores NIST	58
16	O escore global (S) e a média (Fm) de cada grupo.	58
17	Compreensibilidade	61
18	Compreensibilidade sem erros	61
19	Fidelidade	61
20	Cobertura	62
21	Erros de tradução	62
22	Slots sentence	65

23	Slots RADLEX	67
24	Achados	79
25	Prevalência dos achados	81
26	Resultados	82
27	Acuracia	82
28	Concordância	82
29	ERROS	83

LISTA DE ABREVIATURAS E SIGLAS

CID	– Classificação Internacional das Doenças
DECS	– Descritores em Ciências da Saúde
IR	– Recuperação de Informações (do inglês " <i>Information Retrieval</i> ")
KB	– Base de Conhecimentos (do inglês " <i>Knowledge Base</i> ")
MEDLEE	– Medical Language Extraction and Encoding System
MESH	– Medical Subject Headings
PLN	– Processamento de Linguagem Natural
REA	– Regra de Encadeamento Anterógrado
RER	– Regra de Encadeamento Retrógrado
REP	– Registro Eletrônico do Paciente
RRT	– Relatórios de Radiografia Torácica
SUS	– Sistema Único de Saúde
TA	– Tradução Automatizada
UMLS	– Unified Medical Language System
XML	– Extended Markup Language
XSL	– XML Stylesheet Language
XSLT	– XML Stylesheet transformation

RESUMO

O registro eletrônico do paciente (REP) está sendo gradativamente implantado no meio médico hospitalar. Grande parte das informações essenciais do REP está armazenada na forma de texto narrativo livre, dificultando operações de procura, análise e comparação de dados. O processamento de linguagem natural (PLN) refere-se a um conjunto de técnicas computacionais, cujo objetivo é a análise de texto através de conhecimentos léxicos, gramaticais e semânticos. O presente projeto propõe a criação de uma ferramenta computacional de investigação clínico-epidemiológica aplicada a textos narrativos médicos.

Como metodologia propomos a utilização do processador de linguagem natural especializado em medicina MEDLEE desenvolvido para textos em Inglês. Para que seu uso seja possível textos médicos em Português são traduzidos ao Inglês automaticamente. A tradução automatizada (TA) é realizada utilizando o aplicativo baseado em regras SYSTRAN especialmente configurado para processar textos médicos através da incorporação de terminologias especializadas. O resultado desta seqüência de TA e PLN são informações conceituais que serão investigadas à procura de achados clínicos pré-definidos, através de inferência lógica sobre uma ontologia.

O objetivo experimental desta tese foi conduzir um estudo de recuperação de informações em um conjunto de 12.869 relatórios de radiografias torácicas à procura de vinte e dois achados clínicos e radiológicas. A sensibilidade e especificidade médias obtidas em comparação com referência formada pela opinião de três médicos radiologistas foram de 0,91 e 0,99 respectivamente.

Os resultados obtidos indicam a viabilidade da procura de achados clínicos em relatórios de radiografias torácicas através desta metodologia de acoplamento da TA e PLN. Conseqüentemente em trabalhos futuros poderá ser ampliado o número de achados investigados, estendida a metodologia para textos de outras modalidades, bem como de outros idiomas.

ABSTRACT

The Electronic Medical Record (EMR) is gradually replacing paper storage on clinical care settings. Most of essential information contained on EMR is stored as free narrative text, imposing several difficulties on automated data extraction and retrieval. Natural language processing (NLP) refers to computational linguistics tools, whose main objective is text analysis using lexical, grammatical and semantic knowledge. This project describes the creation of a computational tool for clinical and epidemiologic queries on narrative medical texts.

The proposed methodology uses the specialized natural language processor MEDLEE developed for English language. To use this processor on Portuguese medical texts chest x-ray reports were Machine Translated into English. The machine translation (MT) was performed by SYSTRAN software, a rule based system customized with a specialized lexicon developed for this project. The result of serial coupling of MT and NLP is tagged text which needs further investigation for extracting clinical findings, which was done by logical inference upon an ontology.

The experimental objective of this thesis project was to investigate twenty-two clinical and radiological findings on 12.869 chest x-rays reports. Estimated sensitivity and specificity were 0.91 and 0.99 respectively. The gold standard reference was formed by the opinion of three radiologists.

The obtained results indicate the viability of extracting clinical findings from chest x-ray reports using the proposed methodology through coupling MT and NLP. Consequently on future works the number of investigated conditions could be expanded. It is also possible to use this methodology on other medical texts, and on texts of other languages.

1 INTRODUÇÃO

Atualmente passamos por um período de transição do modelo de armazenamento de informações no meio de saúde. Gradativamente, observa-se a substituição do meio físico de papel para o meio digital sob a forma do registro eletrônico do paciente (REP). Uma das grandes vantagens do armazenamento digital é a possibilidade de se realizar buscas e consultas às informações nele armazenadas. Ao ramo da ciência cujo objeto é a procura de elementos em algum meio em resposta a uma consulta de um usuário é dado o nome de Recuperação de Informações¹.

Grande parte das informações clínicas do REP está armazenada na forma de texto narrativo livre, dificultando operações de procura, análise e comparação de documentos. O processamento da linguagem natural (PLN) refere-se a um conjunto de técnicas computacionais cujo objetivo é a análise de texto narrativo para extração de informações, classificação e comparação de dados.

Esta capacidade de interação direta com o texto narrativo livre leva a uma nova abordagem na pesquisa médica modificando a relação com que lidamos com as informações acumuladas em instituições. O crescente aumento das capacidades computacionais disponíveis em nosso ambiente possibilita a ampliação significativa do acesso a estes recursos analíticos outrora indisponíveis. Isto reflete a atual maturidade da moderna sociedade de informação marcada pelo uso intenso de recursos de rede onde se destacam estruturas de busca e indexação de informação e conteúdo na Internet.

Um tipo de estudo clínico implementado com sucesso no REP é a investigação de achados específicos em grande quantidade de relatórios de exames por imagem de forma automatizada². Geralmente, o PLN extrai informações relevantes dos textos dos relatórios para, a seguir, serem executadas tarefas de recuperação nesses documentos. Alguns dos projetos mais bem sucedidos foram desenvolvidos para a língua inglesa. Então, analisando os possíveis modelos de abordagem a serem investigados nos deparamos com a possibilidade de se lançar mão da tradução automatizada (TA). Como o próprio nome já indica, TA designa a tarefa computacional de

transformar o texto de um determinado idioma em outro. Apesar da TA não produzir traduções com qualidade literária apresenta o melhor desempenho em textos especializados³. É então proposta a abordagem de traduzir textos médicos em português para inglês e assim processá-los com uma ferramenta de PLN já validada e executar experimentos de extração e recuperação de informações nestes documentos, de uma forma multilingüe.

Textos médicos têm caráter eminentemente descritivo e narrativo compostos por um número de termos técnicos geralmente de uso restrito ao domínio médico em questão. Além disto, o conteúdo semântico das representações médicas da medicina tem caráter global fundamentado no conhecimento publicado de forma dispersa na literatura médica mundial predominantemente em inglês. Isto faz com que o inglês possa ser considerado como a *lingua franca* da Medicina⁴.

Nossa abordagem propõe então utilizar a TA e a PLN de forma seriada para extração e recuperação de informações de textos médicos. Propomos que a TA converta os textos médicos em português para o inglês de forma controlada, isto é, fundamentada na utilização de um dicionário de tradução especializado no domínio escolhido. Desta forma, a TA funcionará como uma fase de pré-processamento, normalizando e ajustando o texto antes de ser submetido ao PLN. O resultado do PLN ainda necessita de processamento adicional para a determinação da presença dos achados clínicos investigados. Geralmente isto é feito através de sistemas de inteligência artificial, o que também será abordado dentro do projeto. Uma visão esquemática da metodologia empregada está mostrada na figura 1.

Dentro desta linha de pesquisa restringiremos nosso domínio de estudo a textos de relatórios de radiografias de tórax. Dentro deste domínio, investigaremos e desenvolveremos a busca automatizada por diversos achados cardíacos e pulmonares comumente presentes em relatórios de radiografias torácicas. Nesta proposta apresentamos as seguintes investigações:

- Construção de um dicionário para tradução Português-Inglês especializado no domínio da radiologia do tórax.
- Avaliação qualitativa da TA de textos médicos utilizando um sistema baseado em regras associado ao dicionário especializado elaborado para este projeto.
- Aplicação e avaliação do uso do texto automaticamente traduzido do Português para o Inglês como entrada para um processador de linguagem natural médico desenvolvido para textos em Inglês.
- Estruturação de uma ontologia reutilizável no domínio da radiologia torácica para realizar operações de raciocínio lógico visando a identificação de padrões clínicos de ocorrência.

- Aplicar os tópicos acima em um experimento de extração e recuperação de informações de textos médicos, isto é executando o instrumento de investigação clínico epidemiológica em Cardiologia.

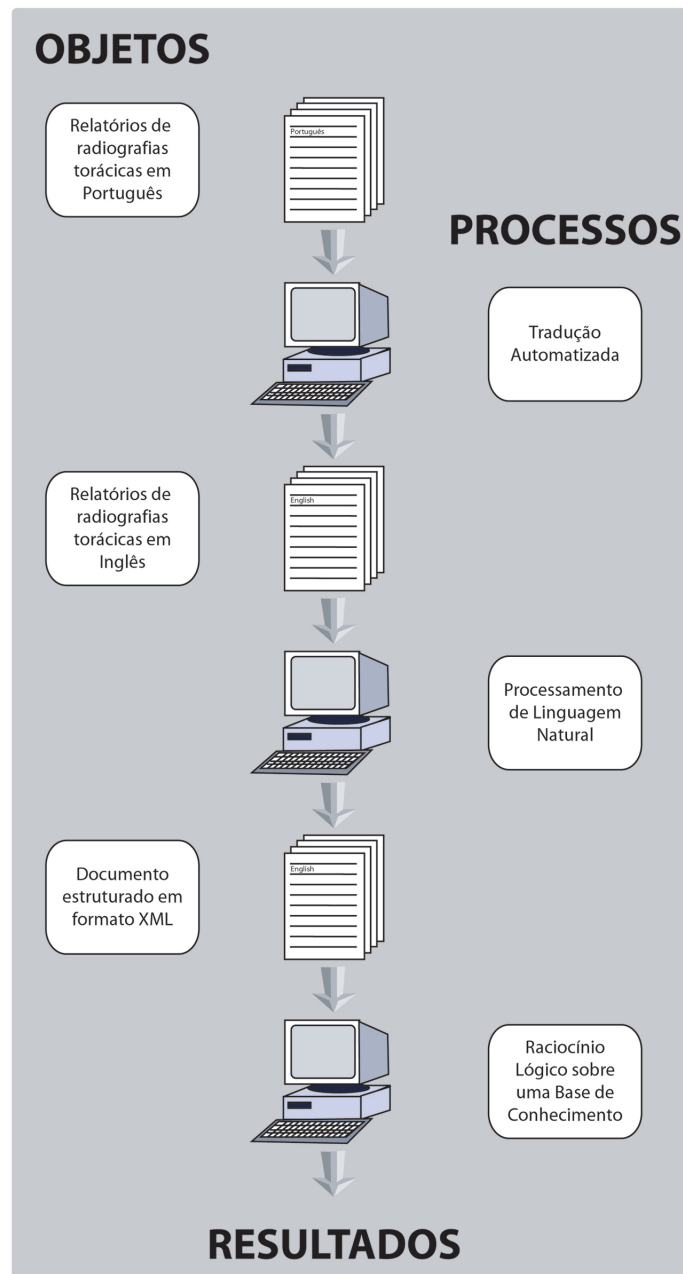


Figura 1: Visão geral da metodologia proposta para esta tese.

Além de conduzir estudos de caráter clínico, outra possível aplicação desta metodologia é a recuperação de informações de uma maneira multilingüe, isto é a possibilidade de realizar perguntas em um determinado idioma sobre um conjunto de documentos em um idioma diverso, e representar o resultado em um idioma pré-determinado. No modelo aqui apresentado as perguntas podem ser realizadas em Inglês e Português, com consultas em documentos em Português e podendo apresentar os resultados em ambos os idiomas.

Esta proposta de projeto se justifica pela necessidade do desenvolvimento de ferramentas adequadas para investigação clínica de forma automatizada nos prontuários eletrônicos já existentes. Grande quantidade de informação é gerada de forma contínua formando um fértil terreno para o desenvolvimento deste tipo de ferramenta. A partir deste tipo de pesquisa, grandes conjuntos de texto médico podem funcionar como base de conhecimentos (KB) estruturado. As informações contidas nas radiografias de tórax podem ser úteis para avaliação de tratamentos, gerenciamento de recursos, avaliação de qualidade e desenvolvimento de protocolos clínicos⁵.

A abordagem utilizada através da TA se justifica pelo desempenho que estes sistemas apresentam atualmente principalmente em domínios restritos³. A funcionalidade deste sistema já foi incorporada inclusive como ferramenta de navegação de Internet^{6,7}. O fato de a língua inglesa ser dominante no meio médico permite-nos supor que não haverá perda de expressividade do conteúdo dos textos ao se realizar esta operação de tradução. Textos em inglês nos permitirão usar o sistema de PLN MEDLEE⁸, que é um processador específico para a este idioma com desempenho significativo, e validado no domínio de relatórios de radiografias torácicas².

1.1 Objetivos

O objetivo experimental desta tese é conduzir um estudo de recuperação de informações em um conjunto de 12.869 relatórios de radiografias torácicas à procura de vinte e dois achados clínicos e radiológicos. Dentro deste objetivo apresentamos as hipóteses investigadas e as questões a serem respondidas:

1. Hipótese 1 – Pode-se traduzir automaticamente textos médicos com alta qualidade.
 - Qual o desempenho da TA de acordo com métricas reconhecidas de avaliação?
2. Hipótese 2 – Podemos associar a tradução automatizada a um processador de linguagem natural para o Inglês a fim de recuperar e extrair informações de relatórios de radiografias de tórax. Um processador de linguagem natural pode ser aplicado em textos em Português após a tradução automatizada.
 - Consegue o PLN ser aplicado em textos automaticamente traduzidos?
3. Hipótese 3 – Esta abordagem pode extrair e recuperar informações de textos médicos em Português com desempenho semelhante aos de observadores humanos.
 - Qual o desempenho comparativo entre observadores humanos e nossa abordagem em reconhecer achados clínicos específicos em relatórios de radiografias torácicas?

- Os dados coletados são confiáveis?
- Qual o desempenho do método automatizado em pesquisar uma grande quantidade de textos médicos?

1.2 Delineamento da Tese

No capítulo 2 será apresentada uma revisão teórica das diversas técnicas empregadas neste projeto de pesquisa. No capítulo seguinte, apresentaremos um breve estudo descritivo de laudos de radiografias torácicas, o objeto desta tese.

Optamos por estruturar este projeto de tese de acordo com o tipo de experimento realizado. Como a metodologia é composta por fases modulares independentes, optamos por apresentar as metodologias e descrição dos resultados de acordo com a modalidade do experimento. A discussão englobará todos os tópicos. Tal apresentação tem como modelo a tese elaborada por Mendonça⁹ que possui temática e estruturação semelhante.

No capítulo 4 será descrita a metodologia de desenvolvimento do vocabulário específico do domínio utilizado para a TA. O capítulo 5 é dedicado aos experimentos de tradução automatizada priorizando os aspectos do desempenho da tradução. O capítulo 6 é dedicado a descrição do sistema de raciocínio lógico em uma ontologia especializada em radiologia criado para este projeto e que será utilizado para analisar os resultados do processamento de linguagem. O capítulo 7 mostra experimentos de extração e recuperação de informações de textos médicos utilizando a metodologia proposta.

No capítulo 8 discute-se a metodologia, os resultados e as contribuições da tese, bem como suas limitações, e apresenta possíveis extensões do trabalho. O capítulo 9 aponta as conclusões do trabalho, ressaltando os objetivos alcançados, o trabalho futuro e as contribuições científicas.

2 REVISÃO TEÓRICA

Neste Capítulo faremos uma breve revisão teórica da literatura, das técnicas utilizadas e de experimentos relacionados com a proposta da tese. Na primeira seção apresentaremos dicionários e vocabulários médicos, fundamentais para o estudo da linguagem neste domínio. A próxima seção refere-se ao PLN, com breve descrição dos seus fundamentos teóricos e principalmente questões relativas a sua avaliação em aplicações clínicas. A seguir, apresentaremos a TA que é o primeiro passo dentro da metodologia com enfoque nas técnicas de avaliação do desempenho. Logo a seguir, descreveremos os mecanismos de inteligência artificial empregados para executar consultas aos relatórios processados. Apresentaremos, então, uma descrição de ontologias, ferramentas para seu desenvolvimento e de cálculo lógico. As metodologias de avaliação estatística de experimentos de recuperação de informações são mostradas a seguir, e logo depois, apresentaremos as diversas formas que os dados se apresentam no projeto. Por fim, descreveremos alguns sistemas de PLN empregados em pesquisa clínica, com maior atenção ao MEDLEE utilizado nesta tese.

2.1 Terminologia Médica

O uso de terminologias e classificações em medicina data do século XVII, inicialmente empregados para se tabular as causas de mortalidade de Londres da época¹⁰. Desde então, o objetivo das diversas terminologias criadas tem sido, entre outras coisas, coletar nomes de substâncias, qualidades, estruturas e processos utilizados tanto no ambiente de pesquisa como na prática clínica. As terminologias médicas refletem a diversidade dos campos biomédicos e principalmente os motivos que nortearam seu desenvolvimento. As terminologias, ou melhor, os léxicos, são o ponto de partida para estudos lingüísticos em medicina como os que suportam este projeto.

Certamente a padronização léxica e semântica permitindo o intercâmbio de informações em diversos níveis é uma das aplicações mais importantes das terminologias médicas. Nesse sentido não podemos mais pensar em terminologias isoladas não intercambiáveis ou terminologias

restritas a um único idioma¹⁰.

Tomemos como exemplo algumas das terminologias médicas mais utilizadas no meio médico-hospitalar brasileiro. As tabelas de Procedimentos e a Classificação Hierarquizada de Procedimentos, ambas da Associação Médica Brasileira, as Tabelas de Procedimentos dos Sistemas de Informação Ambulatorial e Hospitalar do SUS são exemplos de terminologias com uso específico para faturamento de procedimentos.

Já de uso global, citamos a Classificação Internacional das Doenças (CID) da Organização Mundial da Saúde, que tem com objetivo a classificação de doenças para os mais diversos fins. O *Medical Subject Headings* (MESH) e sua versão em Português os Descritores em Ciências da Saúde (DECS) são utilizados fundamentalmente para classificar artigos científicos. O *Unified Medical Language System* (UMLS) é um sistema que propõem unificar diversas terminologias em múltiplas línguas através de uma estrutura semântica comum. Uma iniciativa recente é a criação de um léxico específico para área de radiologia o RADLEX.

A seguir faremos uma breve descrição das terminologias relevantes para o trabalho aqui desenvolvido. É difícil fazer uma clara distinção entre os temas "terminologias" e "ontologias". Portanto, alguns aspectos contínuos a estes aqui descritos serão retomados em seção mais adiante.

2.1.1 UMLS

O UMLS é um projeto da *National Library of Medicine do National Institute of Health* (Bethesda, EUA) que integra diferentes fontes de conhecimento de origens diversas em único banco de dados. Seu objetivo é ser um vocabulário padrão associado a um conjunto de ferramentas para o desenvolvimento de sistemas médicos, como por exemplo, sistemas de indexação de artigos ou codificação de textos contidos no REP. O UMLS não é um vocabulário próprio e sim uma congregação de mais de duas dezenas de terminologias especializadas como o MESH, SNOMED, CID, DECS organizados em uma estrutura única. Embora o UMLS tenha sua própria estrutura semântica, hierárquica e de relacionamento, preserva a estrutura e os relacionamentos de cada vocabulário incorporado¹¹.

A principal parte do UMLS é o *Metathesaurus* que é conjunto central que unifica diversos outros vocabulários e classificações médicas em um banco de dados complexo em nove línguas.

A cada termo é designado um identificador único (SUI) que é então mapeado a um só identificador de conceito (CUI). Um conceito representa uma definição semântica e possui também um identificador único do conceito (CUI). O relacionamento entre conceitos e termos está no

formato de 1 x N, significando que cada conceito tem diversas representações do termo. Estes termos podem estar em diversos idiomas, e, portanto, este agrupamento semântico nos permite supor que os termos em diversas línguas são equivalentes e apropriados ao uso multilíngüe. Estudos já utilizaram esta propriedade para melhorar a qualidade de um sistema de TA estatístico¹². Esta estrutura permite o reconhecimento das diversas formas que um conceito pode assumir na língua escrita e uma aplicação pode identificar o termo como um conceito, qualquer que seja a língua.

Tabela 1: Termos em diversos idiomas mapeados ao conceito de *chewing gum*. Este conceito pertence ao tipo semântico *substance*.

SUI	STR	LAT
S2402143	kauwgum	DUT
S2402142	kauwgom	DUT
S0024341	chewing gum	ENG
S0024341	chewing gum	ENG
S0363185	chewing gum	ENG
S0024342	chewing gums	ENG
S0046367	gum, chewing	ENG
S0046374	gums, chewing	ENG
S1858463	purukumi	FIN
S0229117	gomme a macher	FRE
S0275481	chewing gum	FRE
S0275482	pate a macher	FRE
S1508737	kaugummi	GER
S2083076	gomma da masticare	ITA
S0435564	goma de mascar	POR
S1111589	zhevatel'naia rezinka	RUS
S0453178	goma de mascar	SPA

Uma importante característica do conceito é o tipo semântico. Este classifica os diversos conceitos do UMLS em 134 tipos que estão por sua vez relacionados através da *Semantic Network*, a segunda parte do UMLS. A tabela 2 mostra alguns tipos semânticos.

Tabela 2: Alguns tipos semânticos do UMLS.

Acquired Abnormality
Body Location or Region
Body Part, Organ or Organ Component
Body Space or Junction
Tissue

Do ponto de vista da língua Portuguesa o vocabulário é adequado para aplicações de cunho mais genérico como indexação de artigos, porém, não é profundo bastante para ser usado em

aplicações que requerem vocabulários mais específicos¹¹. A mesma deficiência já foi observada para o domínio da radiologia em relação aos termos em inglês¹³.

No atual domínio em estudo, relatórios de radiografias de tórax, muitos conceitos requeridos não existem nas definições do UMLS. Mesmo em inglês, é possível ocorrer perda de desempenho utilizando-se o UMLS ao invés de vocabulários proprietários em tarefas de recuperação de informações¹⁴. Um ponto positivo no uso do sistema UMLS é que este permite a inclusão de novos vocabulários. O RADLEX (vide abaixo) é uma terminologia especializada em radiologia e existe a previsão de ser incorporada ao UMLS.

2.1.2 Terminologias em Radiologia

São diversos os aspectos da linguagem de relatórios radiológicos que influenciam a terminologia específica deste domínio. Primeiramente, existe uma grande variabilidade em relação a terminologia utilizada em relatórios de radiografias torácicas. Uma análise deste tipo de texto encontrou em pacientes cardiopatas até 14 termos para edema/infiltrado e 23 sinônimos para designar a presença de um achado¹⁵. Porém, apesar desta profusão de termos, as terminologias médicas mais importantes como o UMLS, CID, e SNOMED apresentam um pequeno número de termos do domínio da radiologia¹³, como já mencionado.

Podemos citar algumas características próprias dos termos empregados em radiologia. Alguns tipos de termos da radiologia têm fortes características epistemológicas, isto é, requerem interpretação além de seu conteúdo de imagem. Por exemplo, o termo "lesão em moeda" tem um inegável poder descritivo da sombra produzida, porém, carece de qualquer vínculo real com a natureza da lesão no corpo do paciente¹⁶. Tipos de termos como este podem trazer alguma dificuldade na integração com outras terminologias e ontologias.

2.1.3 Glossário Fleischner

A Sociedade *Fleischner* de Radiologia propôs uma nomenclatura no domínio da radiologia torácica com o objetivo de padronizar o emprego de termos para descrição de achados radiográficos. Publicado em 1984¹⁷ contém termos em forma de dicionário ortográfico, como no exemplo mostrado a seguir. Os termos são agrupados em uma ou mais de uma categoria, neste caso, formando uma categoria complexa. As categorias são "*Radiology*", "*Radiology Physics*", "*Pathophysiology*", "*Pathology*", "*Physiology*" e "*Anatomy*". Abaixo mostramos o item referente ao hilo (pulmonar).

hilum, n, -a. **1.** *Anat.* A depression or pit in that part of an organ where the vessels and nerves enter. **2.** *Radiol.* The composite shadow at the root of each lung produced by bronchi, arteries and veins, lymph nodes, nerves, bronchial vessels, and associated areolar tissue. -Synonyms: hilus, -i, lung root. -Evaluation: **2.** A specific element of radiologic anatomy. *Hilum* and *hila* are preferred to *hilus* and *hili*. -adj. **hilar**.

Posteriormente foram publicados em inglês¹⁸ e em português¹⁹ dicionários análogos, porém, focados em termos relativos a tomografia computadorizada do tórax.

2.1.4 RadLex

O léxico RADLEX, uma iniciativa da *Radiological Society of North America* é um vocabulário especializado em radiologia cujo objetivo é oferecer uma estrutura uniforme para capturar, indexar, e recuperar uma variedade de fontes de informação neste domínio, tais como arquivos de ensino, pesquisa de dados, e relatórios de exames. Este léxico especializado pretende unificar e suplementar termos de radiologia de outras terminologias²⁰. O ponto inicial de desenvolvimento do RADLEX foi o *American College of Radiology Index of Radiological Diagnosis*, que é uma classificação numérica com um identificador anatômico e outro patológico.

No RADLEX estes campos são organizados de forma hierárquica em 12 superclasses distintas chamadas categorias. A lista destas superclasses e o exemplo de um ramo da estrutura em árvore estão mostrados tabela 3.

Tabela 3: Categorias da hierarquia mais alta do RADLEX.

Patient Identifiers
Clinical History
Image Acquisition/Processing/Display
Location on the Image
Image Quality
Anatomic Location
Findings
Relationships
Uncertainty
Conclusions
Recommendations
Teaching Attributes

A versão inicial do RADLEX está disponível na Internet²¹. Porém neste trabalho utilizamos a versão preliminar que contempla termos no domínio da radiologia do tórax, possibilitando

sua utilização neste projeto. Sua estrutura hierárquica presta-se para incorporação em uma ontologia. Em capítulo a frente será demonstrado o uso do RADLEX na forma de ontologia para realizar operações de raciocínio lógico no domínio da radiologia. A tabela 4 abaixo mostra uma análise da versão preliminar.

