

VIVIANE POSSAMAI

**ASSOCIAÇÕES SINTAGMÁTICAS COMO UNIDADES DE
TRADUÇÃO ESPECIALIZADAS E PROPOSTA DE
FERRAMENTA DE APOIO AO TRADUTOR**

**PORTO ALEGRE
2010**

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE LETRAS
PROGRAMA DE PÓS-GRADUAÇÃO EM LETRAS
ÁREA: ESTUDOS DA LINGUAGEM
ESPECIALIDADE: TEORIAS DO TEXTO E DO DISCURSO
LINHA DE PESQUISA: LEXICOGRAFIA E TERMINOLOGIA: RELAÇÕES
TEXTUAIS**

**ASSOCIAÇÕES SINTAGMÁTICAS COMO UNIDADES DE
TRADUÇÃO ESPECIALIZADAS E PROPOSTA DE FERRAMENTA DE
APOIO AO TRADUTOR**

VIVIANE POSSAMAI

ORIENTADORA: PROFa. DRa. MARIA JOSÉ BOCORNY FINATTO

Tese de Doutorado em TEORIAS DO TEXTO E DO DISCURSO, apresentada como requisito parcial para a obtenção do título de Doutor pelo Programa de Pós-Graduação em Letras da Universidade Federal do Rio Grande do Sul

**PORTO ALEGRE
2010**

*Dedico esta tese ao Felipe, meu marido, companheiro e incentivador,
porque ele me estende sempre a mão,
com sabedoria, disposição e amor.*

AGRADECIMENTOS

É muito bom chegar ao término de um trabalho como este e ir lembrando da ajuda e do apoio que recebi de tantas pessoas queridas.

A “primeira fatia do bolo” eu ofereço à minha orientadora, Maria José Finatto, porque ela tem sido, desde sempre, uma incentivadora especial, ajudando a dar forma a meus pensamentos incompletos. Agradeço pela acolhida dos meus anseios com esta tese, pela atenção, pelos aconselhamentos, ensinamentos e questionamentos (esses três, para ficar só nos “mentos”). Agradeço também à amiga MaryJo, pelos banhos de piscina com bolos gostosos, pelos almoços, pelas dicas de filmes e trocas de comentários sobre programas de TV, os quais não podemos revelar. Sem a tua confiança no meu trabalho, a tua amizade e a graça dos teus comentários tudo poderia ter sido mais difícil, senão impossível.

Agradeço à minha orientadora do estágio “sanduíche” na Universidade de Birmingham, Gabriela Saldanha, pelos encontros, pela leitura do trabalho e pelas ótimas e imprescindíveis indicações de leitura. Além disso, agradeço pela oportunidade de assumir uma de suas turmas da disciplina de Tradução e Tecnologia durante um trimestre.

Às amigas das horas calmas e turbulentas – em especial, Cristiane Kilian, Cecília Flores, Luciane Leipnitz, Simone Sarmiento, Sandra Loguércio, Janisa Antoniazzi, Susana Kerschner, Fabiana Schwarstzhaupt e Patrícia Chittoni – agradeço o carinho, o incentivo, a confiança e a companhia. Agradeço imensamente à Susi Azeredo pela dedicação, rapidez e tranquilidade durante a revisão. Kelly Oliveira, obrigada pela mão à distância com o livro do Hoey.

À professora e amiga, Cleci Bevilacqua, que lerá partes desta tese pela terceira vez e que fez a primeira sugestão de uso das unidades de tradução neste trabalho, na ocasião do meu Exame de Qualificação.

Um agradecimento muito especial para Teddy, Fiona, Lisandro e Valeria: *thank you for the opportunities and for understanding my needs while I was involved with this study;*

Adri Viegas e Gaby Silioni, minhas “assistentes de pesquisa”, foi e é muito bom contar com vocês; Fausto Silveira e Dea Vigna, obrigada pelas dicas e ajudas; Salete Cechin, Lu Sarti e Cibele: agradeço pelas palavras de incentivo, pelo carinho e pelos momentos de descontração; Dé, Aline, Letícia: valeu a parceria, a atenção e a compreensão; queridos Dave S., Merita e Sole: *I miss you so much, thank you for making hard days better and happier!*

Ao Thiago Pardo, Leandro H. M. de Oliveira, Sandra Aluísio e Aline Villavicencio pela ajuda no início da empreitada, nas minhas primeiras tentativas de aproximação com o Processamento da Linguagem Natural.