Tabela 4: Principais características do RADLEX.

Características do RADLEX	Quantidade
Conceitos	1326
Relacionamentos de parentesco	1302
Conceitos diferentes com strings idênticas	17
Conceitos mapeados para outros vocabulários	333
Termos mapeados com o UMLS	216
Termos mapeados com o vocabulário da ACR	93
Termos mapeados no próprio RadLex	24
Sinônimos	24

2.2 Processamento da Linguagem Natural

Segundo Chomsky²² uma linguagem é um "conjunto finito ou infinito de sentenças, cada uma de tamanho finito, todas construídas através de um alfabeto finito de símbolos". A partir deste conceito básico podemos também definir uma série de outros pontos importantes ao estudo. O termo Linguagem Natural refere-se àquelas que as pessoas falam espontaneamente como o Inglês e o Português, em oposição às linguagens artificiais como o Esperanto, linguagens de programação, ou de lógica.

Já o Processamento de Linguagem Natural (PLN) é definido como um conjunto de técnicas computacionais, que podem tanto ser de *hardware* ou *software* de análise ou síntese de linguagem escrita ou falada. O PLN escrita envolve o uso de conhecimentos do léxico, sintáticos, e semânticos do idioma bem como outras informações reais necessárias. Oferece para a pesquisa médica, uma possibilidade de, segundo Hripcksac "libertar o conteúdo clínico de relatórios narrativos"²³.

O PLN utiliza-se de técnicas lingüísticas para sua execução. A análise sintática objetiva a transformação de seqüências lineares em estruturas de elementos inter-relacionados. A análise semântica é a codificação dos elementos sintáticos em objetos da base de conhecimentos e a determinação das suas inter-relações. Em outras palavras, o conhecimento semântico é utilizado para determinar os significados das palavras e como se relacionam para dar sentido ao texto. O conhecimento do domínio contém informação a respeito do objeto em estudo. A integração do

discurso refere-se ao estudo da dependência entre as sentenças e a coerência global do texto.

Existe uma grande variabilidade nos componentes e nos tipos de conhecimentos incorporados, bem como na maneira que estão integrados os diversos sistemas PLN já aplicados à Medicina. Poucos são os sistemas que incorporam capacidades multilíngües. Do ponto de vista computacional o PLN geralmente é realizado em múltiplas etapas fundamentado na noção que o texto é dividido em parágrafos, períodos, orações, palavras, sílabas e caracteres. Nessa noção de divisibilidade do texto temos ainda que incluir outros elementos como símbolos, números e caracteres de pontuação.

Tokens são unidades significativas do texto como palavras, números ou sinais e pontuação definidos de acordo com a aplicação implementada. O *Tokenizer* é uma ferramenta básica que realiza as operações de separação nestas unidades. A tabela 5 mostra um exemplo da separação em *tokens* de uma frase, considerando inclusive a pontuação.

Tabela 5: Sentença dividida em *tokens*.

Alteração	morfológica	do	7º	arco	costal	posterior	à
direita	,	compatível	com	fratura	consolidada	.	

A delimitação de sentenças tem como tarefa a identificação de orações e parágrafos. De uma maneira simples pode-se utilizar os caracteres de pontuação habituais, mas pode estar associado a erros, como a presença de pontos para identificar abreviações.

Um *Tagger* é uma ferramenta capaz de associar cada uma das palavras do texto a um rótulo. Este rótulo pode ser tanto uma classe gramatical, por exemplo, como um outro tipo de classe definida de acordo com a implementação. A tabela 6 mostra uma frase devidamente rotulada de acordo com classes gramaticais. O significado destas classes está mostrado na tabela 10, no próximo capítulo.

Tabela 6: Sentença dividida em *tokens* rotulados.

aorta/NN	discretamente/RB	alongada/JJ	e/CC	calcificações/NNS
laminares/JJS	na/IN	croça/NN		./.

O *Stemmer* é o elemento computacional de PLN que relaciona a palavra com seu radical, isto é, com sua forma canônica. Geralmente funciona retirando sufixos ou afixos das palavras de acordo com regras morfológicas estabelecidas. Tem como função o estudo morfológico das palavras a partir do princípio da conservação semântica da palavra mesmo após derivação ou inflexão.

Desta forma podemos entender a multiplicidade de ações que são executadas para desenvolvimento de um projeto de PLN. Também são muito relevantes tarefas mais complexas como identificação de negações e desambiguação de acordo com o contexto.

2.2.1 Avaliação de PLN em Medicina

Apesar de estar presente em publicações há algum tempo ainda podemos dizer que as aplicações de PLN em Medicina estão ainda nas fases iniciais de desenvolvimento. A avaliação deste tipo de projeto é difícil ainda não são muito numerosos os estudos que possam servir de comparação. Tendo em vista estes fatores Friedman²⁴ sugere uma lista de vinte tópicos que a avaliação de sistema clínicos de PLN em medicina deve contemplar. São tópicos a respeito da minimização de viés, do estabelecimento de padrões de referência, da descrição dos métodos de avaliação, e da apresentação e discussão dos resultados.

Em relação ao viés podemos minimizar as interferências através das seguintes medidas:

1. O desenvolvedor não pode ver o conjunto de documentos de teste.
2. Se especialistas do domínio forem escolhidos para estabelecer o padrão de referência, estes não podem participar do projeto e desenvolvimento do sistema.
3. O desenvolvedor não deve realizar a avaliação.
4. O sistema de PLN deve ser fechado antes do início da fase de testes.
5. Se a capacidade de generalização do sistema está sendo testada, o desenvolvedor não deve conhecer os detalhes do estudo.
6. Idealmente, o profissional que projeta a avaliação não deve ser o mesmo que desenvolve o sistema.

Para o estabelecimento de um padrão de referência temos:

1. Se especialistas são usados para determinar um padrão de referência, deve haver um número suficiente deles para assegurar a variabilidade do padrão de referência.
2. O conjunto de testes deve ser grande o suficiente para que seja possível distinguir níveis de desempenho.
3. Se forem usados especialistas no estabelecimento do padrão de referência, o tipo de especialista deve ser apropriado (ex: radiologista x clínico).

Em relação aos métodos de avaliação:

1. O método usado para a obtenção do padrão de referência deve ser descrito claramente, principalmente se houver o envolvimento de especialistas.
2. A forma de escolha dos documentos de teste deve ser descrita.
3. Os métodos usados para calcular o desempenho devem ser descritos de forma clara, ainda mais no caso da utilização de métodos não padronizados.

Referente à apresentação de resultados:

1. Os resultados de desempenho devem ser relacionados com o conjunto inteiro de testes.
2. No caso da utilização de especialistas deve haver a troca de pareceres entre eles.
3. Os intervalos de confiança devem ser fornecidos para toda medida.

Já em relação à discussão das conclusões temos:

1. As limitações do estudo devem ser discutidas.
2. Os resultados devem ser discutidos segundo os requisitos da aplicação.
3. A generalização de resultados deve ser evitada.
4. Uma análise das falhas do sistema deve ser fornecida juntamente com a discussão das dificuldades em se obter as devidas correções.

2.3 Tradução Automatizada

O termo Tradução Automatizada (TA) refere-se a sistemas de lingüística computacional responsáveis por produzir traduções com pouca ou nenhuma interação humana. São diversas as abordagens para a realização da TA ³. A mais antiga é chamada tradução direta onde o sistema é desenhado nos seus detalhes especificamente para um par de idiomas. A segunda estratégia é a abordagem ontológica através de uma interlíngua, onde o texto é convertido em uma representação sintática comum às duas línguas. A terceira abordagem é a de transferência que utiliza duas representações abstratas intermediárias para língua de origem e destino ao contrário da anterior que só utiliza uma. Estas três abordagens citadas são chamadas abordagens baseadas em regras ³.

Existem também outras metodologias que se baseiam em coleções de textos e suas traduções pareadas ²⁵. São as chamadas abordagens baseadas em corpos bilíngües ou multilíngües paralelos. A tradução baseada em exemplos envolve o encontro de uma nova entrada num banco de dados de exemplo. A abordagem estatística depende de uma modelagem matemática de palavras nos textos de origem e destino através de aspectos de estimativa e parâmetros estatísticos de modelos lingüísticos.

Uma análise mais detalhada dos fundamentos da TA está além do objetivo deste trabalho. Por outro lado é de extrema importância o estudo dos métodos de avaliação do seu desempenho tendo em vista a necessidade de se obterem traduções de boa qualidade.

2.3.1 Avaliação da Tradução Automatizada

Podemos realizar a avaliação da TA de uma maneira manual ou automatizada. Na forma manual recursos humanos são alocados para avaliação através de métricas específicas. Avaliação manual pode ser subdividida em métricas dependentes do texto de destino somente ou dependentes tanto do texto de destino como de origem ^{26,27}.

Dentre as métricas dependentes somente do texto de destino, destacamos a legibilidade, compreensibilidade, coerência, coesão, todas estas de caráter subjetivo. Legibilidade refere-se à facilidade de como a sentença é lida. Compreensibilidade refere-se ao entendimento do texto, e a coerência ao grau de individualização das sentenças dentro do texto. Coesão refere-se ao encadeamento dos elementos textuais.

Dentre as métricas dependentes tanto do texto de origem como do destino encontram-se a cobertura, e fidelidade e vários outros critérios de boa morfologia do idioma de destino. Cobertura refere-se a quantificação dos erros e acertos de tradução dentro do domínio em questão. Fidelidade é um critério subjetivo que avalia a reprodução no texto de destino do conteúdo semântico do texto de origem. Podemos avaliar também os textos de acordo com critérios de boa escolha dos fenômenos gramaticais, morfológicos, sintáticos e léxicos da língua.

Apesar da avaliação por observadores humanos ser realmente a melhor solução ela apresenta algumas falhas, como o alto custo e problemas de reprodutibilidade de resultados. Isto motivou o desenvolvimento de técnicas automatizadas de avaliação. Geralmente são realizadas utilizando-se traduções de referência para comparação com os textos automaticamente traduzidos. As principais metodologias partem da análise da concordância de termos e seqüências de termos (N-grama¹) entre o texto avaliado e o texto referência. Dentre as chamadas Avaliações

¹Um N-grama de um termo é um unigrama, de dois é um bigrama e assim por diante.

de Co-Ocorrência de N-gramas destacamos o algoritmo BLEU e seu derivado NIST. Ambas calculam a precisão da tradução comparando-a com traduções de referência e incorporando uma penalidade de comprimento. Alguns estudos demonstraram que existe boa correlação entre o resultado destas avaliações e o julgamento por humanos²⁸.

Avaliações que utilizam estatísticas de ocorrência de N-grama requerem um corpo de avaliação alinhado com uma ou mais traduções de referência de alta qualidade. O resultado é obtido tabulando-se a fração de N-grama na tradução teste que ocorrem também na tradução de referência. O algoritmo BLEU mede a qualidade nos termos de uma soma ponderada das contagens de N-gramas que co-ocorrem segmento a segmento. Este algoritmo também inclui uma penalidade para as traduções cujos comprimentos diferem significativamente das traduções de referência. O escore NIST é uma modificação do escore BLEU original. Suas alterações incluem o uso de uma média aritmética ao invés da geométrica. Uma mudança na penalidade de brevidade foi realizada para diminuir o impacto de pequenas variações de comprimento das seqüências.²⁹ Ambos escores fornecem resultados com valores absolutos, sendo que no caso do BLEU é diretamente proporcional ao número de referências.

Um outro algoritmo de avaliação da tradução automatizada foi proposto mais recentemente. O sistema METEOR calcula a precisão e a recuperação de unigramas entre textos automaticamente traduzidos e traduções de referência. Esta avaliação é realizada em duas fases, antes e após *stemming* do tipo Porter. Este algoritmo também apresenta uma penalidade de diferença de tamanho e seu resultados encontram-se entre 0 e 1³⁰ que ao contrário dos dois algoritmos anteriores facilita a interpretação de resultados isolados.

2.4 Ontologias

Nesta parte definiremos uma ontologia, descreveremos suas principais características e abordaremos questões do desenvolvimento que podem influenciar positivamente tanto no desempenho como no poder de representação e intercâmbio com outras ontologias. Também introduziremos alguns aspectos relevantes para ontologias no campo do diagnóstico por imagem em medicina.

Uma ontologia no campo da Inteligência Artificial é 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 do domínio no sentido de que deve ser reconhecido da mesma forma por todos aqueles que o utilizam e, por ser um modelo, não faz referência às suas instâncias. Uma ontologia associada a um conjunto de instâncias individuais constitui uma base de conhecimen-

tos.

A representação do domínio é feita através de tipos (também chamados de classes ou *frames*) que são coleções de instâncias com características compartilhadas. As propriedades destes tipos (também chamados de slots) descrevem os atributos e características dos conceitos. As propriedades podem possuir restrições em seus possíveis valores.

Os tipos são o foco principal da maioria das ontologias e devem representar a realidade. Os tipos existem na realidade através das instâncias como os tipos de imagens, de processamento de imagens, de regiões cerebrais, de procedimentos clínicos, de protocolos, e assim por diante. Certamente os diferentes tipos têm instâncias consideradas limítrofes, mas certamente a maioria deve ser de instâncias claras deste tipo. Podemos então definir também uma ontologia como uma representação modular cujas unidades têm como objetivos representar os tipos na realidade e as relações entre estes tipos. Imagens médicas representam instâncias da realidade. Uma importante distinção que deve ser feita é a de conceito, que por definição, faz referência a uma representação do pensamento. Ontologias referem-se então a tipos da realidade e não a conceitos que são funções do raciocínio individual. Nesta distinção fundamenta-se o caráter globalizante e unificador das ontologias.

As ontologias também podem ser sistemas de representação baseados em *frames*. Este conceito de representação de conhecimento utiliza esta noção de frames introduzida por Minsky³¹. Um *frame* é uma estrutura que representa um tipo concreto ou abstrato. Uma importante aplicação destes sistemas baseados em *frames* são as lógicas descritivas que capturam a parte declarativa dos *frames* utilizando uma semântica lógica. Características dos sistemas baseados em *frames* são a organização em hierarquias. *Frames* também são compostos de atributos chamados de *slots*, que também podem representar relações. Apresentam também hereditariedade onde propriedades são herdadas de *superframes*, isto é, *frames* hierarquicamente superiores.

As ontologias permitem o desenvolvimento de programas capazes de lidar com informações complexas e heterogêneas como a informação médica. São úteis quando as distinções semânticas humanas são fundamentais para implementação, como nos sistemas especialistas. Estes podem ser aplicados diretamente na atenção clínica ou indiretamente auxiliando a pesquisa científica. Um tipo de aplicação que vem se destacando são as ontologias lógicas que combinam ontologias com regras de raciocínio como, por exemplo, a lógica de primeira ordem³². Essas são capazes de resolver uma série de problemas na recuperação de informações médicas, inclusive sendo empregada neste projeto de pesquisa. Esta combinação associa a representatividade da ontologia com a expressividade da sintaxe lógica.

A criação de ontologias passa por uma série de fases. Primeiramente é feita a definição

das classes, seu ordenamento em uma hierarquia taxonômica e a definição de suas propriedades e restrições. Podemos olhar o processo de criação de ontologias na área biomédica por dois aspectos. A criação *ad-hoc* de ontologias de aplicação, ou o desenvolvimento de ontologias de referência. Uma ontologia de referência é análoga a uma teoria científica, pois procura otimizar a adequação representativa do objeto em estudo. Uma ontologia de aplicação é como uma ferramenta de software, construída para objetivos específicos.

2.4.1 Princípios de Ontologias em Radiologia

As seguir apresentaremos alguns aspectos que julgamos fundamentais na construção de ontologias no domínio da radiologia e diagnóstico por imagem que recentemente foram debatidos em evento específico³³. Podemos inicialmente descrever as entidades que compõem a realidade em dois tipos principais os continuantes e os ocorrentes. Continuantes, como o nome sugere possuem existência contínua no tempo, preservando sua identidade através de mudanças de estado e que existem na totalidade sempre que existirem. Ocorrentes possuem partes temporais, desdobram-se em fases sucessivas e existem somente em suas fases. Os continuantes podem ser independentes como organismos, órgãos, arquivos, dependentes como características, atributos, qualidades, papéis, funções e um terceiro grupo de natureza espacial como ambientes ou espaços. A figura 2 exemplifica estas entidades.

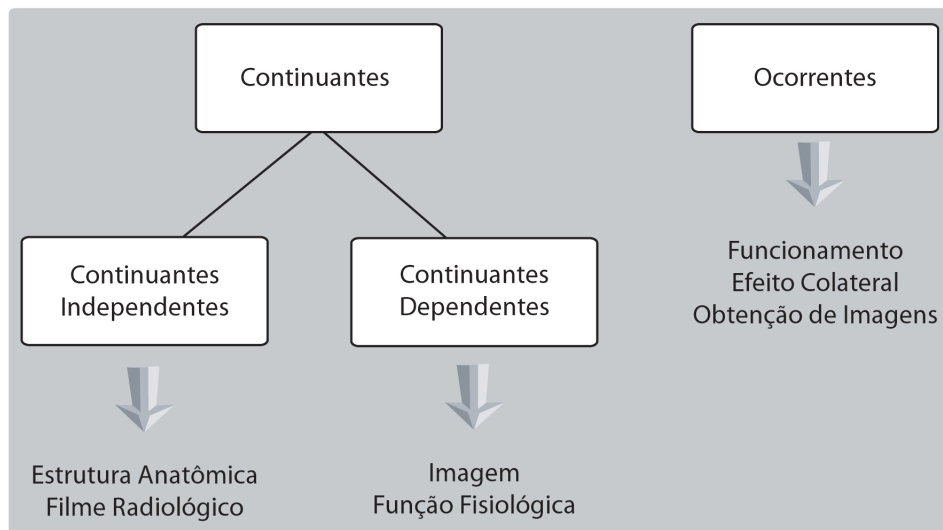


Figura 2: Continuantes e Ocorrentes.

Algumas entidades dependentes são monádicas isto é têm um único portador ou portador independente como um sentimento de depressão, uma medida de temperatura, uma qualidade da forma ou uma doença. Já algumas entidades dependentes são relacionais com mais de um portador, como um ato de fotografar (ocorrentes) que, por exemplo, relaciona uma câmera

(filme) a objetos fotografados, dois tipos de continuantes independentes. Já na Radiologia temos uma relação de referência (sobre) entre imagens que são continuantes dependentes e o objeto em estudo, um continuante independente, como no nosso caso o tórax. Alguns continuantes dependentes são executáveis como a projeção de uma película, a aplicação de uma terapia, o curso de uma doença, a execução de um algoritmo ou a expressão de um gene. Também temos que diferenciar a função do funcionamento. A função é um continuante dependente, enquanto o funcionamento é um ocorrente, como por exemplo, a função cardíaca versus o ato de bombear de um coração específico.

Quando relatamos os aspectos de uma imagem utilizamos termos que representam tipos para capturar o que nós aprendemos previamente através de exemplos, isto é, elementos da realidade, que a partir daqui denominaremos de instâncias. Na verdade, os termos são a representação de tipos para capturar o que nós sabemos sobre estes. Em radiologia médica observamos a imagem ao todo, suas regiões, os atributos de cada região, a parte da realidade que esta imagem representa, como representantes de tipos, ou melhor como instâncias.

Dentro desta perspectiva pode-se definir a interpretação de uma imagem como um mapeamento de regiões de segmentação da imagem a um par \langle instância, tipo \rangle , bem como determinar as relações entre as instâncias. Do ponto de vista formal a imagem radiológica é um padrão de radio-opacidades e radio-luscências, cristais de prata numa radiografia ou pixels no mundo digital. O que está além é apenas interpretação, isto é mapear as regiões da imagem aos tipos. Então a interpretação de imagens é baseada na informação fornecida pelas regiões da imagem e refinada pela informação subordinada de outras regiões dentro da imagem ou de identificadores dos pacientes, e principalmente pela história clínica.

Desta forma, temos uma distinção fundamental entre as entidades da imagem, que são representações bidimensionais de volumes, versus uma interpretação mais provável atribuída a elas pelo observador. Isto serve também para justificar que uma ontologia de referência para o domínio da radiologia deve ser sobre a imagem especificamente, e deve ser agregada a ontologias de referência para anatomia e patologia.

Duas regiões anatômicas podem ter dois tipos de distinção entre elas. Vamos tomar como exemplo nosso objeto de estudo, a radiografia de tórax, como mostrada na figura 3, uma radiografia de tórax. Temos aquelas onde há clara distinção anatômica entre as partes como entre o pulmão e o coração. Por outro lado existem regiões cuja distinção é arbitrada, como por exemplo, a segmentação pulmonar ou a divisão entre as câmaras cardíacas. Não existem pontos anatômicos claros de referência para se delimitar estas regiões na radiografia. Nossa representação não contempla critérios para esta diferenciação. No campo da interpretação é adicionado



Figura 3: Derrame pleural direito e Cardiomegalia. A identificação do achado de derrame pleural representa identificar este tipo e seu atributo de lado, no caso o direito. O mesmo se aplica a cardiomegalia. Porém o estabelecimento da relação causal, isto é, insuficiência cardíaca levando a transudação de líquido para o espaço pleural é uma tarefa de interpretação do observador.

conhecimento de outra natureza, não representado na imagem, neste caso o entendimento da anatomia brônquica e das câmaras cardíacas.

2.4.1.1 Relações em Ontologias de Diagnóstico por Imagem

A partir dos aspectos acima apresentado partimos para o estudo dos relacionamentos entre os diversos tipos representados neste domínio³⁴. Inicialmente ressaltamos a importância de relações comuns entre as ontologias, pois o sucesso do alinhamento entre ontologias exige que relações, *is_a* ou *part_of*, por exemplo, tenham os mesmos significados nas diferentes ontologias a serem alinhadas.

A seguir apresentaremos alguns aspectos das relações em ontologias nos domínios da anatomia e do diagnóstico por imagens. Pode-se definir três relações binárias básicas entre tipos e instâncias³⁵. São elas:

- <classe, classe> - por exemplo *is_a* - pulmão *is_a* órgão lobular
- <instancia, classe> - por exemplo *instance_of* - pulmão em particular *instance_of* classe pulmão
- <instancia, instancia> - relação no nível das instâncias como por exemplo *part_of* - instancia em particular do lobo inferior do pulmão direito *part_of* instancia particular do pulmão direito.

Dentre os relacionamentos estruturais entre entidades anatômicas temos relações de limite (*bounded_by*), partonomia (*part_of*, *regional_part_of*, *constitutional_part_of*) e relações espaciais que incluem as de posição (*located_in*, *contained_in*, *adjacent_to*), orientação (coordenada, lateralidade) e conectividade (*continuous_with*, *attached_to*). Um fato que devemos notar é que as relações que cruzam a divisão entre continuantes e ocorrentes nunca são relações *part_of*.

O estudo da radiologia leva a formulação dos seguintes pontos relevantes para investigação:

- Como passamos das imagens ao diagnóstico?
- Quais as características da imagem que levam ao diagnóstico?
- Existe alguma doença que tenha achados idênticos?
- Quais critérios são evidências de um diagnóstico?

A partir destes questionamentos podemos definir as seguintes relações entre os diferentes tipos:

- T *image_of* T1 - relação básica que une dois continuantes, onde T e T1 podem ser tipos anatômicos ou patológicos.
- T *has_feature* T1 - relação de atributo que une dois continuantes, onde T é uma entidade anatômica ou patológica da imagem, T1 é um atributo.
- T *has_location* T1 - relação básica de posição que une dois continuantes, onde T é uma característica visual ou imagem patológica e T1 é uma entidade anatômica da imagem.

A inter-relação entre atributos de imagem, patologias e anatomia como exemplificada nesta seção está representada na figura 4.

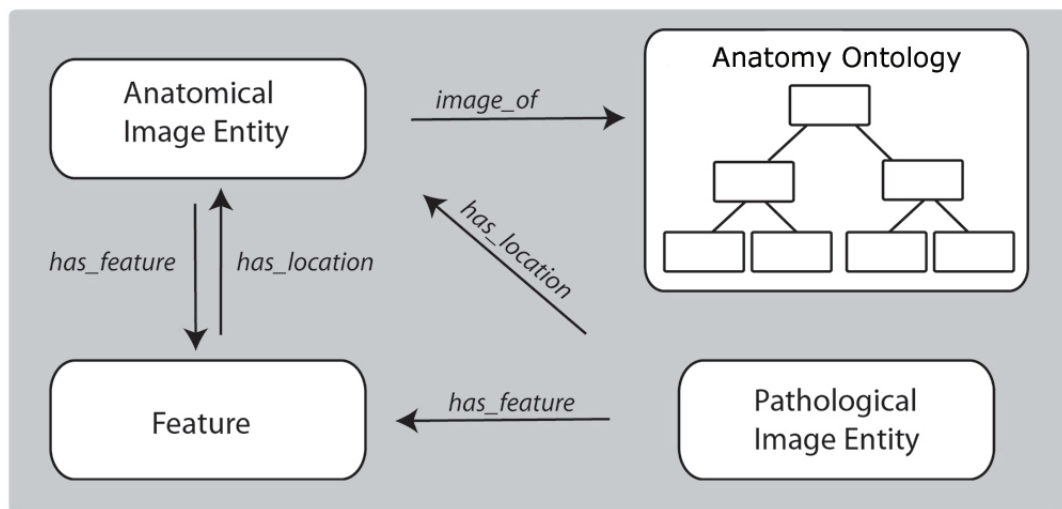


Figura 4: Representação das relações entre entidades da imagem, patologias e anatomia como apresentado nesta seção.

Podemos ter relações que vinculam propriedades a entidades anatômicas ou patológicas. Primeiramente temos rótulos de características visuais. Entre estas temos as relações morfológicas:

- T *has_shape* T1 - Adrenal [Imagem anatômica] *has_shape* redondo [Atributo: Forma]
- T *has_size* T1 - Corpo vertebral [Imagem anatômica] *has_size* reduzido em altura [Atributo: Tamanho]
- T *has_composition* T1 - Tumor [Imagem patológica] *has_composition* cística [Atributo: Composição]

Relações referentes a intensidade da imagem de acordo com a modalidade:

- T *has_density* T1 - Fígado [Imagem anatômica] *has_density* hipodensa [Atributo: Densidade]

E também relações de natureza geral:

- T *has_amount* T1 - Nódulo Pulmonar [Imagem patológica] *has_amount* múltiplos [Atributo: Quantidade]

Podemos então representar várias situações clínicas com ontologias que possuam representação da anatomia, patologia e atributos de imagem a partir deste arcabouço de entidades e relações em diagnóstico por imagem. A tabela 7 mostra alguns exemplos destas relações entre tipos conforme descrito.

Tabela 7: Relações propostas para ontologias na área de radiologia e alguns exemplos.

Tipo	Relação	Tipo
[Imagem patológica] Embolia pulmonar	<i>has_feature</i> <i>has_timing</i>	[Aspecto Geral] aguda
[Imagem patológica] [Imagem patológica] Massa	<i>has_location</i> <i>has_feature</i> <i>has_location</i> <i>has_margin</i> <i>has_shape</i> <i>has_composition</i>	[Entidade de Imagem Anatômica] [Aspecto Morfológico] lobo superior do pulmão esquerdo espiculada arredondada sólida Isto configura evidências de malignidade
[Imagem anatômica] Tireóide	<i>has_feature</i> <i>has_size</i>	[Aspecto Morfológico] aumentada
[Imagem anatômica] [Imagem patológica] Linfonodo hilar Granuloma	<i>has_feature</i> <i>has_location</i> <i>has_composition</i> <i>has_location</i>	[Aspecto Morfológico] [Imagem anatômica] calcificado lobo superior do pulmão esquerdo isto leva a evidência de tuberculose <i>has_timing</i> antiga

Assim temos entidades em dois domínios. No domínio referente ao corpo que inclui a anatomia e as doenças, e aqueles relacionados aos atributos das imagens, que são entidades dependentes do primeiro grupo. Portanto para a construção de uma ontologia no domínio do diagnóstico de imagem é necessária a definição de entidades que devem contemplar as entidades anatômicas, patológicas e atributos de imagem.

2.4.2 Protégé

O Protégé-2000 é um editor de ontologias, isto é, um programa com interface gráfica que permite realizar a construção de uma ontologia, criação de formulários personalizados para entrada de dados, e também a entrada dos mesmos dados. Também pode ser usado como uma biblioteca para outros programas, ou seja, permitir o acesso, consultas e operações em bases de conhecimento³⁶.

A partir da definição das entidades envolvidas na ontologia passamos a sua criação como subclasses de :THING. Temos que definir todas as classes ou tipos e sua estrutura hierárquica de parentesco. Com esta ferramenta também podemos criar os atributos (*slots*) das instâncias declaradas. Uma importante propriedade é a hereditariedade de *slots*. Caso criemos um *slot* em uma superclasse todas as suas subclasses herdarão automaticamente este *slot*. Também podemos definir o tipo de dados que o valor do *slot* pode assumir como número, texto ou ainda outras classes. Também define-se a cardinalidade de cada *slot*, isto é o número possível de valores que pode ser armazenado. Esta ferramenta também permite que uma série de aplicativos possam ser acoplados, como por exemplo mecanismos de raciocínio lógico. A figura 5 mostra a interface do programa.

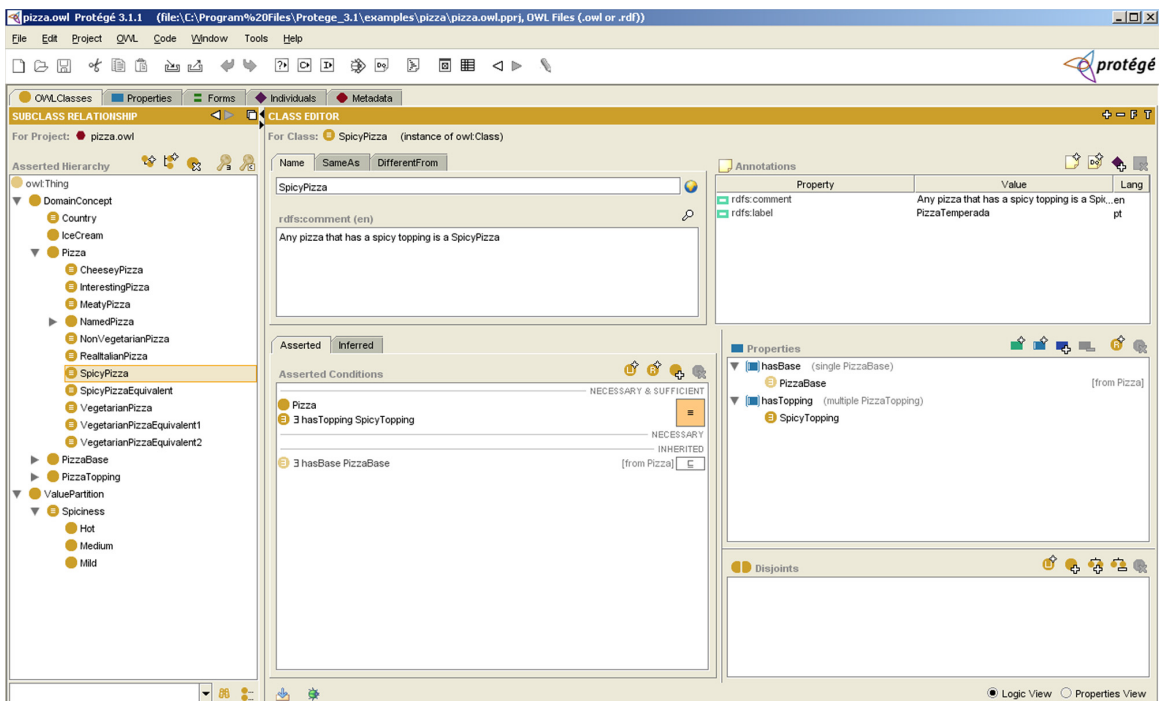


Figura 5: Interface gráfica do aplicativo Protégé

Com o aplicativo podemos criar uma ontologia e armazená-la em um arquivo específico. Este então poderá ser lido por uma série de aplicativos que terão acesso às classes, atributos, instâncias e relacionamentos declarados.

2.5 Agentes Lógicos

Não basta apenas processar textos narrativos médicos para se obter informações clínicas úteis do ponto de vista prático. Os resultados geralmente são informações codificadas que necessitam de conhecimento e raciocínio adicional. Podemos fazer uma analogia com o diagnóstico médico. Tanto a interpretação de resultados codificados como de sintomas estão fundamentados em um conjunto de conhecimentos que denominaremos como KB.

Num modelo de lógica mais simples esta KB pode ser representada por uma série de sentenças. Estas devem estar de acordo com uma sintaxe, constituindo-se de uma linguagem, de acordo com a definição de Chomsky. A sintaxe é definida por um conjunto de regras que estabelece a forma da linguagem, isto é se determinada sentença é bem formada ou não. Uma lógica também deve definir uma semântica, isto é o significado de cada sentença que engloba o conteúdo prático da mesma e que irá exercer algum efeito dentro do modelo, e deverá ser escrita com a sintaxe adequada. Mais precisamente definirá a condição de verdade ou falsidade da sentença dentro de cada mundo possível. Este mundo possível é denominado de modelo, isto é os ambientes potencialmente reais. Modelo é uma abstração matemática que representa o mundo possível que define o estado de verdadeiro ou falso de cada sentença. Em resumo, uma sentença escrita com uma determinada sintaxe possui uma semântica que será verdadeira ou falsa dentro de nosso modelo. Desta forma podemos introduzir o raciocínio lógico. Isto envolve a consequência lógica que refere-se à idéia que uma sentença possui uma continuidade de outra sentença. Desta forma se α tem consequência lógica a β se e somente se α for verdadeiro, β também é em todos os modelos. Inferência lógica refere-se ao uso das consequências lógicas para extrair conclusões. Um algoritmo de inferência consistente é aquele que deriva apenas sentenças com consequências lógicas. Será completo se conseguir derivar todas as consequências lógicas possíveis. A figura 6 ilustra a relação entre o mundo real e o virtual que os modelos lógicos se propõe a representar.

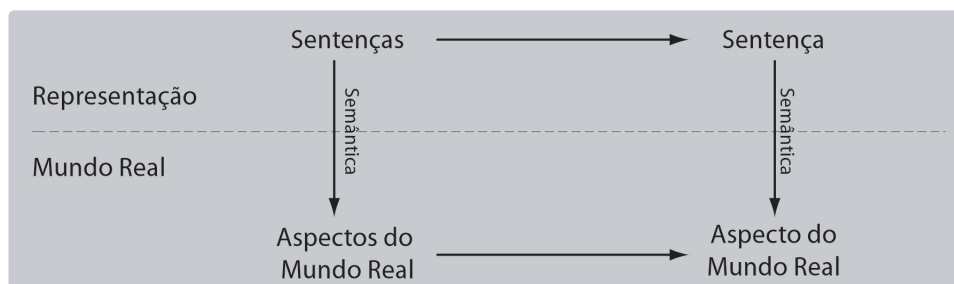


Figura 6: Mundo real e sua representação virtual

No estudo da lógica podemos descrever o comprometimento ontológico e o comprometi-

mento epistemológico. O comprometimento ontológico refere-se à natureza da realidade, isto é que existe no mundo e como pode ser descrito. O comprometimento epistemológico refere-se aos possíveis estados de conhecimento referente aos elementos da realidade, ou seja, quais os possíveis valores destes elementos.

Um modelo simples de lógica é a chamada lógica proposicional. Seu fundamento é a descrição simples de fatos e de seus valores. Desta forma, o comprometimento ontológico desta lógica é apenas a descrição dos fatos e o epistemológico suas condições, verdadeiro, falso ou ignorado.

A lógica proposicional, porém carece de capacidade de representar bem através de fatos modelos que tenham uma diversidade de objetos. Comparativamente podemos lançar mão da representatividade da linguagem como descrição da realidade.

Possíveis achados radiológicos da pneumonia são opacidades, aerobroncogramas e derrame pleural.

Analisando esta frase notamos que existem diversos elementos como objetos (doença e achados) e relações exercidas pelo verbo que une os objetos presentes nesta afirmação. Ainda temos um elemento de possibilidade que descreve algum grau de incerteza em relação a veracidade de toda informação ou de seus constituintes. Trazendo esta noção ao estudo da lógica podemos então descrever a lógica de primeira ordem que traz no seu comprometimento ontológico além de apenas fatos, também objetos e relações mantendo comprometimento epistemológico semelhante ao da lógica proposicional. Porém, neste tipo, a KB não é só de regras apenas, mas é constituída principalmente pela descrição dos objetos e suas relações. Em nosso modelo esta KB tem a forma de uma ontologia.

Neste projeto utilizaremos uma linguagem de programação lógica de primeira ordem chamada de *Algernon Abstract Machine* (MVA)³⁷. Esta implementação dá nome de frames às descrições de classes ou tipos. Os *frames* podem definir uma variedade de objetos tanto concretos como abstratos. Os *frames* são definidos essencialmente por seus relacionamentos com outros *frames*. Os relacionamentos entre frames são representados usando *slots*. Se um frame *f* estiver em um relacionamento *r* a um frame *g*, então nós definimos o valor *g* no slot de *r* de *f*. A figura 7 representa uma árvore genealógica de uma determinada família.

O frame que descreve Adam pode ser descrito como:

Adam

sexo: masculino

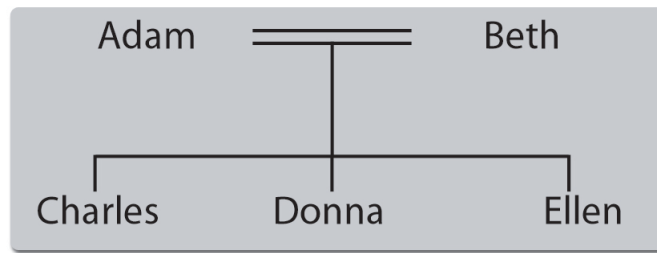


Figura 7: Família do Adam

cônjuge: Beth

filho: (Charles Donna Ellen)

Os elementos sexo, cônjuge, e filho são slots. Note que um único slot pode conter diversos valores. A árvore genealógica seria descrita então pelo menos por sete frames, descrevendo os indivíduos Adam, Beth, Charles, Donna, Ellen, masculino, e feminino. Esta representação através de frames pode ser considerada uma maneira conveniente de representar um conjunto dos predicados aplicados a símbolos constantes. Por exemplo, o frame acima podia ser escrito assim:

sexo (Adam, masculino)

sexo (Beth, feminino)

cônjuge (Adam, Beth)

filho (Adam, Charles)

filho (Adam, Donna)

filho (Adam, Ellen)

Os objetos relacionados a um frame podem ser acessados através da procura por um slot do frame sem a necessidade de procurar a KB inteira. Na MVA define-se um trajeto de acesso de rede de frames, como uma seqüência de frames cada um diretamente acessível, isto é, aparecendo em um slot de seu predecessor. Uma seqüência dos predicados define trajeto do acesso se qualquer variável que apareça como primeiro argumento de um predicado tenha aparecido previamente na seqüência. A irmã do pai de John pode ser expressada na MVA como o trajeto:

((pai John ?x) (irmã ?x ?y))

Aqui as variáveis são representadas em Algernon com nomes que começam com ?. Neste exemplo o pai de John será representado pela variável ?x e a irmã deste por ?y.

As implicações lógicas são representadas como regras. Por exemplo, a afirmação lógica que qualquer que sejam x e y , se x é cônjuge de y , y é cônjuge de x isto é:

$$\forall x; \forall y : [conjuge(x;y) \rightarrow conjuge(y;x)]$$

Podemos representar esta notação lógica habitual pela regra de encadeamento anterógrado (REA):

$$((conjuge ?x ?y) \rightarrow (conjuge ?y ?x))$$

Intuitivamente, tal regra diz que sempre que definir-se uma relação (cônjuge f g) deve-se imediatamente concluir (cônjuge g f). Tais regras chamadas de regras de encadeamento anterógrado são disparadas quando instâncias são acrescentadas a KB. A partir de fatos conhecidos todas as regras cujas premissas podem ser satisfeitas são disparadas adicionando as conclusões aos fatos conhecidos. Tomemos por exemplo a regra acima. Caso o seguinte fato seja adicionado à KB:

$$(conjuge \quad Adam \quad Beth)$$

Automaticamente a regra adicionaria a seguinte informação à KB:

$$(conjuge \quad Beth \quad Adam)$$

Algernon também permite regras de encadeamento retrógrado, como por exemplo:

$$((tia \ John \ ?y) \leftarrow (pai \ John \ ?x) \ (irmã \ ?x \ ?y))$$

Intuitivamente, esta regra diz que se você necessitar encontrar uma tia de John então você deve procurar um pai de John, e então uma irmã desse pai. Observe que o antecedente desta regra é um trajeto de acesso. Todas as regras na MVA devem definir trajetos de acesso.

2.5.1 Inferência em Algernon

São duas as operações básicas em uma KB: consultas e afirmações. As consultas recuperam o conhecimento da KB aplicando regras de encadeamento retrógrado ou através de procura por tipos instanciados. As afirmações por outro lado adicionam fatos à KB ou aplicam REA. Uma

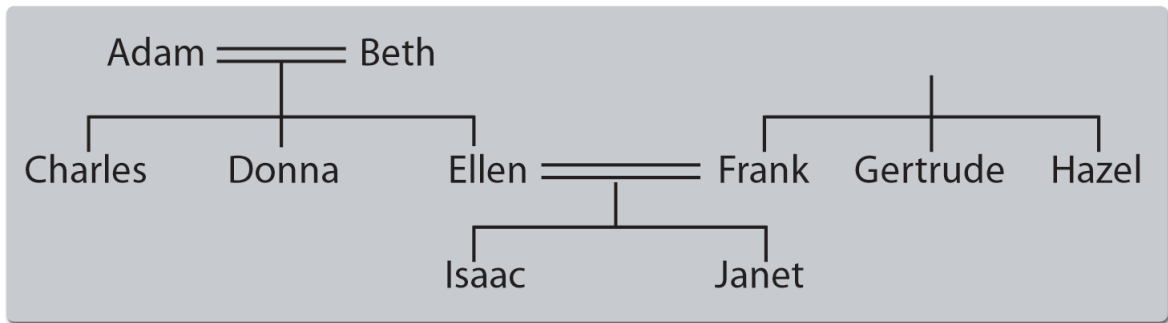


Figura 8: Família de Ellen e Frank

REA bem sucedida adiciona seu conseqüente na KB, enquanto os antecedentes de uma regra de encadeamento retrógrado (RER) são consultados. Consideremos então a genealogia mostrada na figura 8, os frames e regras que a compõe.

Existe também o frame PESSOAS, representando o conjunto de todos os indivíduos pessoas. Definimos a relação *is_a* entre cada frame de pessoas na genealogia e o frame PESSOAS. Por exemplo poderíamos declarar para todos indivíduos deste genealogia:

```
(isa PESSOAS Adam)
(isa PESSOAS Beth)
```

Podemos a seguir definir a seguinte regra para os indivíduos:

```
((tia ?z ?y) < - (pai ?z ?x) (irmã ?x ?y))
```

Formalmente isto significa:

$$(\forall z;x;y : (isa(z;PESSOAS) \rightarrow [tia(z;y) \leftarrow pai(z;x)irma(x;y)]))$$

Operacionalmente, isto quer dizer que toda a consulta ao slot tia de um frame que tenha uma relação *isa* com o frame PESSOAS resultará na aplicação desta regra. Uma regra é aplicada primeiramente consultando seu antecedente e então, se o antecedente suceder, definindo o seu conseqüente. A consulta ao antecedente de uma regra encontra todos os conjuntos consistentemente ligados às variáveis do antecedente. Isto significa que a recuperação guiada por um trajeto de acesso ramifica nos valores em um slot que fornecem ligações possíveis para uma variável. Por exemplo, se nós consultarmos:

```
(tia Janet ?x)
```

A regra acima será aplicada com ?z referido a Janet. O trajeto ramificará imediatamente nos valores no slot pai de Janet, ligações de ?x a Ellen ou Frank. Se ?x é referido a Ellen, então Algernon consultará:

(irmã Ellen ?y)

que liga ?y a Donna, assim que concluirão e definirão:

(tia Janet Donna).

Entretanto, ao longo do ramo onde x é referido a Frank, recuperação

(irmã Frank ?y)

ramificará em duas possibilidades, assim que Algernon também vai concluir e incluir na KB:

(tia Janet Gertrude)

(tia Janet Hazel)

Associar regras com os conjuntos dos indivíduos, ou com os slots que representam relações específicas, impõe-se uma organização semanticamente orientada na KB, esclarecendo a estrutura de seu conhecimento. Limita-se também o acesso a regras, reduzindo desse modo o número de vezes que seus antecedentes devem ser testados ganhando a eficiência.

2.6 Representação das Informações

A maneira como os dados são representados na forma digital é fundamental para determinar o que pode ser realizado computacionalmente. Nesta seção apresentaremos os diversos formatos com os quais os dados podem aparecer em nosso projeto de pesquisa, já mostrando alguns exemplos reais. Também apresentaremos uma ferramenta utilizada em determinado momento da metodologia para a transformação de dados numa forma adequada para o processamento subsequente.

Um relatório de radiografia torácica é um texto narrativo elaborado pelo médico radiologista decorrente de uma interpretação de uma imagem. Tem forma adequada para a compreensão do médico solicitante. Mesmo um indivíduo leigo consegue ler o texto ainda que não tenha um entendimento completo do domínio.

RADIOGRAFIA DE TÓRAX

TÉCNICA:

Exame realizado em equipamento de alta potência, com técnica de alta kilovoltagem, que fornece escala de cinza longa (baixo contraste).

DESCRIÇÃO:

Velamento do seio costo-frênico esquerdo.
 Pequenas calcificações hilares à direita.
 Vascularização pulmonar normal.
 Esternorrafia.
 Imagem cardíaca normal.
 Aorta normal.
 Arco médio normal.
 Opacidade tênue na projeção do lobo médio.

Figura 9: Exemplo de um relatório de radiografia do tórax na forma de texto livre.

Esta forma, porém, não é apropriada para o processamento computacional. Não há como um sistema de informática reconhecer quais elementos neste texto narrativo são relevantes. Esta informação teria de alguma forma ser rotulada. Veja na figura 10 como poderíamos representar este relatório como um documento XML. Este tipo de arquivo é semelhante a linguagem de hipertexto HTML da qual é feita grande parte das páginas da Internet. Um documento XML apresenta uma estrutura hierárquica de elementos, que por sua vez pode apresentar atributos e conteúdo textual.

```
<?xml version="1.0" encoding="UTF-8"?>
<relatorio nome="Fulano" tipo="radiografia" regioao="torax">
<tecnica> Exame realizado em equipamento de alta potência,
com técnica de alta kilovoltagem, que fornece escala de
cinza longa (baixo contraste).</tecnica>
<descricao>Velamento do seio costo frênico esquerdo.</descricao>
<descricao>Pequenas calcificações hilares à direita.</descricao>
<descricao>Vascularização pulmonar normal.</descricao>
<descricao>Imagem cardíaca normal.</descricao>
<descricao>Aorta normal.</descricao>
<descricao>Arco médio normal.</descricao>
<descricao>Opacidade tênue na projeção do lobo médio.</descricao>
<descricao>Esternorrafia.</descricao>
</relatorio>
```

Figura 10: Exemplo de um documento XML contendo o relatório acima apresentado.

Observe que um computador já poderia reconhecer quais as partes constituintes deste relatório de exemplo. Porém a informação clínica ainda está contida dentro do texto narrativo livre. Para isto que lançamos mão do PLN para obter conceitos e rótulos bem definidos. Veja por exemplo na figura 11 o resultado do processamento do MEDLEE das duas primeiras sentenças,

já traduzidas, da descrição do laudo de exemplo.

```

<?xml version="1.0"?><!DOCTYPE medlee SYSTEM "medlee.dtd">
<medlee>
<section c = "report description item">
<structured form = "xml">
<problem v = "shadow" idref = "p29">
<bodyloc v = "costophrenic angle" code = "UMLS:C0230151_costophrenic angle" idref = "p37">
<region v = "left" idref = "p35"></region>
<code v = "UMLS:C0504100_left costodiaphragmatic recess" idref = "p35 p37"></code>
</bodyloc>
<parsemode v = "model"></parsemode>
<sectname v = "report description item"></sectname>
<sid idref = "s3"></sid>
</problem>
<problem v="calcification" code="UMLS:C0175895_pathologic calcium deposition" idref="p55">
<bodyloc v = "hilum" code = "UMLS:C0929176_hilum" idref = "p57">
<region v = "right side" idref = "p63"></region>
<code v = "UMLS:C0929176_hilum" idref = "p57"></code>
</bodyloc>
<descriptor v = "small" idref = "p53"></descriptor>
<parsemode v = "model"></parsemode>
<sectname v = "report description item"></sectname>
<sid idref = "s5"></sid>
<code v = "UMLS:C0744900_hilar calcification" idref = "p55 p57"></code>
</problem>
</structured>
</section>
</medlee>

```

Figura 11: Resultado do processamento do MEDLEE das duas primeiras sentenças.

Observe que agora as informações relevantes estão representadas de forma estruturada, inclusive com referência a conceitos definidos no vocabulário controlado UMLS. Poderíamos então com os dados nesta forma utilizá-los para uma série de processos computacionais. Como já apresentado, o conteúdo destes arquivos será empregado para cálculos de lógica de primeira ordem. Desta forma transformaremos o arquivo XML de cada um dos laudos em instruções de lógica através de linguagem XSL que transforma documentos XML em outros tipos de documentos. Por exemplo, a figura 12 mostra uma parte do arquivo XSLT utilizado para transformar o resultado do MEDLEE em instruções de lógica. Então as duas primeiras sentenças do relatório-exemplo são transformadas pelo código ilustrado na figura 12 nas instruções de lógicas representadas na figura 13.

2.7 Recuperação e Extração de Informações

Pode se definir a recuperação de informações (IR) como parte da ciência da informação que estuda a recuperação de informação, e não de dados, a partir de uma coleção de documentos escritos. Estes documentos recuperados devem satisfazer alguma necessidade por parte do usuário que usualmente expressa sua vontade através da linguagem natural. Já a extração denota um campo do PLN voltada à busca de informações úteis dentro de textos eletrônicos através do uso

```

<?xml version="1.0"?>
<xsl:stylesheet version="1.0" xmlns:xsl="http://www.w3.org/1999/XSL/Transform">
<xsl:output method="text"/>
<xsl:template match="section">
<xsl:for-each select="structured/problem">
<xsl:text>(tell ((:add-instance (?x sentence) (tipo ?x "problem")
(tipo_value ?x "</xsl:text>
<xsl:apply-templates select="@v"/>
<xsl:text>")</xsl:text>
<xsl:for-each select="descriptor">
<xsl:text> (descriptor ?x "</xsl:text>
<xsl:apply-templates select="@v"/><xsl:text>")</xsl:text>
</xsl:for-each>
<xsl:for-each select="bodyloc">
<xsl:text> (bodyloc ?x "</xsl:text>
<xsl:apply-templates select="@v"/><xsl:text>")</xsl:text>
</xsl:for-each>
<xsl:for-each select="sid">
<xsl:text> (suid ?x "</xsl:text>
<xsl:apply-templates select="@idref"/><xsl:text>")</xsl:text>
</xsl:for-each>
<xsl:for-each select="region">
<xsl:text> (region ?x "</xsl:text>
<xsl:apply-templates select="@v"/><xsl:text>")</xsl:text>
</xsl:for-each>
<xsl:for-each select="problemdescr">
<xsl:text> (problemdescr ?x "</xsl:text>
<xsl:apply-templates select="@v"/><xsl:text>")</xsl:text>
</xsl:for-each>
<xsl:choose>
<xsl:when test="certainty">
<xsl:for-each select="certainty">
<xsl:text> (certainty ?x "</xsl:text>
<xsl:apply-templates select="@v"/>
<xsl:text>")</xsl:text>
<xsl:choose>
<xsl:when test="@v='no' ">
<xsl:text> (value2 ?x :FALSE)</xsl:text>
</xsl:when><xsl:otherwise>
<xsl:text> (value2 ?x :TRUE)</xsl:text>
</xsl:otherwise></xsl:choose>
</xsl:for-each></xsl:when>
<xsl:otherwise>
<xsl:text> (value2 ?x :TRUE)</xsl:text>
</xsl:otherwise></xsl:choose>
<xsl:text>))) &#xA;</xsl:text>
</xsl:for-each></xsl:template>
</xsl:stylesheet>

```

Figura 12: Instruções que transformam o elemento *problem* do documento XML do MEDLEE em instruções de lógica de primeira ordem do *Algernon*.

```

(tell ((:add-instance (?x sentence) (tipo ?x "problem") (tipo_value ?x "shadow")
(bodyloc ?x "costophrenic angle") (suid ?x "s3") (value2 ?x :TRUE))))
(tell ((:add-instance (?x sentence) (tipo ?x "problem") (tipo_value ?x "calcification")
(descriptor ?x "small") (bodyloc ?x "hilum") (suid ?x "s5") (value2 ?x :TRUE))))

```

Figura 13: Instruções de lógica na sintaxe do *Algernon* referente às duas primeiras sentenças do relatório de exemplo.

de termos relevantes. Desta forma podemos entender que estas duas áreas são correlatas visto que por definição poderíamos extrair informações relevantes de texto objetivando a recuperação de documentos.

A metodologia utilizada para avaliação da IR incorpora os conceitos de precisão e recuperação. A recuperação é uma medida usada para avaliar a quantidade de informação relevante recuperada por um sistema com relação à informação relevante existente na base, enquanto que a precisão é uma medida utilizada para avaliar a quantidade de informação realmente correta retornada por um sistema com relação ao total de informação recuperado pelo sistema. Estes conceitos são muito utilizados em medicina diagnóstica, pois são conhecidos respectivamente por sensibilidade e valor preditivo positivo. Do ponto de vista estatístico o resultado de nosso processamento será tratado como um método diagnóstico (identificação de achados com base no texto) que terá sua performance comparada com um padrão de referência elaborado a partir de observadores humanos especialistas.

2.7.1 Acurácia e Reprodutibilidade de Testes Diagnósticos

Investigações epidemiológicas freqüentemente necessitam classificar cada indivíduo estudado a alguma variável binária, a mais importante em relação à presença de alguma doença³⁸. Estes procedimentos de classificação são chamados de procedimentos diagnósticos. A avaliação da acurácia destes testes envolve a utilização de padrões de referência. São dois os tipos de erros que podem ocorrer em um teste diagnóstico. Vamos considerar a tabela de contingência (tabela 8) referente a um teste diagnóstico *versus* uma referência padrão ouro.

Tabela 8: Tabela de Contingência de dados de um teste diagnóstico versus uma referência.

<i>Resultado do Teste</i>	<i>Estado verdadeiro</i>		<i>Total</i>
	<i>Positivo</i>	<i>Negativo</i>	
<i>Positivo</i>	<i>a</i>	<i>b</i>	<i>a+b</i>
<i>Negativo</i>	<i>c</i>	<i>d</i>	<i>c+d</i>
<i>Total</i>	<i>a+c</i>	<i>b+d</i>	<i>n</i>

Apesar de nos referirmos a presença de doenças, este modelo pode ser atribuído a qualquer situação de definição de um resultado binário como, por exemplo, investigar a presença ou não de determinados elementos semânticos presentes em um texto. A partir da tabela podemos definir quatro tipos de situação em relação aos resultados do teste diagnóstico.

- Verdadeiro Positivo(VP): quando um objeto é classificado corretamente como positivo.

- Verdadeiro Negativo(VN): quando um objeto é classificado corretamente como negativo.
- Falso Positivo(FP): quando um objeto é classificado erroneamente como positivo.
- Falso Negativo(FN): quando um objeto é classificado erroneamente como negativo.

Então de acordo com a tabela 8 podemos calcular as seguintes taxas de erro.

$$FN = \frac{c}{a+c} \quad (2.1)$$

$$FP = \frac{b}{b+d} \quad (2.2)$$

Assim sendo, temos a taxa de fazer a correta decisão sempre que o indivíduo estudado tenha a doença, chamada de Sensibilidade (S), e a de fazer a decisão correta quando a pessoa não tem a patologia em questão chamada de Especificidade (E).

$$S = \frac{a}{a+c} \quad (2.3)$$

$$E = \frac{d}{b+d} \quad (2.4)$$

Temos que ressaltar que estas duas equações indicam respectivamente a frequência relativa de decisões corretas entre os verdadeiros positivos e os verdadeiros negativos. Também se pode calcular a taxa de um indivíduo ser portador da doença caso o teste seja positivo e não portador se o teste for negativo. Estas taxas são chamadas de valor preditivo positivo (VPP) e valor preditivo negativo (VPN) respectivamente.

$$VPP = \frac{a}{a+b} \quad (2.5)$$

$$VPN = \frac{d}{c+d} \quad (2.6)$$

Dado que as quatro medidas acima citadas são proporções pode-se calcular os intervalos de confiança para cada uma delas através da equação 2.7, onde p é a proporção, n o tamanho da amostra e V um valor crítico da distribuição normal com valor de 1,96 para um intervalo de confiança de 95%:

$$p \pm 2V \sqrt{\frac{p(1-p)}{n}} \quad (2.7)$$

2.7.2 Medidas da Concordância

Medidas de concordância têm como objetivo medir a similaridade de duas ou mais observadores em relação a alguma tarefa². Esta similaridade pode ser entre os padrões de referência nos chamados estudos de medida, semelhantes aos estudos de confiabilidade descritos abaixo. Também é possível realizar estudos de demonstração quando as referências são utilizadas em comparação para se medir o desempenho de um sistema. Um terceiro uso é medir a concordância entre sistemas testados³⁹.

Entre estas medidas de concordância podemos calcular uma proporção entre dois observadores ou concordância observada (p_o), seja de referência ou de teste. Porém, podemos supor que esta concordância seja causada por um fator do acaso. Um índice chamado kappa (κ) foi desenvolvido para medida da concordância que considera tanto a concordância observada com a concordância do acaso (p_a)⁴⁰. Está mostrado na equação 2.8

$$\kappa = \frac{p_o - p_a}{1 - p_a} \quad (2.8)$$

Consideremos então a tabela de contingência acima apresentada (tabela 8). A concordância entre os observadores é dada pela equação:

$$p_o = \frac{a + d}{n} \quad (2.9)$$

Enquanto que a concordância do acaso é calculada por:

$$p_a = \frac{(a+b)(a+c) + (d+c)(d+b)}{n^2} \quad (2.10)$$

2.8 Processamento de Linguagem Natural em Medicina

Apresentadas as técnicas empregadas neste projeto faremos uma descrição de experimentos semelhantes relatados na literatura que utilizam o PLN para extração e recuperação de informações de textos médicos livres.

Diversos tipos de sistemas foram concebidos para o domínio médico alguns mais restritos e

outros mais amplos. Projetos foram bem sucedidos em diversos idiomas como Inglês, Holandês, Alemão, Japonês e em Francês⁴¹. Para a língua portuguesa recentemente Martha⁴² e Bacic¹¹ demonstraram a possibilidade de se extrair tipos diretamente de relatórios de alta médica e de RRT respectivamente.

O desenvolvimento de sistemas de PLN aplicados à área médica tem como um dos objetivos fundamentais melhorar a funcionalidade do REP, extraíndo dados possíveis de comparação e análise por aplicações computacionais⁴³. Sistemas de PLN mostraram-se ferramentas eficazes em converter texto narrativo em dados codificados, podendo ser inclusive utilizados em tempo real. Conforme já explicado estes sistemas apresentam resultados de saída ainda complexos sendo necessário interpretação adicional. Aplicativos de inteligência artificial podem ser utilizados para converter os dados de PLN em interpretações clínicas com sentido claro.

O PLN foi utilizado em medicina em uma série de tarefas de reconhecimentos de estados clínicos em variados tipos de textos narrativos. Podemos citar codificação automática em termos do CID⁴⁴⁻⁴⁶ e SNOMED⁴⁷; o encontro de achados específicos em laudos de radiografias de tórax^{2,48}; mamografia, eletrocardiograma e ecocardiograma⁴³; encontro de achados específicos sugestivos de tuberculose em radiografias de tórax⁴⁹; classificação de severidade de pneumonias em radiografias de tórax⁵⁰; resgate de artigos da literatura a partir de dados clínicos de pacientes específicos⁹; análise de textos em urologia⁵¹; encontro de condições e estados negativos⁵²; identificação de terminologia anatômica em exames de cateterismo cardíaco⁵³; extração de informações de relatos de cirurgias⁵⁴; extração de informações de história e exame físico⁵⁵; e na resolução de ambigüidades textuais. A tabela 9 traz um resumo das principais características destes sistemas.

Para a língua portuguesa existem projetos de PLN, porém ainda é restrito o seu uso na área médica. Em uma tese recente foi descrito o uso de um aplicativo chamado SIRIMED (Sistema de Recuperação de Informações Médicas). Este aplicativo é composto de um algoritmo de semelhança semântica com o DECS e de semelhança ortográfica baseada em um algoritmo de *stemming*, ou seja de redução da palavra ao seu elemento raiz⁴². Recupera textos de relatórios de alta hospitalar a partir de consultas elaboradas pelo usuário. Em um projeto associado a esta tese Bacic¹¹ desenvolveu uma ferramenta para rotular textos de RRT em tipos do UMLS e RADLEX fundamentado. Seu desenvolvimento compartilhou os mesmos grupos de treinamento com este aqui apresentado.

A radiologia e todas as outras especialidades médicas que realizam exames por imagem são um fértil campo para o desenvolvimento de aplicativos de PLN. Estas especialidades têm como objeto de trabalho a avaliação de imagens médicas com elaboração de um relatório clínico a

partir dos achados e interpretações encontrados. Este laudo pode ser considerado a representação escrita do exame de imagem de acordo com uma semântica e conhecimento prévio do domínio específico. Uma série de tarefas de extração e recuperação de informações pode ser realizada com o uso de técnicas de PLN diretamente sobre o texto dos relatórios promovendo de uma maneira indireta acesso ao conteúdo de imagem que o texto representa. A seguir descreveremos três sistemas, todos desenvolvidos para a língua inglesa com aplicação em relatórios de exames de imagem.

2.8.1 SAPHIRE

O sistema SAPHIRE⁵⁶ foi desenvolvido inicialmente para codificar documentos como artigos científicos em conceitos relevantes. O texto é codificado através de métricas de similaridade ortográfica com o UMLS. Apesar de originalmente ter sido desenvolvido para codificar textos de literatura científica como artigos e livros foi também empregado para processar uma variedade de relatórios de radiologia geral como de tomografias, ressonância, radiografias torácicas, e cintilografias ósseas. Este foi o único projeto que apresentou também aspectos multilíngües. Como emprega o UMLS como codificação de saída já foi configurado para fornecer resultados em vários idiomas⁵⁷.

2.8.2 SimText

O *SimText*⁵ é um projeto desenvolvido para a língua inglesa e foi criado inicialmente para codificar informações em relatórios de radiografias de tórax com aplicações posteriores em relatórios de admissão hospitalar e relatórios de cintilografia pulmonar. O *SimText* possui componentes sintáticos e semânticos. A informação sintática é executada como gramáticas computacionais enquanto o conhecimento semântico do sistema é armazenado em três redes bayesianas. O elemento gramatical tem acesso à informação semântica. A primeira rede bayesiana modela os dispositivos que são descritos com frequência como cateteres e sondas. A segunda rede representa achados radiológicos como, por exemplo, opacidades, derrame pleural ou alargamento mediastinal. A terceira rede modela as doenças que são descritas diretamente como pneumonia, cardiomegalia ou atelectasia. O sistema tenta fazer uma interpretação para cada sentença no relatório usando o modelo hierárquico implicado nas redes bayesianas como modelo para as palavras e conceitos. O desenvolvimento de aplicações clínicas segue a sequência de processamento do texto seguido de algum tipo de inferência para a tomada de decisão da presença de achados clínicos selecionados.

Este sistema foi utilizado para a extração de tipos de relatórios de radiografias de tórax relacionados com pneumonia. O sistema apresentou acurácia similar a médicos especialistas em reconhecer estes tipos, e desempenho superior a pessoas leigas ou a procura simples de elementos textuais.

2.8.3 MEDLEE

Especificamente em radiologia destaca-se o aplicativo MEDLEE⁵⁸ que será utilizado como ferramenta de PLN em nossa metodologia. Inicialmente foi desenvolvido como um aplicativo de apoio à decisão clínica para relatórios e radiografias de tórax. O resultado de seu processamento é um documento XML que consiste nos achados primários clínicos junto com elementos modificadores.

Este aplicativo é composto por uma série de fases⁵⁹. Na fase de pré-processamento ocorre o reconhecimento dos limites da sentença e a busca léxica para categorização de termos, bem como sua desambiguação. Na etapa de "parsing", ou de análise gramatical, as definições gramaticais e léxicas identificam e interpretam a estrutura da sentença e elaboram uma estrutura intermediária fundamentada na gramática. A fase de composição é necessária para contruir e normalizar segmentos de múltiplas palavras para facilitar a recuperação de informações nos casos que as palavras individuais da frase foram separadas umas das outras nos textos de entrada. A codificação é fase de mapeamento dos achados a termos de algum léxico definido como o próprio do Medlee ou outro como, por exemplo, o UMLS. Por fim utiliza-se de um reconhecedor de unidades sintáticas estruturais de alto nível ("*chunker*") para aumentar a sensibilidade através de estratégias alternativas de segmentação de sentenças em caso de falha no "*parsing*" inicial. A figura 14 mostra os principais componentes do sistema.

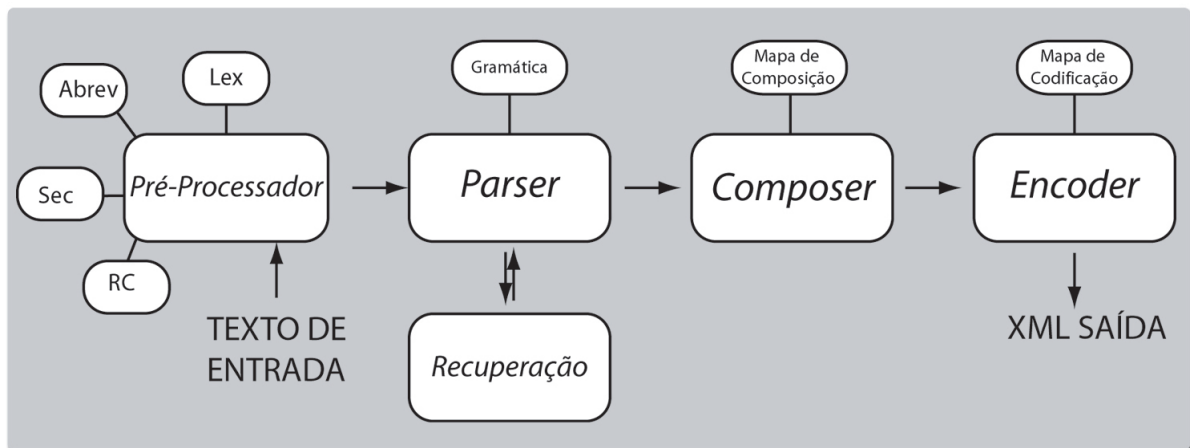


Figura 14: Diagrama dos componentes principais do MEDLEE.

O MEDLEE foi inicialmente desenvolvido para o processamento de relatórios de radiografias de tórax. Progressivamente houve a ampliação dos domínios compreendidos no sistema. Um estudo mostrou o processamento automático de 889.921 relatórios de radiografias de tórax². A análise amostral demonstrou uma sensibilidade de 81% e especificidade de 95% com encontro de achados clínicos e associações em frequências compatíveis com a literatura. Os outros domínios incorporados ao MEDLEE foram ampliados para toda radiologia, mamografia, relatórios de alta, eletrocardiografia, ecocardiografia e patologia. Uma avaliação independente do MEDLEE mostrou-o efetivo no reconhecimento de situações clínicas selecionadas ⁶⁰, bem como capacidade de processar textos de outras instituições com pequenos ajustes.

Tabela 9: Alguns sistemas de PLN em Medicina. Entrada - como o sistema é informado do que procurar. Saída - terminologia utilizada para na apresentação dos resultados. Ref. - referência bibliográfica.

Autor	Ref.	Sistema	Composição	Textos	Entrada	Saída	Achados Pesquisados	Resultados
Martha	42	SIRIMED	Similaridade semântica e ortográfica	Resumos de Alta	Consulta de usuário	DECS	Diversos	Sensibilidade – 46%
Bacic	11		Análise gramatical, léxica e semântica	Radiografias de tórax	termos do UMLS e RADLEX	UMLS e RADLEX	Diversos	Sensibilidade – 76%
Hersh	56	SAPHIRE	Similaridade ortográfica	Radiologia variados (RX, CT, RM, MN)	termos do UMLS	UMLS	Diversos	Sensibilidade – 63 % VPP – 63%
Fizman	48	SYMTEXT	Análise gramatical, léxica e semântica	Radiografias de tórax	Diagnóstico pré-estabelecido	Proprietário	Pneumonia	Sensibilidade – 95% Especificidade – 85%
Hrpicsak	2	MEDLEE	Análise gramatical, léxica e semântica	Radiografias de tórax	Diagnóstico pré-estabelecido	Proprietário ou UMLS	Diversos	Sensibilidade – 81 % Especificidade – 99%
Mendonça	59	MEDLEE	Análise gramatical, léxica e semântica	Radiografias de tórax	Diagnóstico pré-estabelecido	Proprietário ou UMLS	Pneumonia natal	Sensibilidade – 71% Especificidade – 99%

3 RELATÓRIOS DE RADIOLOGIA TORÁCICA

Como já apresentado, o objeto de estudo é a interpretação escrita feita por um médico especialista de radiografias torácicas. Estas radiografias são realizadas em diversas incidências únicas ou combinadas. As mais comuns são a pósterio-anterior ou frontal e a incidência lateral ou em perfil. A incidência de escolha é a frontal e é realizada quando possível com o paciente em posição ortostática. A escolha do exame é feita pelo médico assistente, e não houve distinção entre os diversos tipos na coleta de dados.

O grupo de relatórios utilizado neste experimento e no treinamento da metodologia é composto por 1132 RRT elaborados no Instituto do Coração nos anos de 2003 e 2004. Para cada experimento foi selecionado um subgrupo destes seja de relatórios ou de frases isoladas. Representam no total a produção de dois meses já subtraído o grande número de exames liberados sem relatório. A maioria dos exames é de pacientes ambulatoriais refletindo, portanto o padrão nosológico de um hospital de cardiologia de alta complexidade. Neste capítulo apresentaremos um experimento de análise de uma amostra destes relatórios, derivado de experimentos conduzidos em um trabalho correlato de nosso grupo¹¹.

Uma amostra de sentenças únicas do grupo terá suas palavras manualmente rotuladas por um especialista. Este grupo funcionará como treinamento para um rotulador automático (*Tagger*). Este aplicativo por sua vez realizará a rotulação de um outro grupo de sentenças. O objetivo aqui é estudar a distribuição das classes gramaticais dos termos presentes em nossos textos de pesquisa.

3.1 Metodologia

Após a extração da base de dados do hospital, os textos foram pré-processados e organizados por linha de laudo, isto é, cada frase é armazenada separadamente das demais, mas com uma identificação do laudo e da seqüência de apresentação no texto. Para extrair as frases foi

usado um algoritmo simples executado por ferramenta léxica baseada em expressões regulares que identificam o ponto final em cada frase do texto.

A estrutura dos dados permitiu a observação de muitas frases idênticas entre os laudos, sendo assim optou-se por fazer o experimento a partir de sentenças únicas ao invés de fazer por laudos, para minimizar o esforço manual de anotação.

As frases do grupo de teste foram anotadas manualmente com os rótulos utilizados no *Brown Corpus*⁶¹. Como estes rótulos foram criados para a língua inglesa, nem todos são apropriados para o português havendo necessidade de alguma adaptação. A lista dos rótulos utilizados está mostrada na tabela 10. A anotação foi realizada por um profissional da área de letras com experiência na criação e anotação de corpus, inclusive na área de biomedicina.

Tabela 10: Rótulos utilizados para anotar relatórios de radiografias torácicas

Rótulos	Descrição
/(Marcador de abre parêntesis
/)	Marcador de fecha parêntesis
/,	Marcado de vírgula como sinal de pontuação
/.	Marcado de ponto como sinal de pontuação
/?	Marcado de interrogação como sinal de pontuação
/AT	Artigo
/CC	Conjunção coordenada
/CD	Número, cardinal
/DTI	Determinante/qualificador singular ou plural
/IN	Preposição
/JJ	Adjetivo no singular
/ADJP	Adjetivo no plural
/NN	Substantivo no singular
/NNS	Substantivo no plural
/OD	Número ordinal
/PPS	Pronome possessivo
/RB	Advérbio
/VBG	Verbo presente/Particípio/Gerúndio
/VBN	Verbo no particípio passado

Foi empregado o pacote *Natural Language Tool Kit* (NLTK) que consiste em um conjunto de ferramentas lingüísticas sob a forma de uma extensão para linguagem de programação *Python*. O corpus anotado é utilizado para treinar o *Brill Tagger*¹ existente no NLTK adaptado para o português⁶². A adaptação deve levar em consideração os rótulos utilizados no corpus anotado, mas com as devidas alterações referentes aos rótulos usados para a língua portuguesa.

¹Trata-se de um algoritmo baseado em aprendizado por transformações, que faz uma rotulação inicial do texto e em um segundo passo faz a correção dos rótulos através de regras. Por se tratar de um método de aprendizado supervisionado, é necessário treinar o algoritmo com dados anotados previamente.

Após treinar o rotulador com o corpus anotado, o mesmo é usado para anotar o segundo grupo de sentenças, o grupo de teste. Além do rotulador O NLTK tem uma série de outras ferramentas de análise de texto que permitem, por exemplo, a contagem dos rótulos encontrados.

3.2 Resultados

Abaixo mostramos um exemplo de frase extraída de um laudo que foi anotado, incluindo os sinais de pontuação.

abaulamento/NN do/IN arco/NN médio/JJ./.

Foram selecionados 1132 relatórios para o experimento. São constituídos por 907 frases de grafias únicas. A repetição da mesma grafia entre frases de laudos diferentes é devido a utilização de um dicionário comum entre os radiologistas da instituição. Das 907 frases diferentes, foram anotadas manualmente 508, enquanto 399 foram pelo *Tagger*. A distribuição das classes de ambos grupos de sentenças está mostrado nas tabelas 11 e 12.

Tabela 11: Distribuição das classes do grupo de treinamento

Classes	Quantidade
Artigo	16
Conjunção	78
Número	30
Preposição	645
Adjetivo no Singular	788
Adjetivo no Plural	238
Substantivo no Singular	995
Substantivo no Plural	249
Pronome	6
Advérbio	51
Verbo	27

3.3 Discussão

Textos de relatórios de radiografias são geralmente divididos em partes. A única parte realmente fundamental e invariavelmente presente é a descrição dos achados. Dados clínicos, técnica empregada e conclusões podem ou não estar presentes. Em nosso meio os relatórios têm caráter predominantemente descritivo. Nosso modelo de estudo considera o relatório como um

Tabela 12: Distribuição das classes do grupo de teste

Classes	Quantidade
Artigo	7
Conjunção	58
Número	12
Preposição	495
Adjetivo no Singular	660
Adjetivo no Plural	223
Substantivo no Singular	813
Substantivo no Plural	143
Pronome	1
Advérbio	25
Verbo	7

conjunto finito de frases, cada uma delas relacionadas com um determinado aspecto anatômico e radiológico, representando a descrição dos achados de imagem. A somatória destas frases corresponde a um determinado significado clínico.

O conteúdo semântico das sentenças é basicamente constituído por tipos anatômicos ou patológicos relacionados a aspectos de imagem. Então podemos formular as seguintes questões relevantes para a caracterização dos conhecimentos necessários para a interpretação do relatório.

1. Quais relatórios suportam determinada patologia?
2. Quais relatórios suportam determinada região anatômica?
3. Quais relatórios suportam determinado achado?
4. Quais achados relacionam-se a determinada região anatômica?
5. Quais achados relacionam-se a determinada patologia?
6. Quais patologias relacionam-se a determinada região anatômica?

O texto do laudo é simples e objetivo segundo a construção linguística. A maioria das sentenças não possui verbos. Foram apenas encontrados 27 verbos em 508 sentenças analisadas. A sentença da figura 9 mostra a simplicidade das construções gramaticais presentes na descrição dos relatórios. A maioria das frases sem verbos é uma afirmação de um achado presente.

Em uma análise do conjunto de laudos, percebemos também que o vocabulário utilizado é mais ou menos constante. Um aspecto relevante é o fato que operacionalmente os relatórios são construídos a partir de um padrão normal.

Observamos que de um conjunto de 1132 laudos, foram obtidas 14120 frases totais sendo 930 únicas. Uma simples observação mostrou que a maioria das frases encontradas são as relacionadas com os achados da normalidade. É de extrema relevância para este tipo de projeto incluir no grupo de treinamento as frases normais. Tudo isto também reforça nossa modelagem do sistema que procura uma diferença do estado da normalidade.

4 **DICIONÁRIO PARA TRADUÇÃO AUTOMATIZADA**

Este projeto pode ser descrito como um estudo de recuperação multilingüe de informações cuja abordagem é a tradução dos documentos investigados em oposição às abordagens de tradução das consultas⁶³. A abordagem da TA por regras aqui empregada necessita de um léxico especializado, representado por um dicionário de tradução. O aplicativo SYSTRAN possui uma interface especializada para gerenciamento de terminologias especializadas. Neste capítulo apresentaremos os componentes do aplicativo de gerenciamentos dos dicionários do programa SYSTRAN, e posteriormente descrevermos a elaboração do dicionário especializado desenvolvido para o projeto.

4.1 **Dicionários do SYSTRAN**

O aplicativo gráfico de gerenciamento de dicionários tem o nome de *User Dictionary Manager* (UDM) e apresenta três tipos distintos que são dicionário do usuário, de normalização e memórias de tradução.

Os dicionários do usuário (DU) são compostos por uma tabela de origem e destino dos termos bem como outros atributos, o mais importante a categoria, que é um misto de classe gramatical e sintática. Outra importante característica é que este dicionário é unidirecional só permitindo a tradução de uma língua para a outra não permitindo o inverso. Em caso de necessidade de tradução de sentido contrário é preciso a elaboração de um novo dicionário, não apenas realizando a inversão dos termos, mas também os ajustes referentes às diferenças que os idiomas apresentam, principalmente no que se refere as derivações e inflexões. As categorias são denominadas sujeito, nome próprio, verbo, advérbio, seqüência, acrônimo, preposição e adjetivo. A definição desta categoria pode ser feita tanto pelo usuário, como também pelo programa, de forma automatizada, ainda que sujeita a erros. Também é possível a indicação de quais entradas que não devem ser traduzidas, e que ficam armazenadas em uma tabela distinta.

A figura 15 mostra a interface gráfica do aplicativo de gerenciamento com um dicionário de usuário aberto.

Português	Inglês	Categoria	Confabilidade
✓ broncogenicas	bronchogenic	Adjetivo	██████
✓ broncogenico	bronchogenic	Adjetivo	██████
✓ broncogenicos	bronchogenic	Adjetivo	██████
✓ broncografia	bronchography	Sujeito	████████
✓ broncografico	bronchographic	Adjetivo	██████
✓ broncograma	bronchogram		██████
✓ broncogramas	bronchograms		██████
✓ broncolitase	broncholithiasis		██████
✓ broncopleural	bronchopleural		██████
✓ broncopulmonar	bronchopulmonary		████████
✓ broncopulmonares	bronchopulmonary		████████
✓ broncoscopia	bronchoscopy	Sujeito	██████
✓ broncoscopico	bronchoscopic	Adjetivo	██████
✓ broncoscopista	bronchoscopist	Sujeito	██████
✓ broncovascular	bronchovascular	Adjetivo	██████
✓ broncovasculares	bronchovascular	Adjetivo	██████
✓ bronquica	bronchial	Adjetivo	██████
✓ bronquicas	bronchial	Adjetivo	██████
✓ bronquico	bronchial	Adjetivo	██████

Figura 15: Imagem da tela do gerenciador de dicionários do usuário. Nota-se a coluna com os registros em português à esquerda, seus equivalentes em inglês e a classe gramatical.

Os DU permitem uma melhor análise do texto de entrada ampliando o léxico em uso da língua de origem melhorando conteúdo de saída no idioma de destino. Podem também ser utilizados para cancelar os significados de uma palavra ou expressão de SYSTRAN, configurando uma não-tradução. Outra utilidade é o tratamento unitário de expressões, que são grupos de palavras sem verbo.

Os dicionários de Normalização (DN) podem ser de dois tipos, de acordo com o objeto, o texto de entrada ou texto de saída. Tem como objetivo, como o próprio nome indica, a normalização, ou melhor, a substituição de variantes de palavras em um termo comum conferindo homogeneidade ao texto. A normalização do texto de destino adapta a saída da tradução às necessidades de consistência da terminologia. Oferece uma maneira de substituir as expressões escolhidas pelo programa da tradução com expressões definidas pelo usuário. O terceiro tipo de dicionário são as Memórias da Tradução (MT). Estas consistem de bases de dados de sentenças pré-traduzidas próprias para processar aquelas de ocorrência freqüente nos textos processados.

4.2 Metodologia

O dicionário especializado foi elaborado em diversas etapas seguindo-se um fluxo próprio para o desenvolvimento de dicionários para tradução automatizada⁶⁴. Em princípio foram definidos os objetivos léxicos de acordo com o domínio em estudo. Tendo em vista que o objetivo é a tradução de textos médicos especializados o dicionário deve contar com grande número de substantivos e adjetivos específicos de acordo com os achados apresentados no capítulo anterior. Isto refletiu diretamente nas fontes de termos empregados para sua elaboração.

A primeira fonte escolhida foram os próprios textos de relatórios de radiografias do tórax, cujas palavras inicialmente em Português foram traduzidas para o Inglês manualmente. A segunda fonte foram revistas especializadas em radiologia. Estes termos foram obtidos do trabalho de Chang⁶⁵, que coletou termos dos periódicos ©*Radiology* e ©*Radiographics* com o objetivo de criar um dicionário de correção ortográfica para processadores de texto eletrônicos. Estas palavras foram traduzidas do Inglês para Português. Incluímos também o léxico especializado em radiologia RADLEX²¹ cuja versão preliminar aborda termos da radiologia de tórax em Inglês, que também foram traduzidos para o Português.

Por fim selecionamos palavras de termos em Português do UMLS, que por sua vez foram traduzidas para o Inglês. As palavras então foram incorporadas em uma planilha, classificadas, corrigidas, eliminadas as repetições e inconsistências, constituindo-se o dicionário básico. A seguir, o dicionário foi incorporado ao sistema de TA e traduções nas diversas fases do projeto foram realizadas. Ao final de cada experimento e com base na adequação das traduções, foram feitas as correções e os ajustes no dicionário.

4.3 Resultados

Foram configurados dois dicionários, um UD e outro MT. No UD foram incluídos os termos coletados nas fontes acima descritas. A sua versão mais atual contém um total de 5810 entradas sendo 2927 sujeitos, 2810 adjetivos, 100 advérbios, 57 acrônimos e 46 verbos.

Já a memória de tradução foi configurada para processar as frases relativas aos achados normais algumas delas presentes invariavelmente em todos os laudos. A tabela 13 mostra estas sentenças e as respectivas versões em Inglês utilizadas neste projeto.

Tabela 13: Memória de tradução para sentenças do laudo normal.

Português	Inglês
Aorta normal.	<i>The aorta is normal.</i>
Arco médio normal.	<i>The left atrium auricle silhouette is not enlarged.</i>
CONCLUSÃO:	<i>Impression:</i>
Estruturas ósseas e tecidos moles da parede torácica, sem alterações.	<i>The osseous structures and soft tissues are normal.</i>
Exame realizado em equipamento de alta potência, com técnica de alta kilovoltagem, que fornece escala de contraste longa (baixo contraste).	<i>High kvp low contrast technique radiography.</i>
Hilos de configuração, topografia e dimensões normais.	<i>The pulmonary hila have normal form and dimensions.</i>
Imagem cardíaca normal.	<i>The cardiac silhouette is within normal limits in size.</i>
Mediastino sem alterações.	<i>The mediastinum is normal.</i>
Parênquima pulmonar de transparência normal.	<i>The lungs are clear.</i>
TÉCNICA:	<i>Technique:</i>
Tórax normal.	<i>Normal thorax.</i>
Vascularização pulmonar normal.	<i>Normal pulmonary circulation.</i>
Seios e cúpulas diafragmáticas livres.	<i>The costophrenic sulci are normal. The hemidiaphragms are normal.</i>

4.4 Discussão

Apesar do aplicativo de tradução SYSTRAN possuir um dicionário médico próprio este não foi utilizado. É um dicionário proprietário e não possuímos acesso ao seu conteúdo. Primeiramente não foi possível avaliar o conteúdo e sua adequação ao projeto. Além do mais traduções pré-experimentais mostraram que este era apenas capaz de traduzir termos médicos genéricos e não termos mais específicos, principalmente aqueles flexionados. Este parece mais adequado para traduzir termos médicos presentes em textos gerais não especializados. O desempenho deste dicionário foi abaixo do esperado para este projeto, tendo sido descartado. Desta forma passamos para a elaboração de um dicionário bilíngüe para tradução automática de textos do Português para o Inglês.

A tarefa de traduzir textos médicos em alta qualidade requer um conjunto de termos e expressões relacionados ao domínio, no caso termos relacionados a estruturas anatômicas, patologias e achados de imagem. Um grande número de substantivos e adjetivos é necessário. A avaliação inicial mostra um grande número de erros de tradução (21%) e as altas proporções destas classes gramaticais nos relatórios como demonstrado no capítulo anterior confirmam isto.

A elaboração deste dicionário foi contínua durante o desenvolvimento desta tese. A cada fase de foram realizados ajustes e correções de acordo com os erros de tradução observados. Isto prossegui até a fase experimental quando o dicionário foi finalizado para a realização de experimento de conclusão desta tese.

5 TRADUÇÃO AUTOMATIZADA DE TEXTOS MÉDICOS

Neste capítulo serão apresentados os experimentos conduzidos para avaliação do desempenho da TA de textos médicos. Parte-se da premissa que a expressividade semântica dos textos é mantida com a tradução automatizada do Português para o Inglês. Para que estes textos sejam adequadamente avaliados pelo sistema de PLN, esta tradução deverá ser de boa qualidade com risco de prejuízo global da performance. A avaliação desta qualidade foi realizada através de três experimentos chamados de avaliação inicial, avaliação automática e avaliação humana. Estes experimentos refletem diferentes fases do desenvolvimento do projeto principalmente em relação ao do dicionário especializado empregado. A seguir serão apresentados a metodologia e os resultados enquanto a discussão será abordada em capítulo específico.

Para esta fase do projeto era necessário um sistema de TA disponível, com maior tempo de uso e validação possível, de desempenho reconhecido, e passível de ser adequado a diferentes domínios. Tendo em vista que esta tese não pretende fazer uma avaliação das diferentes metodologias ou do desempenho específicos de algum sistema, qualquer um que atendesse os requisitos apresentados poderia ser empregado. A escolha recaiu pelo aplicativo SYSTRAN (Systransoft, San Diego) na sua versão Premium⁶⁶.

5.1 Avaliação Inicial

Primeiramente procuramos estabelecer um padrão de desempenho de base do aplicativo de TA escolhido. Este estudo avaliou o programa em sua versão original, obtendo um desempenho mínimo esperado, objeto de comparação posterior. Nesta fase não foi empregado o dicionário especializado.

5.1.1 Metodologia

A fonte dos dados foi um subconjunto de 589 relatórios de radiografias de tórax realizadas no Instituto do Coração. O conteúdo do texto foi recuperado sem os dados que permitiram a identificação do paciente. Seu conteúdo foi separado em 507 frases únicas e distintas. Nesta fase foram corrigidos os erros ortográficos. O conjunto de frases foi então submetida a tradução automatizada. Apesar do sistema oferecer um dicionário médico este não foi utilizado. Somente foi utilizado o dicionário geral, ou seja, a opção mínima de vocabulário não possível de ser desabilitada.

As frases originais e traduzidas foram pareadas lado a lado para a avaliação. Foram estudados parâmetros de cobertura e compreensibilidade. A cobertura refere-se à habilidade em lidar com os fenômenos lingüísticos multilingües, isto é a presença de erros de tradução. A proporção de palavras não traduzidas e as erroneamente traduzidas foi a métrica utilizada. Palavras morfológicamente iguais (ex: superior, anterior, infraclavicular) em ambas as línguas que não necessitam de tradução, foram contabilizadas como palavras corretamente traduzidas. A compreensibilidade refere se à qualidade do entendimento do texto. A métrica utilizada foi binária classificando o texto como compreensível e não compreensível. Para esta avaliação foram utilizadas somente as frases que não possuíam palavras não traduzidas, excluindo-se as entre parênteses. Todas as avaliações foram feitas pelo autor.

5.1.2 Resultados

Das 507 frases, a grande maioria delas, 391(77%) apresentava palavras não traduzidas, e foram incluídas no grupo I. A presença de erros de tradução excluía da avaliação quanto a compreensibilidade. Já 116 frases (23%) que não possuíam erros de tradução foram incluídas no grupo II.

Para avaliação dos erros de tradução foram considerados os dois grupos. Para seu cálculo contabilizamos o número de ocorrências de palavras não traduzidas no grupo I, e o número de palavras erroneamente traduzidas no grupo II. Foram respectivamente 772 e 45 ocorrências, ou seja, 817 (21%) do total de 3884 palavras.

Opacidade heterogênea em base direita.

Heterogeneous opacity in right base.

Figura 16: Sentença considerada compreensível.

Linhas finas presentes no lobo médio.
Fine lines gifts in the average wolf.

Figura 17: Exemplo de sentença considerada incompreensível.

As frases do grupo II, ou seja, aquelas sem palavras não traduzidas, foram submetidas à avaliação de compreensibilidade. Das 116 avaliadas 78 (67%) foram consideradas compreensíveis, enquanto 38 (33%) não foram.

5.1.3 Discussão

O primeiro experimento constituiu da avaliação da tradução de sentenças de relatórios de radiografias torácicas com o aplicativo SYSTRAN em sua configuração original sem nenhum dicionário adicional habilitado. Podemos considerar que o desempenho do tradutor automatizado foi satisfatório tendo em vista que 79% das palavras foram corretamente traduzidas e que 67% das frases sem palavras não traduzidas foram consideradas compreensíveis. Excluímos da avaliação de compreensibilidade frases com palavras não traduzidas. Como o objetivo deste experimento era avaliar as possibilidades e não realizar um rigoroso teste de desempenho, consideramos irrelevante avaliar o entendimento de frases que possuíam termos ignorados pelo sistema. Partimos da premissa que o sistema original enfrentaria sérios problemas com a terminologia. O aplicativo estando desprovido de informação gramatical ou de sintaxe de um determinado termo fica impossibilitado de determinar seu papel na frase prejudicando a composição do texto de saída. Além disto a presença de um termo em Português numa frase em Inglês inviabiliza seu entendimento por observadores da língua inglesa, sejam eles humanos ou sistemas de PLN, principalmente quando se tratar de um termo relevante.

A maioria absoluta das palavras não traduzidas foi de termos técnicos da radiologia do tórax. As palavras erroneamente traduzidas geralmente tratavam-se de homônimos como, por exemplo, lobo que foi traduzido a *wolf* ao invés de *lobe*.

Com estes resultados nota-se a possibilidade de traduzir automaticamente relatórios de radiografias de tórax com bom desempenho. Para que isto é necessária a incorporação de um dicionário especializado. Os experimentos seguintes foram realizados com o uso do dicionário de tradução especializado.

5.2 Avaliação Automática

Continuando a avaliação da tradução automatizada passaremos a um estudo específico a respeito da capacidade do dicionário especializado em lidar com termos específicos do domínio médico⁶⁷.

Em um projeto de tradução automatizada a elaboração de um novo dicionário especializado é uma fase com grande custo. Este custo será ainda maior se lançarmos métodos de avaliação manual da tradução automatizada para auxiliar este desenvolvimento. Neste experimento propomos uma metodologia para avaliar o dicionário especializado através de técnicas automatizadas de avaliação da TA de termos médicos. Necessitamos de uma ferramenta de avaliação de baixo custo e elevada reprodutibilidade, em comparação com avaliação humana. Pode se supor que um bom desempenho na tradução de termos médicos de um vocabulário controlado pode mensurar a evolução do dicionário dentro do domínio.

Esta metodologia propõe utilizar o UMLS e sua característica multilingüe intrínseca para se construir um ambiente apropriado para avaliação de tradução de termos médicos. Nesta fase já será utilizado o dicionário especializado.

5.2.1 Metodologia

O experimento inicia-se com a seleção dos conceitos de UMLS que formarão o ambiente de estudo. Foram escolhidos de acordo com os objetivos léxicos e semânticos do domínio em estudo e apenas aqueles que possuam termos em Português. Foram criados 5 grupos com todos os conceitos com termos em Português dos tipos semânticos Acquired Abnormality (AA), Body Location or Region (BLR), Acquired Abnormality (AA), Body Location or Region (BLR), Body Part Organ or Organ Component (BPOC), Body Space or Junction (BSJ) e Tissue (T). O sexto grupo foi formado por termos vinculados a conceitos relacionados ao tórax. Um algoritmo procurou pelos conceitos filhos do conceito tórax (CUI - C0039992), de acordo com os relacionamentos do MESH representados dentro de UMLS.

Este experimento utiliza os algoritmos de avaliação da TA BLEU/NIST e METEOR. Porém como apresentado anteriormente os primeiros fornecem apenas resultados absolutos não adequados em avaliações isoladas. Esta é a característica principal desta metodologia, que avalia o dicionário especializado através da criação de diversos grupos de termos utilizados como teste ou referência, de acordo com as propriedades multilingües do UMLS. Dos conceitos acima selecionados serão extraídos termos vinculados em Português, Inglês e outras línguas. Alguns

serão traduzidos automaticamente e avaliados, enquanto outros serão referências, simulando os limites teóricos de desempenho superior e inferior.

Cada conceito selecionado é um segmento unitário de avaliação e dele obteremos quatro termos distintos. O processo será realizado para os todos conceitos de cada um dos seis grupos. Primeiramente são coletados dois termos distintos em Inglês, um enviado do arquivo REFERENCIA (referência global) e o outro ao arquivo SUPERIOR (limite de desempenho superior). A seguir, é selecionado um termo em língua distinta do Inglês e do Português que será enviado ao arquivo INFERIOR. Finalmente, é selecionado um termo em Português que será submetido a TA duas vezes, com e sem o uso do dicionário especializado sendo armazenado nos arquivos TESTE e ORIGINAL respectivamente. Os arquivos TESTE, SUPERIOR, INFERIOR e ORIGINAL são comparados então ao arquivo REFERENCIA obtendo-se assim quatro escores. Para cada um dos seis grupos de conceitos selecionados é realizado este processo. A representação gráfica dessa metodologia é descrita na figura 18.

O princípio deste experimento é de gerar dois grupos de comparação o SUPERIOR e INFERIOR, dois experimentais TESTE e ORIGINAL, que serão todos avaliados de acordo com o grupo REFERÊNCIA. Todos os grupos contém o mesmo número de termos e conteúdo semântico. Para grupo SUPERIOR por ser formado por termos em Inglês no mesmo idioma do grupo REFERÊNCIA é esperado o melhor desempenho. Para o grupo INFERIOR formado por termos em idiomas diferentes é esperado o menor desempenho. Já os grupos TESTE e ORIGINAL terão seus termos automaticamente traduzidos do Português para o Inglês. Seus escores então refletirão o desempenho da tradução.

A experiência foi realizada com a versão 2004AD do UMLS. As avaliações foram realizadas pela versão 11 do algoritmo BLEU/NIST e pela versão 0.4.3 do algoritmo METEOR. O sistema tradução utilizado é o SYSTRAN Premium 4.0 (Systransoft, San Diego).

5.2.2 Resultados

A quantidade de conceitos em cada um dos grupos esta apresentada na tabela 14.

O algoritmo de busca de conceitos relacionados ao tórax procurou até o quinto nível de relacionamento. Foram selecionados originalmente 3240 conceitos, porém somente 116 deles apresentavam termos em Português, que compuseram o sexto e último grupo.

Recapitulando a estrutura dos resultados, existem quatro documentos comparados com a referência global em Inglês que possuem o mesmo conjunto de conceitos. O grupo INFERIOR é composto por termos em idiomas distintos do Inglês e Português, o ORIGINAL é composto

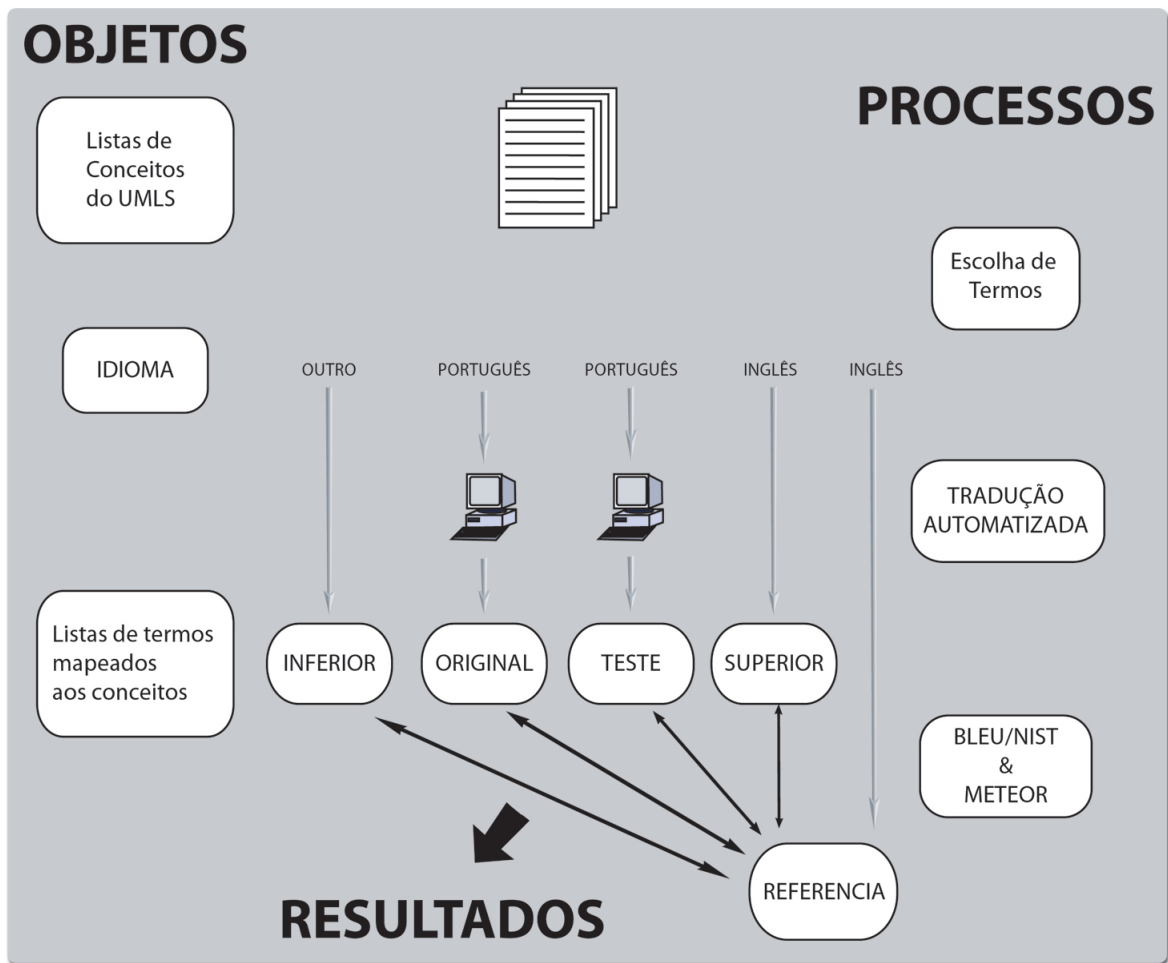


Figura 18: Representação gráfica da metodologia. Cada grupo do conceito de UMLS gera 5 conjuntos de termos usados na experiência. O grupo REFERENCIA será comparado com os outros quatro originais (TESTE, SUPERIOR, INFERIOR e ORIGINAL) por ambos algoritmos.

por termos em Português traduzidos automaticamente sem o dicionário, o TESTE é igual ao anterior, só que traduzidos com o uso do dicionário especializado, e o grupo SUPERIOR formado por termos em Inglês distintos, quando possível, dos termos em Inglês da referencia global. Os resultados da avaliação pelo algoritmo BLEU/NIST estão na tabela 15, enquanto os resultados do METEOR estão na tabela 16, cujo score varia entre 0 e 1.

5.2.3 Discussão

O UMLS oferece uma estrutura onde termos morfologicamente distintos, em línguas diversas e de fontes diferentes, são agrupados sob um único conceito, fazendo possível seu uso como equivalentes da tradução. Isto permitiu a comparação entre os scores dos grupos automaticamente traduzidos, o ORIGINAL e TESTE que representam dois estágios de TA diferentes, antes e após a incorporação do dicionário especializado. São comparados com os scores dos

Tabela 14: Quantidade de termos em Português disponíveis para teste em cada grupo de conceitos selecionados para o experimento.

GRUPO	n
Acquired Abnormality (AA)	675
Body Location or Region (BLR)	57
Body Part, Organ or Organ Component (BPOC)	898
Body Space or Junction (BSJ)	71
Tissue(T)	94
Thorax Child Concepts(TCC)	116

Tabela 15: Escores NIST acumulados de cada documento testado.

GRUPO	SUPERIOR	TESTE	ORIGINAL	INFERIOR
AA	11,56	4,4	2,8	1,14
BLR	6,41	2,08	0,92	0,66
BPOC	11,13	3,5	1,84	1,09
BSJ	7,48	2,24	1,04	0,16
T	8,02	3,06	1,48	0,69
TCC	8,15	3,41	1,35	0,86

Tabela 16: O escore global (S) e a média (Fm) de cada grupo.

GRUPO	SUPERIOR		TESTE		ORIGINAL		INFERIOR	
	S	Fm	S	Fm	S	Fm	S	Fm
AA	0,95	1	0,4	0,56	0,14	0,23	0,1	0,17
BLR	0,99	1	0,5	0,61	0,29	0,38	0,08	0,14
BPOC	0,91	1	0,29	0,48	0,11	0,21	0,11	0,14
BSJ	0,95	1	0,37	0,52	0,17	0,27	0,13	0,19
T	0,93	1	0,42	0,62	0,17	0,27	0,09	0,15
TCC	0,96	1	0,36	0,52	0,12	0,2	0,05	0,07

arquivos SUPERIOR e INFERIOR, representando respectivamente os limites superior e inferior teóricos de desempenho da tradução automática. O grupo SUPERIOR é composto por termos em Inglês que são semanticamente equivalentes e morfologicamente distintos quando possível. O grupo INFERIOR consiste no equivalente dos termos semanticamente corretos, porém morfologicamente não, pois não estão em Inglês. Apesar da incorreção inerente deste grupo, foram obtidos escores acima da nulidade, provavelmente um efeito da similaridade entre palavras das línguas diversas dentro do domínio médico.

Isto ilustra a característica principal deste experimento: a potencialidade de um vocabulário controlado multilingüe em gerar conjuntos de termos para serem traduzidos automaticamente e grupos para servirem de referência. O desempenho dos textos traduzidos sem o uso do dicionário especializado (grupo ORIGINAL) foi melhor que o grupo INFERIOR representando o desempenho mínimo do sistema da TA. O desempenho de grupo TESTE que foi traduzido com uso do dicionário foi maior que o do grupo ORIGINAL, como era esperado. Houve uma melhora significativa dos desempenhos medidos por ambos algoritmos em todos os grupos testados.

As medidas da qualidade da TA envolvem o uso de métricas humanas enquanto os algoritmos do NIST/BLEU medem a similaridade entre documentos, e o METEOR calcula a precisão e a recuperação e medidas derivadas. Em relação aos algoritmos do NIST/BLEU, baixos escores não significam necessariamente traduções de baixa qualidade. Por outro lado escores elevados são mais indicativos de traduções da boa qualidade. Estes escores são absolutos e variam extensamente entre os estudos diversos sendo que são proporcionais ao número das referências usadas^{29, 28, 68}. Já os escores do METEOR variam entre zero e um, pois são proporções.

O UMLS foi usado com sucesso neste experimento como um substrato testado e parâmetro da referência para a avaliação automatizada da TA em termos médicos. Demonstramos de forma automatizada uma melhoria significativa de desempenho após a incorporação do dicionário especializado em um sistema comercial da TA. Esta metodologia além de não empregar observadores humanos que são onerosos e demorados é facilmente reproduzível podendo ser ampliada para outros domínios médicos.

5.3 Avaliação Humana

A última fase da avaliação da TA é a que utiliza o observador humano especialista em oposição ao experimento anterior que emprega algoritmos comparativos. Neste estudo a avaliação manual da TA de sentenças originadas de relatórios de radiografia torácicas é feita através de

parâmetros da cobertura, compreensibilidade e fidelidade. As traduções desta fase foram realizadas com a incorporação do dicionário especializado desenvolvido para o projeto. O desempenho aqui estimado é comparado com os obtidos previamente e descritos neste projeto.

5.3.1 Metodologia

Foram selecionadas 188 sentenças distintas oriundas de um conjunto de 1216 relatórios de radiografias torácicas. Para evidenciar o verdadeiro desempenho não foram feitas correções ortográficas. As sentenças foram então traduzidas pelo aplicativo em uso com a incorporação do dicionário especializado. Os parâmetros de compreensibilidade, fidelidade e da cobertura foram avaliados por um especialista em radiologia não participante do projeto.

A compreensibilidade refere a facilidade com que uma tradução pode ser compreendida, isto é sua clareza ao leitor^{27,69}. A métrica utilizada é binária, classificando o texto compreensível ou não compreensível. Inicialmente o avaliador teve somente acesso à sentença traduzida para avaliar este parâmetro.

A seguir é apresentada a sentença original em Português para avaliação da fidelidade. A fidelidade refere-se à avaliação subjetiva do grau a que a informação contida no texto original foi reproduzida sem distorção na tradução. Esta avaliação foi feita utilizando-se uma escala discreta de 4 pontos^{27,69}. Por este parâmetro a sentença considerada não compreensível pelo parâmetro anterior foi atribuída o grau mais baixo de fidelidade onde nenhuma informação foi passada.

O último parâmetro, este mais objetivo, é a cobertura que se relaciona à tradução correta das palavras. A métrica usada é a porcentagem das palavras traduzidas corretamente em relação ao total das palavras. As palavras não traduzidas, os erros da tradução e os erros morfológicos no texto de origem foram contados como erros no score de cobertura. Palavras morfológicamente similares no Inglês e no Português foram tratadas como traduções corretas.

5.3.2 Resultados

Em relação ao parâmetro de compreensibilidade, das 188 sentenças, 164 (87%) foram considerados compreensíveis enquanto 24 (13%) não, conforme apresentado na tabela 17.

Se forem excluídas as sentenças com erros da tradução a compreensibilidade sobe para 92%.

Em relação à fidelidade, às 24 (13%) sentenças não compreensíveis foram atribuídas o

Tabela 17: Classificação das sentenças de acordo com sua compreensibilidade.

Compreensibilidade	Quantidade	Porcentagem
Sentenças Incompreensíveis	164	87%
Sentenças Compreensíveis	24	13%

Tabela 18: Resultado da compreensibilidade somente nas sentenças sem erros de tradução podendo se notar uma melhora em relação ao grupo total.

Compreensibilidade	Quantidade	Porcentagem
Sentenças Incompreensíveis	13	8%
Sentenças Compreensíveis	141	92%

escore zero (completamente ou quase completamente infiel). A uma sentença (1%) foi atribuído o escore 1 (pouco fiel: menos de 50% da informação original foi passada na tradução), a 8 (4%) foi atribuído escore 2 (razoavelmente fiel: mais de 50% da informação foi passada na tradução) e finalmente a 155 (82%) foi atribuído o escore 3 (completamente ou quase completamente fiel). Estes dados estão sumarizados na tabela 19

Tabela 19: Fidelidade das sentenças automaticamente traduzidas.

Escore	Fidelidade	Quantidade	Porcentagem
0	Infel	24	13%
1	Pouco fiel	1	1%
2	Razoavelmente fiel	8	4%
3	Fiel ou quase completamente	155	82%
Escore Médio		2,56	

Do total de 1308 palavras em 188 frases, 1270 (97%) delas foram corretamente traduzidas enquanto 38 (3%) não foram.

Os erros ortográficos de tradução foram classificados em três tipos. As palavras não traduzidas ocorreram após a tradução automática quando a palavra estava grafada em Português no texto de destino que seria em Inglês. O erro da tradução ocorreu na escolha errada de uma palavra em Inglês no texto de destino. Finalmente o erro de origem foi aquele decorrente de erros de morfologia do Português dos textos de origem. O número de erros está apresentado na tabela 21. Das 188 sentenças, 30 (16%) delas possuíam um erro de tradução enquanto 4 (2%) possuíam 2 erros.

Tabela 20: Cobertura dos textos de destino.

Parâmetro	Quantidade	Porcentagem
Erros	38	3%
Acertos	1270	97% (Cobertura)
Total de Palavras	1308	100%

Tabela 21: Classificação dos erros de tradução.

Tipos de Erros	Quantidade
Palavras não Traduzidas	16
Erros de Tradução	15
Erros de Origem	7

5.3.3 Discussão

O próximo passo da avaliação da TA empregou a observação de um especialista no domínio da radiologia de tórax para avaliar a TA de sentenças de relatórios de radiografias torácicas após incorporação do dicionário pelo aplicativo. Foi observada uma melhora nos resultados de compreensibilidade em comparação com a avaliação do desempenho de base do software da tradução automática⁷⁰. Nessa experiência precedente observamos que 67 % das sentenças sem palavras não traduzidas eram compreensíveis, enquanto que neste experimento observamos 87% e 92%, este último do subgrupo de frases sem erros de tradução.

A análise da fidelidade mostra também resultados favoráveis. A maioria das sentenças foi classificada na categoria superior a respeito da informação passada. A contagem média encontra-se entre 50% e 100% da informação passada de acordo com a escala discreta de avaliação subjetiva.

Foi encontrado um baixo número de erros de origem. Embora pequeno em número, estes parecem degradar mais a compreensibilidade. Das sete sentenças com erros de origem, cinco (71%) foram considerados não compreensíveis. Se considerarmos as sentenças com erros de tradução ou palavras não traduzidas veremos frequências de 28% e 18% respectivamente. Embora o tamanho da amostra seja pequena podemos sugerir o uso de um corretor ortográfico ou um dicionário de normalização para correção dos textos de entrada.

Certamente a maior limitação deste experimento é o fato da avaliação ser feita por apenas um especialista. É sabido que existe uma grande variabilidade na interpretação de diferentes observadores em relação ao mesmo objeto²⁴. Como o emprego de observadores humanos é custoso e demorado, optamos por empregar um número maior deles nos estudos relacionados à

recuperação de documentos médicos. Como o objetivo primário desta tese é a avaliação desta recuperação e não o estudo do desempenho da TA optamos priorizar o uso de especialistas nas fases mais relevantes do projeto.

6 IMPLEMENTAÇÃO DE INFERÊNCIA BASEADA EM ONTOLOGIA

Sistemas de PLN geralmente trazem como resultado dados na forma de tipos semânticos que necessitam de processamento adicional para determinação da presença de achados clínicos. Para esta tarefa implementamos uma abordagem baseada no uso de uma ontologia e de ferramentas de processamento lógico que descrevemos no capítulo 2. A ontologia utilizada neste projeto foi construída a partir de um vocabulário especializado no domínio da radiologia RADLEX²¹. Apresenta um formato hierárquico apropriado para organização na forma de uma ontologia. Esta ontologia será a KB utilizada para a construção de uma série de consultas lógicas implementadas com o uso da Máquina Virtual Algernon (MVA). A ontologia RADLEX será desta forma consultada como KB através do código MVA que contém informações de tipos, que no nosso caso, são obtidos através do PLN com o MEDLEE. Neste capítulo descreveremos a construção da ontologia com o software Protégé e o mecanismo de consulta implementado com a MVA⁷¹.

6.1 Construção da Ontologia

Idealmente seria melhor elaborar uma ontologia com estrutura única para as tarefas de processamento. Porém, as significativas diferenças entre a estrutura de dados e a terminologia do RADLEX e o MEDLEE nos levou a elaborar uma estrutura dupla, recorrendo a regras para promover equivalência e integração entre ambos. Mostraremos então alguns aspectos que julgamos relevantes em cada uma destas estruturas.

6.1.1 Estrutura de entrada

O MEDLEE, como dito, produz um documento XML para cada documento de texto processado. Este ainda não está no formato mais apropriado para a realização das consultas. Propomos, então, uma transformação deste documento através de uma operação XSLT que simplifica a estrutura dos elementos constituintes e também adequa a sintaxe lógica requerida pela MVA. No capítulo 2 mostramos exemplos do documento XML (Figura 10), uma parte do arquivo de transformação XSLT (Figura 12) e o resultado do processamento já na sintaxe da MVA (Figura 13).

Criamos então na nossa KB uma estrutura intermediária elaborada para armazenar temporariamente o resultado do processamento do laudo antes de serem realizadas as consultas clínicas. Criou-se então uma classe *report* que possui uma classe *sentence*. Para cada elemento de primeiro nível do MEDLEE será criada uma instância da classe *sentence* que armazenará todos os atributos relevantes deste elemento em diversos slots. Por elemento de primeiro nível do MEDLEE entende-se aquele que é descendente direto do elemento SECTION. Dentre os diversos tipos de elementos de primeiro nível que o MEDLEE define vamos utilizar *problem*, *normalfinding*, *finding*, *bodyfunction*, *procedure*, e *device*. Vale ressaltar que quando existe a necessidade de explicitar a estrutura anatômica referente a um destes elementos isto é feito através de um elemento filho *bodyloc* ou *region*. Isto não seria necessário, por exemplo, no caso *finding=cardiomegaly*, pois a localização anatômica já está implícita. O slots criados estão descritos na tabela 22.

Tabela 22: Atributos (slots) do tipo (classe) de entrada. EPN - elemento de primeiro nível.

Tipo(classe)	Slot	Descrição
<i>sentence</i>	<i>tipo</i>	Tipo de EPN (<i>finding</i>)
	<i>tipovalue</i>	Valor deste tipo (<i>radiopacity</i>)
	<i>bodyloc</i>	Estrutura anatômica do EPN (<i>lungs</i>)
	<i>certainty</i>	Grau de certeza do EPN (<i>high certainty</i>)
	<i>descriptor</i>	Qualificador geral do EPN (<i>metallic</i>)
	<i>region</i>	Qualificador de região do EPN ou <i>bodyloc(left)</i>
	<i>suid</i>	Sentença do EPN(<i>sI</i>)
	<i>value</i>	Estado de afirmação (:TRUE) ou negação (:FALSE) do EPN

Então resumidamente para cada relatório de radiografia torácica o processamento do MEDLEE fornecerá um conjunto de elementos de primeiro nível que criarão instâncias de *sentence* tantos quantos forem estes elementos.

6.1.2 Estrutura do RADLEX

Como já apresentado o vocabulário RADLEX tem como objetivo oferecer uma estrutura uniforme de dados para operações de informação em radiologia. Uma característica importante é que este apresenta um formato hierárquico apropriado para incorporação em forma de ontologia fundamentado em 12 superclasses distintas chamadas categorias. Por exemplo, a árvore correspondente a estrutura anatômica *lungs* (pulmões) está parcialmente representada na figura 19. Inicialmente utilizaremos somente os tipos relativos aos achados (*findings*) e estruturas anatômicas (*anatomic location*).

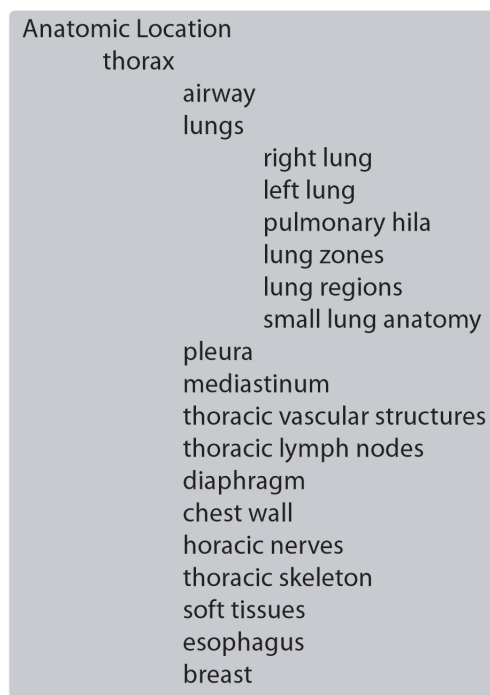


Figura 19: Exemplo de elemento do RADLEX. Estrutura parcial da árvore de estruturas anatômicas.

Cada unidade do RADLEX foi incorporada à ontologia como uma classe. Partimos da premissa que todos os conceitos significativos ao domínio em estudo estão contemplados nesta ontologia, isto é uma hipótese de mundo fechado onde não são admitidas situações, conceitos e estados não descritos na KB. A estrutura hierárquica original do RADLEX foi preservada. Foram criados atributos alguns comuns e outros restritos a subgrupos de tipos específicos. A tabela 23 mostra um resumo geral dos atributos das superclasses *anatomic location*, *morphologic and physiologic processes*, *diagnoses and etiologies*, e *visual features*.

O frame *"anatomic location"*, que está parcialmente apresentado na figura 19, representa todo o conjunto e estruturas anatômicas consideradas para processamento.

Tabela 23: Atributos (slots) das superclasses da estrutura ontológica do RADLEX.

Tipo(classe)	Slot	Descrição
<i>anatomic location, diagnoses and etiologies, visual features, morphologic and physiologic processes</i>	<i>instancia</i>	Identificacao do elemento do texto a qual pertence(geralmente sentença)
	<i>value</i>	Estado de afirmação (:TRUE) ou negação (:FALSE)
<i>anatomic location</i>	<i>normal</i>	Estado de normalidade da instância anatômica
<i>morphologic and physiologic processes, diagnoses and etiologies</i>	<i>affect</i>	Restrito a uma classe de <i>anatomic location</i> . Relaciona-se com a estrutura afetado por estes tipos
<i>visual features</i>	<i>imageof</i>	Restrito a tipos das classes de <i>morphologic and physiologic processes</i> e <i>diagnoses and etiologies</i> ou de <i>anatomic location</i> . Relaciona-se com seu aspecto visual.

6.2 Raciocínio Lógico

A MVA transforma o código de programação lógica de primeira ordem em linguagem de máquina própria para processamento computacional. Pode acessar bases de conhecimento na forma de frames como as ontologias descritas pela ferramenta *Protégé*. Vamos descrever então a utilização da MVA mostrando suas operações fundamentais envolvidas nesta metodologia.

As definições, as afirmações, as regras, inclusive as clínicas e as consultas são incluídas num arquivo texto utilizando a sintaxe própria da MVA. Este arquivo será então processado e sua saída também se dá na forma de um arquivo texto. A unidade de processamento é o conteúdo de um relatório de radiografia torácica.

A ordem de comandos enviados é primeiramente carregar as regras, em seguida adicionar as instancias a ontologia e por fim executar as consultas. O arquivo de saída também está na forma de arquivo texto e contém respostas das consultas que obtiverem resposta da ontologia. Para uma melhor clareza apresentaremos primeiro a criação e instancias, e as seguir as regras e depois as consultas.

6.2.1 Declaração de Instâncias

A MVA apresenta dois tipos diferentes de comandos básicos de interação com a KB. A primeira é a *"tell"* que representa operações de informação à base de dados de regras e de

adicionar ou editar instâncias. Digamos que o resultado de um determinado processamento de texto traga o conceito pneumonia como verdadeiro em uma determinada frase do relatório investigado. A assertiva seguinte instancia a classe correspondente à patologia em questão, da sentença 1 com valor verdadeiro.

```
((:add-instance (?x pneumonia) (instancia ?x "1") (value ?x :TRUE)))
```

Figura 20: O comando add-instance adiciona uma instância à classe pneumonia.

Note que as variáveis são representadas em Algernon com nomes que começam com ?. A função "*add-instance (?x pneumonia)*" cria uma instância na classe pneumonia. O trajeto a seguir significa que a instância criada pertence à sentença 1 e possui valor verdadeiro, ambos ligados pela variável ?x. Isto representa um trajeto de acesso na sintaxe da MVA.

6.2.2 Declaração de Regras

Podemos representar implicações lógicas na forma de regras, também através da operação "*tell*". Vejamos o exemplo seguinte na figura 21.

```
((:ADD-RULE      "spine disease"
  ((value ?x :TRUE)
  (instancia ?x ?suid)      -> (:add-instance (?y "thoracic spine")
                                (instancia ?y ?suid )
                                (normal ?y :FALSE)
                                (value ?y :TRUE)
  )))
```

Figura 21: Exemplo de uma regra de encadeamento anterógrado.

Por exemplo, a instrução cima cria uma instância de coluna torácica não normal quando uma instância de *spine disease* (doença torácica) é criada. Este tipo de regra é chamada de regra de encadeamento anterógrado (REA) e é disparada quando instâncias são acrescentadas a KB e neste caso levam a execução das instruções à direita da seta. A MVA também permite elaborar regras de encadeamento retrógrado (RER) como o exemplo da figura 22.

Ao pé da letra, esta regra significa que para existência de uma pneumonia descrita em uma determinada unidade de texto é necessária a existência do conceito pulmões (lungs) e opacidade (radiopaque) na mesma frase. Ou, sob outro ponto de vista, se existirem os conceitos pulmão e opacidade na mesma frase ocorrerá pneumonia. Apesar desta regra não ser inteiramente adequada do ponto de vista clínico, pode ilustrar o funcionamento deste tipo de regra.

```

((:add-rule pneumonia
((instancia ?p ?x)      <-  (:INSTANCE lungs ?lungs)
                             (instancia ?lungs ?x)
                             (value ?lungs :TRUE)
                             (:NEQ ?lungs ?radiopaque)
                             (:INSTANCE radiopaque ?radiopaque)
                             (instancia ?radiopaque ?x)
                             (value ?radiopaque :TRUE)
)))

```

Figura 22: Exemplo de uma regra de encadeamento retrógrado.

Não existem regras específicas para quando se utilizar REA ou RER. Porém com o uso heurísticas foram desenvolvidas na semântica destas regras. As REA deduzem propriedades que são consideradas invariantes da KB. Se o fato é verdadeiro ou falso isto estará explícito na KB. As RER são utilizadas na resolução de problemas na maioria dos tipos de inferência .

6.2.3 Consultas

A outra operação básica em uma KB do MVA é a *"ask"*. Através dela podemos recuperar informações da KB. Existem duas maneiras de fazê-la, a primeira aplicando RER. Poderíamos criar uma série de regras de encadeamento retrógrado. A partir de diagnósticos possíveis poderia se investigar a presença ou não de conceitos antecedentes que poderiam determinar sua presença ou não.

Porém, como nosso sistema é desenhado para o processamento de dados pareceu mais apropriado implementar uma solução desenvolvida com REA. O processamento de texto fornece conjuntos de conceitos diversos que serão adicionados como instâncias da KB. A partir disto regras são disparadas povoando diversas classes da ontologia especializada. REA também são chamadas de mecanismos de inferência movidos a dados, enquanto que o modelo fundamentado em RER é chamado de movido a objetivos.

A operação *"ask"* então é utilizada para consultar a KB para verificar a presença ou não de instâncias de classes pré-definidas então respondendo perguntas a clínicas selecionadas. Por exemplo, podemos consultar a presença de uma instância de coluna torácica com o slot *normal* com valor *":FALSE"* com o trajeto mostrado na figura 28. Em caso de sucesso a expressão de texto representando a não normalidade da coluna torácica é adicionada ao arquivo de saída.

6.2.4 Funcionamento da Aplicação

Então tomemos o relatório de radiografia torácica apresentado na figura 9. Realizadas as fases de tradução, PLN e transformação obteremos as seguintes expressões:

```
(tell ((:add-instance (?x sentence) (tipo ?x "problem") (tipo_value ?x "osteophytosis")
(suid ?x "s1.1.1") (value2 ?x :TRUE))))
(tell ((:add-instance (?x sentence) (tipo ?x "problem") (tipo_value ?x "enlarged")
(bodyloc ?x "auricle") (suid ?x "s1.1.6") (certainty ?x "no") (value2 ?x :FALSE))))
(tell ((:add-instance (?x sentence) (tipo ?x "finding") (tipo_value ?x "free")
(suid ?x "s1.1.2") (region ?x "recess") (value2 ?x :TRUE))))
(tell ((:add-instance (?x sentence) (tipo ?x "finding") (tipo_value ?x "elongated")
(bodyloc ?x "aorta") (suid ?x "s1.1.5") (value2 ?x :TRUE))))
(tell ((:add-instance (?x sentence) (tipo ?x "normalfinding") (tipo_value ?x "normal")
(descriptor ?x "form") (bodyloc ?x "pulmonary hila") (suid ?x "s1.1.4")
(certainty ?x "high certainty") (value2 ?x :TRUE))))
(tell ((:add-instance (?x sentence) (tipo ?x "normalfinding") (tipo_value ?x "clear lungs")
(bodyloc ?x "lung") (suid ?x "s1.1.7") (certainty ?x "high certainty") (value2 ?x :TRUE))))
(tell ((:add-instance (?x sentence) (tipo ?x "normalfinding") (tipo_value ?x "normal")
(bodyloc ?x "mediastinum") (suid ?x "s1.1.8")
(certainty ?x "high certainty") (value2 ?x :TRUE))))
(tell ((:add-instance (?x sentence) (tipo ?x "bodyfunc") (tipo_value ?x "circulation")
(bodyloc ?x "lung") (suid ?x "s1.1.3") (problemdescr ?x "normal") (value2 ?x :TRUE))))
```

Figura 23: Exemplo de um relatório completo na forma de assertivas lógicas.

Estas assertivas irão instanciar a classe sentence com diversas instâncias. Esta classe foi criada em nossa KB para receber os dados diretamente do processador de linguagem natural da forma estruturada com cada um dos elementos referindo-se a apenas uma sentença. Previamente a isto ocorre o carregamento das regras. Foram elaborados vários tipos de regras. A primeira são as regras de equivalência de termos entre o MEDLEE e o RADLEX. Tendo em vista que estamos utilizando dois vocabulários distintos tivemos de lançar mão de regras para os termos morfologicamente diferentes. O exemplo da figura 24 mostra a equivalência do termo pulmão entre ambos. Note que o resultado do MEDLEE é no singular enquanto o RADLEX o utiliza no plural.

```
((:ADD-RULE sentence((bodyloc ?x "lung") -> (bodyloc ?x "lungs"))))
```

Figura 24: Equivalência de termos simples do MEDLEE para o RADLEX.

Outro tipo de regra criada também foi de equivalência de termos entre os vocabulários, porém neste caso para o termo do MEDLEE haveria o instanciamento de dois termos do RADLEX como mostrado na figura 25.

O termo *calcified aorta* do MEDLEE leva a criação de instâncias de aorta e calcificação, ambos relacionados pelo valor do slot suid que se refere à sentença que contém este achado.

```
((:ADD-RULE sentence((tipo ?x "problem")(tipo_value ?x "calcified aorta")
  (suid ?x ?suid)(value2 ?x :TRUE) ->
  (:add-instance (?y "aorta")(instancia ?y ?suid)
    (value ?y :TRUE)(normal ?y :FALSE))))))
(:ADD-RULE sentence((tipo ?x "problem")( tipo_value ?x "calcified aorta")
  (suid ?x ?suid)(value2 ?x :TRUE) ->
  (:add-instance (?y "calcification")
    (instancia ?y ?suid)(value ?y :TRUE))))))
```

Figura 25: Equivalência de termos simples do MEDLEE para compostos do RADLEX.

Poderíamos também criar regras onde múltiplos termos de MEDLEE instanciaríamos apenas um tipo do RADLEX, isto através de trajetos de acesso. Ainda não houve a necessidade até o momento.

No caso dos termos que são morfologicamente iguais em ambos vocabulários, o instanciamento do RADLEX é feito através de regras genéricas como exposto na figura 26. Os achados ("*findings*") com exceção daqueles com valor "*free*" no MEDLEE são instanciados nos tipos do RADLEX.

```
((:ADD-RULE sentence ((tipo_value ?x ?nome ) (suid ?x ?suid)
(tipo ?x "finding")(value2 ?x :TRUE)(:NEQ ?y "free") ->
(:add-instance (?y ?nome) (instancia ?y ?suid )(value ?y :TRUE))))))
```

Figura 26: Regra de encadeamento anterógrado para instanciamento genérico de achados.

Regras semelhantes foram criadas para outros tipos de elementos do MEDLEE. No caso do termo "*free*" este será processado juntamente com um tipo de regras, aquelas relacionadas com os achados da normalidade. Um exemplo é esta regra geral de instanciamento de estruturas anatômicas normais:

```
((:ADD-RULE sentence((bodyloc ?x ?nome ) (suid ?x ?suid)
(tipo ?x "normalfinding")(value2 ?x :TRUE) ->
(:add-instance (?y ?nome) (instancia ?y ?suid )
(normal ?y :TRUE)(value ?y :TRUE))))))
```

Figura 27: Regra de instanciamento de estrutura anatômica normal.

Também foram criadas regras de cunho clínico que relacionam classes anatômicas, patológicas e achados de imagem. Um exemplo é a regra da patologia vertebral apresentada na subseção acima, na qual sua presença leva a criação de uma instância não normal da estrutura anatômica representada que no domínio em questão é a coluna torácica.

Por fim são realizadas as consultas. Veja a consulta apresentada na figura 28.

```
((:INSTANCE "thoracic spine" ?z)(normal ?z :FALSE)
(:PRINTLN "Coluna Torácica não normal"))
```

Figura 28: Trajeto de consulta a base de dados à procura de instância não normal de coluna torácica.

Se esta consulta for aplicada após o carregamento do relatório de exemplo, iríamos obter no arquivo de saída a expressão que indica o sucesso da pergunta. A partir deste ponto a determinação da presença e algum achado clínico se resume a procura de expressões específicas conseqüentes às consultas realizadas diretamente no arquivo de saída do relatório analisado.

6.3 Discussão

Poderíamos perguntar porque precisamos de um mecanismo de inferência com o resultado do PLN, visto que são informações codificadas? Não seria suficiente para a determinação dos achados clínicos investigados? A resposta é não. Existem uma série de outros fatores além da busca de achados por semelhança de texto. Estados podem estar negados o que é captado pelo nosso sistema. O que necessita implementação são os estados de certeza intermediária situados entre a negação e a certeza alta. Atualmente operamos com uma escala binária (:TRUE e :FALSE). Os complexos relacionamentos que existem entre os diversos tipos existentes também é fator de complexidade que necessita de processamento adicional.

Em um trabalho⁵ que estuda o resultado de saída do processador SIMTEXT em 150 relatórios de radiografias torácicas foi analisado o desempenho de cinco métodos de inferência na determinação de pneumonia comparado com três observadores humanos (dois radiologistas e um clínico). Os métodos foram regras elaboradas por especialistas, dois tipos de redes bayesianas, uma árvore de decisão e busca por palavras. Com exceção desta última, com pior desempenho, não foram observadas diferenças com a referência humana em decidir pela presença de pneumonia nos textos.

Especificamente para uso com o MEDLEE foi implementado um sistema de inferência baseada em regras para processamento dos resultados de PLN. Esta abordagem utiliza a *Arden Syntax for Medical Modules* que aceita como entrada os achados codificados em tipos e usa um conjunto de regras para determinar a presença ou não de achados clínicos selecionados. Por exemplo, a instrução da figura 29 indica conclusão verdadeira se atelectasia estiver presente.

```

If finding is in ("atelectasis", "discoid atelectasis", "linear atelectasis",
                 "platelike atelectasis", "pressure atelectasis",
                 "segmental atelectasis", "subsegmental atelectasis")
then conclude true

```

Figura 29: Instrução na forma de regras em sintaxe Arden para consulta à atelectasia

Nossa metodologia utiliza uma abordagem conceitualmente diferente. No caso da presença de algum dos tipos de atelectasias descritos, sua classe correspondente seria instanciada na ontologia de acordo com o resultado do processamento. Ao invés de executar consulta a cada um destes tipos é feita uma consulta a super-classe atelectasia como abaixo:

```

((:INSTANCE "atelectasis" ?z) (value ?z :TRUE)
 (:PRINTLN "Atelectasis"))

```

Figura 30: Consulta à presença e atelectasia em sintaxe da MVA.

Consultas são executadas a partir das assertivas oferecidas, das regras preconcebidas e da ontologia como substrato incluindo aí seus tipos e relações. A pergunta carregada leva uma resposta binária da presença de algum elemento instanciado, no caso a interpretação do achado de acordo com sua representação textual do relatório. Este modelo do raciocínio é capaz de verificar a presença ou não das situações clínicas investigadas baseadas em assertivas fornecidas.

Já foi descrito um modelo de processamento utilizando ontologias construídas no Protégé para interpretar achados de PLN com o MEDLEE⁷². O resultado do processamento de PLN em relatórios de alta hospitalar foi transformado através de uma ferramenta XSLT que tinha acesso a uma ontologia descrita com o Protégé. Neste caso, a ontologia que representa os conceitos e as relações intrínsecos ao próprio processador MEDLEE. Diferentemente nosso modelo utiliza uma ontologia construída a partir de um vocabulário aberto não proprietário. Este modelo foi implementado para em sumarização de relatórios, e não para recuperação de documentos.

Um outro aspecto que deve ser ressaltado é a possibilidade de se implementar relações entre os tipos patológicos, anatômicos e de imagem. Podemos por exemplo preencher os *slots* com atributos que determinam as relações. Uma outra maneira já implementada é através de regras. Na figura 31 mostramos a regra que instancia o tipo *lung* em resposta a instância *pneumonia*.

```
(tell ( (:ADD-RULE "pneumonia" ((value ?x :TRUE)
(instancia ?x ?suid) -> (:add-instance (?y "lung")
                        (instancia ?y ?suid ) (normal ?y :FALSE)
                        (value ?y :TRUE))
))))
```

Figura 31: Exemplo de uma regra de encadeamento anterógrado que estabelece uma relação entre uma estrutura anatômica e uma patologia.

7 RESULTADOS DA EXTRAÇÃO E RECUPERAÇÃO DE INFORMAÇÕES

Dando continuidade à metodologia, mostraremos neste capítulo sua aplicação em dois experimentos. No primeiro será apresentado um experimento de extração de informações clínicas de frases de relatórios de radiografias de tórax. O segundo experimento é o mais importante do projeto de pesquisa. Nele todas as fases concebidas serão executadas num conjunto de relatórios. Serão traduzidas, processadas, transformadas e calculadas objetivando a identificação e quantificação de vinte e dois estados clínicos distintos destes relatórios.

7.1 Extração de Informações de Sentenças Médicas

Esta fase do experimento foi a primeira avaliação formal, ainda que muito simplificada, do acoplamento da tradução automatizada com o processador MEDLEE. Além de avaliar o desempenho de extração de conceitos de sentenças de relatórios de radiografias torácicas este experimento tem como objetivo estudar caso a caso o desempenho do processamento do MEDLEE apontando as falhas e corrigindo os erros, principalmente os decorrentes da TA ou de uso de termos específicos. Dado que a natureza deste experimento é dar suporte ao desenvolvimento ela foi realizada pelo autor da tese.

7.1.1 Metodologia

O grupo de 188 frases distintas utilizadas na experiência anterior de avaliação humana da TA foram submetidas ao processamento do MEDLEE em sua interface web de demonstração⁷³. O resultado estruturado foi comparado com as frases em Português e em Inglês. Os resultados foram então classificados utilizando-se uma métrica de quatro classes ordinais de acordo com o processamento semântico correto. Poderíamos fazer uma analogia com a escala de fidelidade

utilizada para avaliação da tradução automatizada. O escore zero (0) foi atribuído quando o processamento não apresentou resultado. O escore um (1) foi atribuído quando o resultado apresentava menos de 50% da informação passada ao resultado. O escore dois (2) foi associado com mais de 50% da informação passada. Por fim o escore três (3) referiu-se a quando a informação foi completamente ou quase completamente passada para o resultado. Na prática foi dado um peso maior à correta identificação anatômica a qual a sentença se refere.

Na segunda fase, as sentenças que não obtiveram o escore máximo foram avaliadas uma a uma à procura dos erros que poderiam afetar o processamento da linguagem textual. Alguns dos erros foram corrigidos e as sentenças novamente processadas pelo MEDLEE, e avaliadas. Esta fase foi concebida para se avaliar os efeitos no desempenho do processamento do MEDLEE após modificações em alguns aspectos da TA. Estas modificações foram incorporadas ao dicionário de tradução e terão efeito na próxima fase do projeto de pesquisa.

7.1.2 Resultados

Das 188 sentenças inicialmente avaliadas 103 (55%) obtiveram o escore 3, 27 (15 %) o escore 2, 21 (11%) o escore 1, e 37 (19%) o escore 0. Um exemplo de sentença que obteve o escore 3 está mostrado na figura 32 abaixo.

```

Calcificação do ligamento longitudinal anterior.
calcification of the anterior longitudinal ligament.
problem:calcification
  bodyloc>> ligament
    idref>> 352
    position>> longitudinal
      idref>> 350
      region>> anterior
        idref>> 348
    code>> UMLS:C0206329^anterior longitudinal ligament structure
      idref>> [348,350,352]
    code>> UMLS:C1280217^entire anterior longitudinal ligament
      idref>> [348,350,352]
  idref>> 342
  parsemode>> model
  sectname>> report unknown section item
  sid>> [1,19,1]
  code>> UMLS:C0343212^calcification of ligament
    idref>> [342,352]

```

Figura 32: Exemplo de sentença que obteve escore máximo.

A figura 33 mostra exemplos de sentenças que obtiveram escore 0.

Na segunda rodada de avaliação foram processadas 85 sentenças. Destas 50 delas (59%)

Aorta à direita.

aorta on the right side.

Figura 33: Exemplo de sentença que não obteve resultado de processamento.

obtiveram o escore 3, 14 (16%) o escore 2, 10 (12%) o escore 1, e 11 (13%) o escore 0⁷⁴. Se acumulássemos os escores de ambas as rodadas veríamos que 153 das 188 sentenças obtiveram o escore máximo de conteúdo semântico extraído pelo processador e 14 delas um conteúdo maior que 50%.

7.2 Recuperação de Informações de Textos Médicos

A integração dos processos de TA de PLN e de Raciocínio Lógico de forma serial e nesta ordem irá compor a parte principal da metodologia. Desta forma, iremos recuperar documentos clínicos de acordo com suas informações textuais extraídas das frases. Relatórios de radiografias torácicas serão, então, investigados desta forma à procura de vinte e duas condições clínicas comumente presentes em RRT. Os resultados deste processamento respostas binárias indicando ou não sua presença, serão comparadas com a opinião de dois observadores especialistas (médicos radiologistas).

7.2.1 Metodologia

Foram selecionados RRT realizadas em nossa instituição e selecionados em dois grupos. No primeiro grupo denominado de grupo de treinamento foram incluídos 1362 RRT realizados em 2003 e 2004. Estes foram utilizados nos experimentos anteriormente realizados e também como treinamento do experimento final. Toda metodologia foi desenvolvida e ajustada utilizando este grupo de RRT. No grupo denominado experimental foram incluídos 12.869 relatórios realizados entre janeiro de 2005 e julho de 2006. Em nenhum relatório foram obtidos quaisquer dados que permitissem a identificação do paciente ou alguma informação clínica. Não houve nenhum tipo de restrição em relação às incidências utilizadas, sendo todas incluídas. Também não houve distinção se eram radiografias realizadas em equipamento portátil no leito. A única restrição era a presença de laudo emitido pelo médico radiologista. Também não houve identificação do profissional que havia elaborado o relatório.

A versão do dicionário de tradução utilizada neste experimento foi a obtida com as modi-

ficações sugeridas pelos experimentos anteriores. Toda a metodologia foi realizada no grupo de teste inicialmente. Ajustes foram realizados nos processos de tradução automatizada, transformação do resultado do MEDLEE, estrutura da ontologia e no mecanismo de inferência. A partir deste ponto o sistema foi fechado para a realização dos testes com o grupo experimental.

A cada relatório foi designado um identificador único numérico e seu conteúdo foi colocado em arquivo de texto. O laudo a seguir foi submetido à TA utilizando o aplicativo SYSTRAN configurado com o dicionário especializado. O laudo traduzido em inglês foi então submetido ao processamento pelo MEDLEE e o resultado do processamento armazenado num arquivo XML. Na seção de representação de informação apresentamos um exemplo de relatório submetido a toda metodologia.

O resultado do MEDLEE como apresentado no capítulo 2 é um documento estruturado cuja definição está explicitada no seu documento dtd¹ que apresenta uma estrutura arboriforme em vários níveis para descrever os diferentes aspectos dos textos médicos. Esta estrutura atualmente em uso é bem mais complexa que a publicada originalmente⁷⁵. Para nosso processamento, porém, necessitamos de uma estrutura menos complexa que esta contida na definição atual. Para isso, lançamos mão de uma fase de transformação de elementos do MEDLEE em assertivas lógicas com estrutura simplificada. Basicamente, a estrutura multinível original do documento MEDLEE é achatada para uma estrutura de um nível hierárquico como mostrado na figura 23.

Esta simplificação da estrutura do documento XML do MEDLEE permite a construção de assertivas lógicas diretamente dos resultados do PLN. Como já apresentado foi criada uma estrutura dupla na ontologia. A estrutura de entrada cuja classe principal *sentences* é instanciada para cada elemento de primeiro nível do MEDLEE relatório. Os slots da classe *sentences* estão ajustados à simplificação estrutural realizada pela transformação.

Com a declaração destas instâncias REA criam novas instâncias secundárias na estrutura do RADLEX. Estas povoarão as diversas classes de achados, patologias e estruturas anatômicas permitindo consultas de caráter clínico. Em resumo, o aplicativo consiste em obter de cada relatório uma lista de assertivas que vão instanciar os tipos da estrutura ontológica de sentenças (entrada). Estas por vez vão disparar regras que criam novas instâncias na ontologia RADLEX que serão objeto de consultas clínicas.

As consultas são estabelecidas de acordo com os objetivos clínicos procurados. Constituem-se como uma busca de ocorrências de instâncias positivas mapeadas dentro da ontologia RADLEX. O processamento se dá a partir da consolidação das regras globais e perguntas aos achados dos relatórios transformados em assertivas. Este conjunto de instruções será processado

¹Documento de definição do XML (<http://persival.cs.columbia.edu/medlee/medlee.dtd>)

pela MVA objetivando uma resposta binária (sim ou não) da presença dos achados.

Neste experimento o objetivo é a recuperação de RRT que apresentem vinte e dois achados radiológicos comumente presentes neste tipo de texto descritos na tabela 24.

Tabela 24: Achados clínico radiológicos investigados.

Achado	Descrição
Alargamento de mediastino	Alargamento ou aumento do mediastino
Aneurisma, dilatação ou alteração de aorta	Alterações no calibre, tortuosidade, ectasia ou alongamento da aorta.
Atelectasia	Descrição ou sugestão de atelectasia, opacidade e desvio do mediastino
Ateromatose	Calcificação arterial principalmente em aorta
Cardiomegalia ou aumento de câmara cardíaca	Aumento do coração, cardiomegalia ou de câmara cardíaca. O arco médio refere-se à silhueta da aurícula do atrio esquerdo e pode estar alterada.
Cifose	Descrição de cifose torácica sugerindo acentuação da mesma
Derrame, espessamento pleural ou alteração de seio cardio/costo-frênico	Derrame pleural, opacificação ou obliteração de seio cardio e costo-frênicos. Espessamento pleural
Dispositivo, aparato ou corpo estranho	Marca-passo, clip, sutura, sonda, material de osteossíntese ou qualquer outro tipo de elemento estranho ao corpo.
Doença degenerativa vertebral	Osteofitose, osteoartrose, sindesmófitos, espondilose ou espondilolistese
DPOC	Hiperinsuflação bilateral, bolhas, aumento do diâmetro ântero-posterior, depressão do diafragma, diminuição da circulação pulmonar
Edema ou congestão pulmonar	Aumento hilar, aumento da circulação pulmonar, opacidades peri-hilares ou basais bilaterais ou linha de Kerley
Elevação ou alteração do diafragma	Elevação, lobulação ou paralisia de um ou de ambos diafragmas
Escoliose	Descrição de escoliose
Fratura costal, vertebral ou óssea	Qualquer fratura óssea, incluindo achatamento ou encunhamento de corpo vertebral
Hipertransparência ou hiperinsuflação pulmonar	Aumento da transparência pulmonar ou descrição de hiperinsuflação.
Laudo normal	Relatório sem nenhum tipo de anormalidade. Qualquer achado exclui esta possibilidade
Massas ou nódulos	Nódulos, opacidades nodulares ou massas.
Opacidade pulmonar	Qualquer tipo de opacidade, seja alveolar, intersticial, linear.
Pneumonia	Descrição sugestiva de pneumonia, seja lobar, broncopneumonia, intersticial. Opacidade unilateral pode ser considerada. Caso o relatório sugira ou interrogue considerar como positivo.
Pneumotórax	Descrição de pneumotórax
Projéteis ou fragmentos	Fragmentos ou projéteis de arma de fogo
Tuberculose	Tuberculose sugerida ou interrogada, opacidades apicais, cavitação

Os laudos normais podem ser definidos como aqueles cujas frases relatem apenas achados radiológicos compatíveis com a normalidade. A esta tarefa de definir a normalidade é relegada ao MEDLEE. O MEDLEE tem a capacidade de identificar os achados dentro da normalidade possibilitando a resolução das consultas pelo modelo apresentado.

Os outros achados de imagem procurados são aqueles relativos a implantes, cateteres, equipamentos e outros implementos de natureza cirúrgica, como suturas ou cliques. Tais elementos serão chamados genericamente de dispositivos.

Do grupo experimental 148 relatórios foram selecionados aleatoriamente para serem comparados com a referência. Esta foi elaborada por observadores humanos, no caso três médicos radiologistas que responderam às mesmas perguntas. No caso de mais de um observador indicar a presença do achado este foi considerado positivo. Para esta avaliação foi utilizada uma interface em página de Internet mostrada na figura 34.

O desempenho global do sistema será avaliado por esta amostra. A variabilidade da referên-

RELATÓRIO - 7

tecnic:
exame realizado em equipamento de alta potencia, com tecnica de alta kilovoltagem, que fornece escala de contraste longa (baixo contraste).
laudo:
estruturas osseas e tecidos moles da parede toracica, sem alteracoes.
seios e cupulas diafragmaticas livres.
vascularizacao pulmonar normal.
hilos de configuracao, topografia e dimensoes normais.
imagem cardiaca normal.
discreta alongamento da aorta.
parenquima pulmonar de transparencia normal.
ac.

Por favor indique a presença dos seguintes estados clínico - radiológicos:

Laudo normal	<input checked="" type="radio"/> SIM	<input type="radio"/> NÃO	Tuberculose	<input type="radio"/> SIM	<input checked="" type="radio"/> NÃO
Dispositivo, aparato ou corpo estranho	<input type="radio"/> SIM	<input checked="" type="radio"/> NÃO	Cardiomegalia ou aumento de câmara cardíaca	<input type="radio"/> SIM	<input checked="" type="radio"/> NÃO
Pneumonia	<input type="radio"/> SIM	<input checked="" type="radio"/> NÃO	Ateromatose	<input type="radio"/> SIM	<input checked="" type="radio"/> NÃO
Opacidade pulmonar	<input type="radio"/> SIM	<input checked="" type="radio"/> NÃO	Aneurisma, dilatação ou alteração de aorta	<input type="radio"/> SIM	<input checked="" type="radio"/> NÃO
Atelectasia	<input type="radio"/> SIM	<input checked="" type="radio"/> NÃO	Fratura costal, vertebral ou óssea	<input type="radio"/> SIM	<input checked="" type="radio"/> NÃO
Hipertransparência ou hiperinsuflação pulmonar	<input type="radio"/> SIM	<input checked="" type="radio"/> NÃO	Doença degenerativa vertebral	<input type="radio"/> SIM	<input checked="" type="radio"/> NÃO
DPOC	<input type="radio"/> SIM	<input checked="" type="radio"/> NÃO	Escoliose	<input type="radio"/> SIM	<input checked="" type="radio"/> NÃO
Derrame, espessamento pleural ou alteração de seio cardio/costo-frênico	<input type="radio"/> SIM	<input checked="" type="radio"/> NÃO	Cifose	<input type="radio"/> SIM	<input checked="" type="radio"/> NÃO
Pneumotorax	<input type="radio"/> SIM	<input checked="" type="radio"/> NÃO	Projéteis ou fragmentos	<input type="radio"/> SIM	<input checked="" type="radio"/> NÃO
Massas ou nódulos	<input type="radio"/> SIM	<input checked="" type="radio"/> NÃO	Alargamento de mediastino	<input type="radio"/> SIM	<input checked="" type="radio"/> NÃO
Edema ou congestão pulmonar	<input type="radio"/> SIM	<input checked="" type="radio"/> NÃO	Elevação ou alteração do diafragma	<input type="radio"/> SIM	<input checked="" type="radio"/> NÃO

Figura 34: Interface de avaliação dos laudos pelos observadores especializados

cia foi avaliada pelo modelo de concordância calculado pelo índice kappa (κ), medida entre os observadores dois a dois. O desempenho do sistema na recuperação de laudos foi avaliado com o cálculo da especificidade, sensibilidade, valores preditivos positivo e negativo, incluindo seus intervalos de confiança³⁸. Também foram gerados gráficos ROC para cada uma das duas características investigadas. Também foi empregado o teste de McNemar para se estimar a diferença de desempenho entre o desempenho da metodologia e o padrão de referência.

7.2.2 Resultados

A metodologia foi aplicada no grupo experimental composto por 12.869 relatórios. Foi utilizado um computador pessoal para as operações de tradução, transformação e inferência lógica cabendo o processamento MEDLEE ao servidor do tipo Sun. A operação completa demorou cerca de 2 dias, sendo que a TA foi a fase mais demorada.

A tabela 25 mostra a prevalência dos achados no grupo total de relatórios bem como suas frequências relativas.

Os resultados dos 148 laudos avaliados pelos observadores e comparados a metodologia mostrou uma sensibilidade média de 0,92 (IC 95%: 0,88 – 0,95) e uma especificidade média de 0,99 (IC 95%: 0,99 – 1,00). As medidas de acurácia e seus intervalos de confiança estão

Tabela 25: Prevalência dos achados. A metodologia refere-se ao grupo experimental de 12.869 relatórios. A referência refere-se aos achados dos 148 laudos analisados pelos observadores. O número esperado significa a quantidade de laudos esperados em 148 de acordo as porcentagens obtidas com a metodologia.

Achado	Metodologia(%)	Referência(%)	Número esperado
Alargamento de mediastino	280 (2,18%)	1 (1%)	3
Aneurisma, dilatação ou alteração de aorta	3835 (29,80%)	49 (33%)	44
Atelectasia	85 (0,66%)	1 (1%)	1
Ateromatose	1778 (13,82%)	18 (12%)	20
Cardiomegalia ou aumento de câmara cardíaca	4119 (32,01%)	42 (28%)	47
Cifose	439 (3,41%)	6 (4%)	5
Derrame, espessamento pleural ou alteração de seio cardio/costo-frênico	895 (6,95%)	10 (7%)	10
Dispositivo, aparato ou corpo estranho	2433 (18,91%)	27 (18%)	28
Doença degenerativa vertebral	4063 (31,57%)	40 (27%)	47
DPOC	66 (0,51%)	1 (1%)	1
Edema ou congestão pulmonar	1146 (8,91%)	13 (9%)	13
Elevação ou alteração do diafragma	291 (2,26%)	1 (1%)	3
Escoliose	434 (3,37%)	5 (3%)	5
Fratura costal, vertebral ou óssea	149 (1,16%)	1 (1%)	2
Hipertransparência ou hiperinsuflação pulmonar	48 (0,37%)	3 (2%)	1
Laudo não normal	9214 (71,60%)	106 (71%)	106
Laudo normal	3655 (28,40%)	42 (28%)	42
Massas ou nódulos	481 (3,74%)	9 (6%)	6
Opacidade pulmonar	1318 (10,24%)	9 (6%)	15
Pneumonia	3 (0,02%)	0 (0%)	0
Pneumotórax	4 (0,03%)	0 (0%)	0
Projéteis ou fragmentos	4 (0,03%)	0 (0%)	0
Tuberculose	26 (0,20%)	0 (0%)	0

mostrados na tabela 27. A tabela de contingência 26 contém os resultados da metodologia comparada com os observadores.

Tabela 26: Tabela de Contingência da recuperação de relatórios. Os números referem-se aos 22 achados discriminados pelo sistema e observadores em 148 relatórios

Metodologia	Referência		
	Positivo	Negativo	Total
Positivo	255	18	273
Negativo	23	2960	2983
Total	278	2978	3256

Tabela 27: Acurácia da recuperação de relatórios

	Estimado	Intervalo de Confiança 95%	
		Limite Inferior	Limite Superior
Sensibilidade	0,92	0,88	0,95
Especificidade	0,99	0,99	1,00
VPP	0,93	0,90	0,94
VPN	0,99	0,99	0,99

A concordância entre os três observadores foi calculada com o índice kappa e a porcentagem de concordância pareando-se dois a dois. Os valores estão mostrados na tabela 28.

Tabela 28: Concordância entre os observadores em apontar achados nos relatórios

Parâmetros	Observador 1 vs 2	Observador 1 vs 3	Observador 2 vs 3
Concordância	97%	97%	98%
kappa	0,84	0,84	0,90
(IC - 95%)	0,80 - 0,87	0,80 - 0,87	0,87 - 0,92

Os 148 relatórios comparados com a referência foram analisados em relação a 22 variáveis levando um total de 3256 achados. Foram encontrados 41 erros sendo 23 falsos negativos e 18 falsos positivos, A análise dos motivos que levaram a estes erros está mostrada na tabela 29.

Os erros do MEDLEE referem-se aqueles em que o texto apesar de morfologicamente correto após a tradução não geraram resultado pelo PLN. Não é esperado que o MEDLEE consiga processar todas as expressões e palavras existentes em relatórios oriundos de diferentes instituições. Um exemplo foi a palavra *micronodules*, que apesar de correta e indicar um achado relativamente comum não foi identificado pelo MEDLEE. Erros de morfologia são aqueles relativos à ortografia e construção do texto original em português. Um exemplo é a frase "vascularização pulmonar lombar e periférica aumentada" cuja palavra lombar está claramente fora de

Tabela 29: Erros da metodologia,

Tipo de Erro	Falso Positivo	Falso Negativo
Inferência	11	5
MEDLEE	1	7
Morfologia	-	3
Observador	6	5
Tradução	-	1
Transformação XSLT	-	2
Total	18	23

contexto prejudicando o processamento. Erros de transformação XSLT e de inferência referem-se aqueles oriundos da fase de transformação do resultado do MEDLEE em assertivas lógicas e das regras clínicas contidas na fase de inferência do sistema respectivamente. Já os erros dos observadores foram decorrentes de equívocos na formação da referencia tanto de sobre como subestimação da presença dos achados. Na fase de treinamento foram encontrados mais erros de tradução e interpretação. Por exemplo, podemos citar o ocorrido com a descrição de linhas densas que representam na achados parenquimatosos pulmonares. Na tradução *dense line* foi interpretada pelo MEDLEE como acesso vascular semelhante a *venous line*. Tal achado teve como elemento resultado <device v=line>. Para correta interpretação, a tradução foi modificada para *linear opacity*. Na fase experimental foi encontrado somente um erro de tradução, a não tradução da palavra "pinçamento".

8 *DISCUSSÃO*

Apresentadas as metodologias e os resultados experimentais faremos a discussão dos tópicos principais abordados neste projeto de tese.

O PLN é uma abordagem consistente na obtenção de informações de textos médicos. Este tipo de pesquisa permite, por exemplo, encontrar pacientes portadores de patologias raras ou comuns com o objetivo de conduzir estudos clínicos tanto retrospectivos como prospectivos. O NLP apresenta vantagens em relação a procura de palavras e expressões diretamente no texto. Este tipo de procura direta não permite, por exemplo, a identificação de estruturas sintáticas que influem no conteúdo semântico como negações e ambigüidades. Um estudo⁵ mostrou que o desempenho da procura textual direta teve um desempenho inferior ao NLP na procura de achados em RRT.

A fase inicial de TA é a mais crítica da metodologia. Porém apesar de bem evoluída³ esta tecnologia ainda apresenta falhas. Como o objetivo deste trabalho é investigar a recuperação de informações clínicas, não foi nossa intenção realizar uma análise mais profunda sobre as diversas metodologias de TA que poderiam ser empregadas. Os requerimentos foram de acesso, funcionalidade e desempenho. Optou-se então, pela metodologia de TA baseada em regras. O aplicativo de tradução SYSTRAN foi escolhido, pois atende estes requisitos, possui um histórico na pesquisa da TA⁶⁶ além de estar disponível comercialmente. As metodologias baseadas em corpos paralelos não poderiam ser empregadas devido à inexistência de corpos paralelos de textos médicos em Português e Inglês dentro domínio da radiologia torácica.

Como já mencionado esta abordagem necessita de uma personalização do sistema de TA. Os sistemas baseados em regras empregam conhecimento gramatical e sintático para execução da tradução, estando preparados teoricamente para traduzir qualquer tipo de texto dentro do par de idiomas em questão. A modificação que estes sistemas demandam é a personalização do dicionário. Após a incorporação do dicionário especializado, foi observada uma significativa melhora na qualidade de tradução de acordo com um observador humano como mostrado no capítulo 5.

Após a avaliação da tradução automatizada passamos para os experimentos que processam o texto automaticamente traduzido com o software MEDLEE. O primeiro foi o experimento de extração de tipos de sentenças de relatórios de radiografias torácicas. Consideramos que esta fase é mais uma etapa de desenvolvimento da metodologia do que propriamente uma avaliação de desempenho. O objetivo maior foi o aperfeiçoamento da tradução automatizada com base no resultado do PLN realizado pelo MEDLEE. Foi nesta fase que foram implementadas as memórias de tradução das frases referentes aos achados normais. O resultado do processamento destas frases foi ajustado para a obtenção de elementos pertinentes (*normalfinding*) através de ajustes na tradução.

O mecanismo utilizado pelo sistema automatizado para a determinação da presença dos achados é uma inferência lógica fundamentada em uma ontologia. Esta se contrapõe a outras metodologias^{5,59,60} empregadas em PLN de textos médicos, como inferência por regras, árvores de decisão e redes Bayesianas com resultados tanto probabilísticos como binários.

O documento no formato XML resultado do MEDLEE tem estrutura mais complexa que a necessária para efetuar o raciocínio lógico na atual fase de desenvolvimento. Na simplificação adotada até o momento os elementos relevantes são simplificados no mesmo nível. Durante o treinamento foi o principal fator de erros. Para uma correta identificação de situações clínicas e descrições anatômicas foram necessários ajustes das fases de transformação e inferência. O MEDLEE descreve determinada estrutura anatômica de uma forma estruturada com diversos elementos do XML, enquanto a estrutura do RADLEX, que é utilizada para os mecanismos de inferência, descreve cada estrutura através de apenas um nível.

Os achados radiológicos mais frequentemente observados foram cardiomegalia ou aumento de câmara cardíaca, doença degenerativa vertebral(osteofitose principalmente), dispositivos e aparatos médicos e ateromatose. Supomos que estes relatórios refletem o atendimento ambulatorial da instituição. Dos mais de 40.000 RRT realizados entre janeiro de 2005 e julho de 2006 a grande maioria foi excluída por conter a expressão "por solicitação do paciente / médico assistente, ou devido à urgência do caso, o exame foi entregue sem laudo" ao invés da descrição provavelmente indicando a origem no pronto-socorro e enfermaria destes. Estes achados são concordantes com o padrão de atendimento ambulatorial de um hospital de cardiologia, que possui grande número de pacientes com doenças crônico-degenerativas e de idade mais elevada que a média populacional. Um estudo interessante seria a correlação destes achados com os diagnósticos de atendimentos realizados no ambulatório da instituição. Outro aspecto que corrobora a origem ambulatorial é a maior ocorrência de casos de tuberculose do que de pneumonia.

Os resultados de sensibilidade e especificidade foram semelhantes, se não melhores, que no trabalho com o MedLee de mesmas características publicado pela Columbia University. Hripsak e colaboradores² obtiveram sensibilidade e especificidade de 0,81 e 0,99 enquanto aqui foram obtidos valores de 0,92 e 0,99 respectivamente. Apesar dos textos passarem pelo processo de TA que poderia levar a uma perda no processo isto não foi observado. Provavelmente deve-se a homogeneidade dos relatórios avaliados. Estes oriundos de um hospital especializado refletem um padrão nosológico mais de um Hospital especializados em Cardiologia. Isto leva nos a supor que o treinamento foi eficiente, pois foi realizado num grupo que representou bem o grupo experimental. Uma observação não sistemática confirmou a pouca variabilidade das sentenças destes relatórios. Por outro lado o trabalho citado foi feito sobre um conjunto de laudos de um hospital geral, o que nos faz supor que exista uma maior variabilidade.

Os erros apontados no capítulo anterior acima foram apenas aqueles que prejudicaram o processamento levado a não concordância com a referência. Uma análise não sistemática dos textos originais e traduzidos mostrou erros morfológicos, que apesar de em grande número não afetaram significativamente o desempenho. Os erros morfológicos nos textos em Inglês podem ser divididos em alguns tipos. O primeiro tipo descrito refere-se às palavras que apresentam erros de ortografia no texto de origem em Português. O tradutor automatizado SYSTRAN apresenta a opção de correção ortográfica. Esta correção tem funcionado adequadamente tendo em vista que diferenças de acentuação entre os termos do texto e do dicionário são absolutamente ignoradas e traduzidas corretamente pelo sistema. Esta ferramenta presta-se, por exemplo, a eliminar incorreções referentes ao uso de acentos e outros caracteres especiais funcionando como uma normalização. Outro tipo de erro é a palavra não traduzida pela falta da sua ocorrência no dicionário de tradução automatizada. No desenvolvimento do dicionário todas as palavras presentes nos textos do grupo experimental estavam representadas. Era esperada ocorrência deste tipo de erro durante a fase experimental visto que laudos inéditos foram processados. Em ao menos um laudo submetido à comparação o erro de interpretação foi decorrente de palavra ausente no dicionário de tradução. Certamente no grupo de 12.721 laudos não comparados com a referência mais erros deste tipo tenham ocorrido, sem contar com os erros que não influíram na interpretação.

Certamente o desempenho da metodologia observado deve-se a qualidade da tradução automatizada e a excelência do processador MEDLEE. O primeiro aspecto, discutido acima, deve-se a pouca variabilidade das frases utilizadas na construção dos RRT. Já em relação ao MEDLEE este sistema está em desenvolvimento e uso há pelo menos 13 anos. Além de RRT é capaz de processar uma série de tipos de textos médicos desde outros tipos de textos radiológicos até relatórios de alta médica.

Os resultados obtidos não podem esconder as limitações do modelo aqui apresentado. Um dos problemas neste tipo de metodologia é decorrente da seqüência de eventos desde o paciente até o relatório. A informação deve ser captada pelo filme radiológico, ser identificada e relatada pelo médico, e corretamente digitada ou transcrita pelo profissional digitador.

Também se destaca a cobertura dos achados clínico-radiológicos investigados. Seria praticamente impossível atingir com esta metodologia todos os diagnósticos possíveis de serem elaborados através da RRT. Foram escolhidas sem um critério sistemático 22 situações clínicas freqüentemente encontradas nesta modalidade. Apesar do grande número muitas outras raras e comuns não são contempladas como, por exemplo, derrame pericárdico ou hérnia hiatal. O modo de implementação desta metodologia pressupõe a explicitação da situação investigada, portanto o número de normais falsos positivos poderia ser diminuído com a inclusão de mais diagnósticos. Os dois processos sobre controle, a TA e a inferência lógica deveriam ser ajustadas para isto.

Outra limitação do sistema é a não descrição da região anatômica acometida. Isto não é necessário, por exemplo, no caso de cardiomegalia, quando a estrutura acometida é o coração. Porém no caso de opacidades, derrame pleural, nódulos ou alteração diafragmática poderia haver discriminação do campo pulmonar, lado ou região acometida.

Por fim a maior limitação deste estudo refere-se a generalização dos resultados. O domínio escolhido é representado por um tipo de texto muito simples visando um a prova de conceito. O desempenho satisfatório aqui obtido pode não se repetir no caso de textos mais complexos como, por exemplo, relatórios de ressonância magnética. Certamente será necessário um esforço de adequação do dicionário de tradução e do mecanismo de inferência. Também supomos que o desempenho será melhor se o número de achados investigados for mais restrito. Acreditamos que seria mais eficaz e eficiente se a investigação fosse focada em uma ou duas situações clínicas ao invés de 22 procuradas no trabalho atual.

9 CONCLUSÕES

Dentro do exposto, consideramos que respondemos às questões de investigação propostas como objetivo deste trabalho.

A primeira questão refere-se ao desempenho da TA de acordo com métricas reconhecidas de avaliação. Consideramos que foi satisfatório de acordo com os experimentos de avaliação de desempenho específicos conduzidos. Foi criado um dicionário de tradução especializado e verificamos então a capacidade do sistema em lidar com desempenho significativo tanto os fenômenos gramaticais, sintáticos como a terminologia especializada. A principal contribuição desta fase da metodologia é a elaboração de um léxico específico para o domínio da radiologia torácica bilíngüe Português-Ingês. Este poderá ser utilizado em novos projetos de pesquisa que envolvam recuperação de informações. Também poderá servir de suporte para o desenvolvimento de um processador de linguagem natural próprio para a língua portuguesa. Já está sendo desenvolvida uma versão em Português e Espanhol do léxico RADLEX a partir deste trabalho⁷⁶.

O principal objetivo deste trabalho é uma prova de conceito. Apresentamos uma metodologia que permite esta recuperação de informações através de uma abordagem multilingüe. A viabilidade da hipótese de que se podem traduzir textos em Português e submetê-los ao PLN pela ferramenta MEDLEE foi demonstrada através destas consultas clínicas apresentadas no último experimento.

Uma série de artigos e outras publicações descreve o uso do processamento de linguagem natural para obtenção de dados de textos livres que podem ser processados computacionalmente obtendo informações clínicas como apresentado no capítulo 2. Porém esse trabalho inova ao utilizar a tradução automatizada como pré-processamento para a realização de processamento de linguagem natural.

Foi demonstrado que textos traduzidos automaticamente com boa qualidade podem ser processados e terem seus tipos corretamente identificados e apontados pelos experimentos de extração e recuperação de informações. Dentro dos achados de investigação estabelecidos para

esta tese a semelhança do desempenho da metodologia e da referência pode ser demonstrada. O encontro da maior parte dos achados foi realizado com excelente resultado demonstrando a viabilidade do mecanismo de inferência e da metodologia como um todo.

Dentre as contribuições específicas da fase de PLN destaca-se a utilização do processador MEDLEE em textos originalmente elaborados em outro idioma que não o Inglês. Já havia sido demonstrada sua capacidade em processar textos elaborados em instituição diferente da de desenvolvimento, porém seu uso em idioma diverso é inédito. Ressaltamos que a prova deste conceito abre caminho para utilização do MedLee, e outros processadores de linguagem natural, em outros idiomas além do originalmente planejado.

Em relação ao processo de inferência foi demonstrado o desempenho em se realizar consultas lógicas sobre uma ontologia. Destacamos sua possível aplicação em futuras pesquisas clínicas em novos domínios bastando um esforço de adequação e ajuste. Apesar de ser uma abordagem bem conhecida do campo da inteligência artificial seu uso em medicina com propósitos de investigação clínica em radiologia ainda não é muito difundida.

Como contribuições científicas globais do presente projeto destacamos a introdução do uso do PLN em nosso meio, e a disponibilização para nossa instituição de um novo modelo de investigação e pesquisa adequado ao ambiente informatizado de prontuário eletrônico. Este tipo de metodologia possibilita a extração automática de informações dos dados armazenados em forma digital propiciando a formação de conhecimento. A operação sistemática do REP do INCOR traz a necessidade do desenvolvimento de ferramentas computacionais capazes de oferecer organização e estruturação adequadas, aumentando as capacidades de procura, pesquisa, recuperação e análise de textos livres nele contidos. É também possível ampliação futura para outros domínios, através da incorporação de novos termos e regras lógicas.

O sistema proposto pode oferecer uma ferramenta na análise de casos já inseridos investigando uma grande quantidade em um tempo reduzido, propiciando o estudo de casos retrospectivamente. Poderá também ser útil em estudos prospectivos, classificando automaticamente casos no momento da inserção do texto em tempo real, sendo um auxiliar no seguimento de casos com avaliação da evolução de parâmetros determinados. Pode se também acoplar o sistema a buscas em outros tipos de dados. Por exemplo, o sistema poderia automaticamente procurar um determinado resultado de exame laboratorial em resposta a um achado de imagem representado no texto configurando uma ferramenta de suporte clínico.

Como trabalho futuro sugere-se ampliar a abrangência de textos médicos passíveis de investigação com esta metodologia. Também pode ser aplicada na investigação de achados clínicos visando o rastreamento de patologias específicas. De maior abrangência poderia ser desen-

volvida uma metodologia de processamento automático de textos assim que fossem inseridos na base de dados hospitalares. Assim se formaria um conjunto robusto de dados passíveis de recuperação de informações e produção de conhecimento.

ANEXO

Produção científica relacionada ao projeto

1. **CBIS 2004** - Castilla AC, Furuie SS. Avaliação da Tradução Automatizada de Relatórios de Radiografias de Tórax: Resultados Preliminares. Ribeirão Preto, Brasil; 2004. Pôster apresentado no IX Congresso of Sociedade Brasileira de Informática em Saúde.
2. **MT Summit 2005** - Castilla AC, Bacic AS, Furuie SS. Machine Translation on the Medical Domain: The Role of BLEU/NIST and METEOR in a Controlled Vocabulary Setting. Artigo publicado nos anais: Proceedings of the Tenth Machine Translation Summit. Phuket, Thailand; 2005. p. 47
3. **RSNA 2006** - Castilla AC, Bacic AS, Furuie SS. Performing Logical Reasoning on a RadLex Ontology. Apresentação oral; Radiological Society of North America Scientific Assembly and Annual Meeting. 27/11/2006. Chicago, USA.
4. **Tese de Doutorado** - Castilla AC; Orientador: Furuie SS. Instrumento de investigação clínico-epidemiológica em cardiologia fundamentado no processamento de linguagem natural. Julho de 2007. Universidade de São Paulo, São Paulo, Brasil.
5. **MedInfo 2007** - Castilla AC, Furuie SS, Mendonça EA. Multilingual Information Retrieval in Thoracic Radiology: Feasibility Study. Artigo completo aceito para apresentação oral e publicação em periódico indexado; MedInfo 2007. 23/08/2007. Brisbane, Austrália.
6. **RSNA 2007** - Castilla AC, Bacic AS, Furuie SS. Portuguese and Spanish Versions of RadLex: Term Browser and Report Tagging Application . Exibição educacional aceita pra apresentação; Radiological Society of North America Scientific Assembly and Annual Meeting. Novembro de 2007. Chicago, USA.

REFERÊNCIAS

1. Mitkov R. *The Oxford Handbook of Computational Linguistics*. 1st ed. Oxford, England, UK: Oxford University Press; 2003.
2. Hripcsak G, Austin JHM, Alderson PO, Friedman C. Use of natural language processing to translate clinical information from a database of 889,921 chest radiographic reports. *Radiology*. 2002;224(1):157–163.
3. Hutchins J. 27. In: *Machine Translation: General Overview*. Oxford University Press; 2003. p. 501–511.
4. Hersh W, Li Zhang MS. Teaching English Medical Terminology Using the UMLS Metathesaurus and World Wide Web. *Proceedings of the 1999 AMIA Annual Symposium*. 1999;p. 1078.
5. Chapman WW, Haug PJ. Comparing Expert Systems for Identifying Chest X-ray Reports that Support Pneumonia. *Proceedings of the 1999 AMIA Annual Symposium*. 1999;p. 216–20.
6. Google. *Ferramentas de idiomas*. Mountain View, CA: Google Inc.; 2006. Available from: http://www.google.com/language_tools.
7. Systransoft. *SYSTRAN Language Translation Technology*; 2006. Available from: <http://www.systransoft.com/index.html>.
8. Friedman C, Alderson P, Austin J, Cimino J, Johnson S. A general natural-language text processor for clinical radiology. *Journal of American Medical Informatics Association*. 1994;p. 161–174.
9. Mendonça EA. *Using Automated Extraction from the Medical Record to Access Biomedical Literature*. Columbia University; 2002.
10. Lovis C. Trends and pitfalls with nomenclatures and classifications in medicine. *Int J Med Inform*. 1998;p. 141–8.
11. Bacic AS. *Extração e Documentação de Informação de Laudos Médicos*. Universidade de São Paulo; 2007.
12. Eck M, Vogel S, Waibel A. *Improving Statistical Machine Translation in the Medical Domain using the Unified Medical Language System*. Geneva, Switzerland; 2004.
13. Langlotz CP, Caldwell SA. The completeness of existing lexicons for representing radiology report information. *J Digit Imaging*. 2002;p. 201–5.
14. Friedman C, Liu H, Shagina L, Johnson S, Hripcsak G. Evaluating the UMLS as a source of lexical knowledge for medical language processing. *Proc AMIA Symp*. 2001;p. 189–93.

15. Sobel JL, Pearson ML, Gross K, Desmond KA, Harrison ER, Rubenstein LV, et al. Information content and clarity of radiologists' reports for chest radiography. *Acad Radiol.* 1996;p. 709–17.
16. Marwede D, Fielding M. The epistemological-ontological divide in clinical radiology. *Stud Health Technol Inform.* 2005;p. 749–54.
17. Tuddenham WJ. Glossary of terms for thoracic radiology: recommendations of the Nomenclature Committee of the Fleischner Society. *AJR Am J Roentgenol.* 1984;p. 509–17.
18. Austin JH, Muller NL, Friedman PJ, Hansell DM, Naidich DP, Remy-Jardin M, et al. Glossary of terms for CT of the lungs: recommendations of the Nomenclature Committee of the Fleischner Society. *Radiology.* 1996;p. 327–31.
19. Souza AS, de Araujo Neto C, Jasinovodolinsky D, Marchiori E, Kavakama J, Irion KL, et al. Terminologia para a descrição de tomografia computadorizada do tórax. *Radiol Bras.* 2002;p. 125–128.
20. RSNA. RADLEX: A Lexicon for Uniform Indexing and Retrieval of Radiology Information Resources; 2003. Available from: <http://mirc.rsna.org/radlex/RadLexOverview.pdf>.
21. RSNA. RADLEX: A Lexicon for Uniform Indexing and Retrieval of Radiology Information Resources; 2003. Available from: <http://mirc.rsna.org/radlex/service>.
22. Chomsky N. Three Models for The Description of Language. *IEEE - Transactions on Information Theory.* 1956;.
23. Hripcsak G, Friedman C, Alderson PO, DuMouchel W, Johnson SB, Clayton PD. Unlocking clinical data from narrative reports: a study of natural language processing. *Annals of Internal Medicine.* 1995;p. 681–8.
24. Friedman C, Hripcsak G. Evaluating natural language processors in the clinical domain. *Proc Conference on Natural Language and Medical Concept Representation.* 1997;p. 41–52. Available from: citeseer.ist.psu.edu/friedman97evaluating.html.
25. Somers H. 28. In: *Machine Translation: Latest developments.* Oxford University Press; 2003. p. 512–528.
26. Hovy E, King M, Popescu-Belis A. Principles of Context-Based Machine Translation Evaluation. *Machine Translation.* 2002;17:43–75.
27. vanSlype G. *Critical Methods for Evaluating the Quality of Machine Translation;* 1979. Bureau Marcel van Dijk.
28. Papineni K, Roukos S, Ward T, Zhu W. Bleu: a method for automatic evaluation of machine translation; 2001. Available from: citeseer.ist.psu.edu/papineni02bleu.html.
29. NIST. Automatic Evaluation of Machine Translation Quality Using N-gram Co-Occurrence Statistics. National Institute of Standards and Technology; 2001. Available from: <http://www.nist.gov/speech/tests/mt/doc/ngram-study.pdf>.
30. Lavie A, Sagae K, Jayaraman S. The Significance of Recall in Automatic Metrics for MT evaluation. In: *Proceedings of the 6th Conference of the Association for Machine Translation in the Americas;* 2004. .

31. Minsky M. A Framework for Representing Knowledge. Cambridge, MA, USA; 1974. Available from: <http://web.media.mit.edu/~minsky/papers/Frames/frames.html>.
32. Rogers JE. Quality assurance of medical ontologies. *Methods Inf Med*. 2006;p. 267–74.
33. Smith B. Workshop on Ontology of Images. Stanford , CA; 2006. Available from: http://www.bioontology.org/wiki/index.php/Workshop_on_Ontology_of_Images.
34. Marwede D. Relations in Image Ontologies and Anatomy Ontologies. Stanford , CA; 2006. Available from: http://www.bioontology.org/wiki/index.php/Workshop_on_Ontology_of_Images.
35. Smith B, Ceusters W, Klagges B, Kohler J, Kumar A, Lomax J, et al. Relations in Biomedical Ontologies. *Genome Biology*. 2005;p. 46.
36. Protégé. The Protege Project; 2000. Available from: <http://protege.stanford.edu>.
37. Kuipers B. Algernon for Expert Systems. Austin, Texas; 1994. Available from: <http://www.cs.utexas.edu/users/qr/algy/algy-expsys/algy-expsys.html>.
38. Woodward M. *Epidemiology: Study design and data analysis*. 1st ed. Boca Raton: Chapman and Hall/CRC; 1999.
39. Tiitola M, Kivisaari L, Tervahartiala P, Palomäki M, Kivisaari RP, Mankinen P, et al. Estimation or quantification of tumour volume? CT study on irregular phantoms. *Acta Radiologica*. 2001;p. 101.
40. Kundel HL, Polansky M. Measurement of observer agreement. *Radiology*. 2003;p. 303–308.
41. Friedman C, Hripcsak G. Natural language processing and its future in medicine: Can Computers Make Sense out of Natural Language Text. *Acad Med*. 1999;p. 890–5.
42. Martha AS. Recuperação de informação em campos de texto livre de Prontuários eletrônicos do paciente baseada em Semelhança semântica e ortográfica. Universidade Federal de São Paulo; 2005.
43. Friedman C. A broad-coverage natural language processing system. In: *Proc AMIA Annu Fall Symp*; 2000. p. 270–4.
44. Lussier YA, Shagina L, Friedman C. Automating SNOMED coding using medical language understanding: a feasibility study. *Proc AMIA Symp*. 2001;p. 418–22.
45. Larkey LS, B CW. Automatic Assignment of ICD9 Codes to Discharge Summaries: Technical report. Amherst, MA; 1995. Available from: <http://citeseer.nj.nec.com/larkey95automatic.html>.
46. Delamarre D, Burgun A, Seka LP, Le Beux P. Automated coding of patient discharge summaries using conceptual graphs. *Methods Inf Med*. 1995;p. 345–51.
47. De Bruijn LM, Ver Heijen E, Van Nes FL, Arends JW. Assigning SNOMED codes to natural language pathology reports. *Medical Informatics Europe*. 1996;p. 198–202.

48. Fiszman M, Chapman WW, Aronsky D, Evans RS, Haug PJ. Automatic detection of acute bacterial pneumonia from chest X-ray reports. *J Am Med Inform Assoc.* 2000;p. 593–604.
49. Jain NL, Knirsch CA, Friedman C, Hripcsak G. Identification of suspected tuberculosis patients based on natural language processing of chest radiograph reports. *Proc AMIA Annu Fall Symp.* 1996;p. 542–6.
50. Friedman C, Knirsch C, Shagina L, Hripcsak G. Automating a severity score guideline for community-acquired pneumonia employing medical language processing of discharge summaries. *Proc AMIA Symp.* 1999;p. 256–60.
51. El-Gamal SS, Esmail MM. Understanding clinical narrative text. *Med Inform (Lond).* 1995;p. 161–73.
52. Mutalik PG, Deshpande A, Nadkarni PM. Use of general-purpose negation detection to augment concept indexing of medical documents: a quantitative study using the UMLS. *J Am Med Inform Assoc.* 2001;p. 598–609.
53. Sneiderman CA, Rindflesch TC, Bean CA. Identification of anatomical terminology in medical text. *Proc AMIA Symp.* 1998;p. 428–32.
54. Ceusters W, Spyns P, DeMoor G. From syntactic-semantic tagging to knowledge discovery in medical texts. *Int J Med Inf.* 1998;p. 149–57.
55. Sponsler JL. HPARSER: extracting formal patient data from free text history and physical reports using natural language processing software. *Proc AMIA Symp.* 2001;p. 637–41.
56. Hersh W, Mailhot M, Arnott-Smith C, Lowe H. Selective automated indexing of findings and diagnoses in radiology reports. *J Biomed Inform.* 2001;p. 262–73.
57. Hersh WR DL. SAPHIRE International: a tool for cross-language information retrieval. *Proceedings of the 1998 AMIA Annual Symposium.* 1998;p. 673–677.
58. Friedman C. Towards a comprehensive medical language processing system: methods and issues. In: *Proc AMIA Annu Fall Symp; 1997.* p. 595–9.
59. Mendonca EA, Haas J, Shagina L, Larson E, Friedman C. Extracting Information on Pneumonia in Infants Using Natural Language Processing of Radiology Reports. *Journal of Biomedical Informatics.* 2005;38:314–321.
60. Hripcsak G, Kuperman GJ, Friedman C. Extracting Findings from Narrative Reports: Software Transferability and Sources of Physician Disagreement. *Methods Inf Med.* 1998;37:1–7.
61. Francis WN, Kucera H. *Brown Corpus Manual, Manual of Information to accompany a Standard Corpus of Present-Day Edited American English, for use with Digital Computers, Revised and Amplified; 1979.*
62. Brill E. A Simple Rule-Based Part of Speech Tagger. In: *Proceedings of the Third Annual Conference on Applied Natural Language Processing.* Association for Computational Linguistics; 1992. .

63. McCarley JS. Should we translate the documents or the queries in cross-language information retrieval? In: Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics. Morristown, NJ, USA: Association for Computational Linguistics; 1999. p. 208–214.
64. Dillinger M. Dictionary Development Workflow for MT: Design and Management. In: Proceedings of the Eighth Machine Translation Summit; 2001. .
65. Chang MY, Sun YC, Chang CF, et al. A Free Radiology Dictionary Made From Abstract Corpus Of The Radiology And Radiographics. In: Scientific Assembly and Annual Meeting Program of the Radiological Society of North America. Chicago; 2003. .
66. SYSTRAN. Systran Premium 5.0; 2003.
67. Castilla AC, Bacic AS, Furuie SS. Machine Translation on the Medical Domain: The Role of BLEU/NIST and METEOR in a Controlled Vocabulary Setting. In: Proceedings of the Tenth Machine Translation Summit. Phuket, Thailand; 2005. p. 47.
68. Culy C, Riehemann SZ. The Limits of N-gram Translation Evaluation Metrics. In: Proceedings of the 9th Machine Translation Summit. New Orleans, LO; 2003. .
69. Hovy E, King M, Popescu-Belis A. Principles of Context-Based Machine Translation Evaluation. Machine Translation. 2002;p. 1–33.
70. Castilla AC, Furuie SS. Avaliação da Tradução Automatizada de Relatórios de Radiografias de Tórax: Resultados Preliminares. Ribeirao Preto, Brazil; 2004. Poster presented on the 9th Congress of Sociedade Brasileira de Informatica em Saude.
71. Castilla AC, Bacic AS, Furuie SS. Performing Logical Reasoning on a RadLex Ontology. In: 92nd Scientific Assembly and Annual Meeting Program of the Radiological Society of North America. Chicago; 2006. .
72. Krauthammer M, Hripcsak G. A knowledge model for the interpretation and visualization of NLP-parsed discharged summaries. Proc AMIA Symp. 2001;p. 339–43.
73. Friedman C. WWW MedLee; 1995. Available from: <http://lucid.cpmc.columbia.edu/medlee/demo/cgi-bin/medlee.cgi>.
74. Castilla AC, Furuie SS. Multilingual approach to MedLee using Machine Translation: Initial results on chest radiography reports; 2006. Report.
75. Friedman C, Hripcsak G, Shagina L, Liu H. Representing information in patient reports using natural language processing and the extensible markup language. J Am Med Inform Assoc. 1999;p. 76–87.
76. Castilla AC, Bacic AS, Furuie SS. Portuguese and Spanish Versions of RadLex: Term Browser and Report Tagging Application. In: 93rd Scientific Assembly and Annual Meeting Program of the Radiological Society of North America. Chicago; 2007. .