Agradeço também ao Robert Coulthard, por ter fornecido o cópuz do Jornal de Pediatria, já pedindo desculpas por não ter enviado mais notícias sobre o andamento do trabalho; além do Robert, agradeço também ao Gilson Mattos, que foram os principais tradutores do Jornal de Pediatria durante os anos selecionados para o cópuz.

Ao CNPq, pela bolsa do estágio no exterior, sem a qual, com certeza, os resultados dessa pesquisa teriam sido mais limitados, e à CAPES, pela bolsa concedida durante os outros três anos.

Aos meus pais, agradeço a compreensão nesses últimos meses, por terem feito de tudo para que eu tivesse o ambiente ideal para concluir esta tese. Mas mais importante do que isso, obrigada por serem essa fonte inesgotável e incansável de amor e dedicação. Admiro e amo muito vocês!

À minha querida amiga e sogra, Iná Spencer, por ser uma pessoa tão iluminada e zelosa, por cuidar da gente mesmo à distância, por sua sabedoria de vida sempre tão bem transmitida por meio de palavras doces e certeiras.

E ao Felipe, por tantas coisas... Por me conhecer tão bem e ter dito a frase que me deu coragem para encarar o desafio, no momento em que receava os quatro anos de envolvimento que teria pela frente: “Faz o doutorado agora e pensa nisso por quatro anos, ou não faz e pensa nisso o resto da tua vida”. Obrigada por prezar tanto pela minha felicidade e por me mostrar todos os dias qual é o sentido de amar e estar junto. Não posso deixar de agradecer também por ele ter tornado real o meu objetivo com esta tese, desenvolvendo o protótipo do sistema ASSIM, aqui apresentado.

Aos amigos que não estão aqui mencionados, desde já, se sintam devidamente agradecidos e abraçados. De verdade, a todos, muito obrigada! Que a minha alegria na conclusão de mais esta etapa possa retornar na vida de vocês!

RESUMO

Este trabalho tem dois objetivos principais: o primeiro objetivo é investigar uma hipótese de fundo teórico; o segundo objetivo envolve propor uma aplicação prática para essa hipótese. Todo o trabalho desta pesquisa é feito tendo em vista oferecer uma contribuição para a prática da tradução. Na primeira parte da tese, investiga-se a hipótese de que há, nos textos especializados, associações sintagmáticas que: a) são combinações mais ou menos fixas e frequentes de palavras, b) não são necessariamente contínuas, c) estão associadas no nível sintagmático por uma atração semântica e/ou pragmática e/ou gramatical e d) não necessariamente contêm um termo em sua composição. Tais associações são consideradas Unidades de Tradução Especializadas. Discutem-se conceitos relacionados a associações sintagmáticas, produto e processo tradutório e pondera-se sobre a importância de o tradutor conhecer essas associações e sobre a relação das mesmas com o caráter especializado do texto. Para conduzir nossa investigação, foi utilizado um corpus composto de artigos do Jornal de Pediatria, um corpus de artigos da Revista Química Nova e o corpus Lácio-Web, composto de textos jornalísticos. Foram também empregadas ferramentas da Linguística de Corpus, o software WordSmith Tools e o ConcGram. Na segunda parte do trabalho, é apresentada a ferramenta informatizada de apoio ao tradutor que idealizamos e construímos, a ASSIM, que faz uso das associações sintagmáticas identificadas. Nesta parte, além de demonstrar como foi construído o banco de dados que integra o sistema e o próprio sistema, ressaltamos vantagens e benefícios do uso de subsegmentos, como as associações, em sistemas de apoio ao tradutor. Os resultados obtidos ressaltam a importância do estudo e uso de elementos do texto especializado diferentes de termos e fraseologias, principalmente no que concerne o desenvolvimento de materiais de apoio e referência destinados a tradutores.

ABSTRACT

This study is oriented by two major goals: the first is to investigate a theoretical hypothesis; the second implies the development of a practical application for such hypothesis. This work intends to contribute to the translation practice. In the first part of the thesis we investigate the hypothesis that in specialized texts there are syntagmatic associations that: a) are combinations not so fixed and frequent of words, b) not necessarily contiguous; c) are associated in the syntagmatic level by a semantic and/or pragmatic and/or grammar force and d) do not necessarily contain a term. Such associations are considered Specialized Translation Units. We discuss about concepts related to syntagmatic associations, translation product and process and about the importance of knowing such associations and about their relation with the specialized features of the text. In order to carry out this study, we used three different corpora: Medical research articles from *Jornal de Pediatria*, Chemistry research articles from *Revista Química Nova* and newspaper articles from *Lácio-Web*. We also used Corpus Linguistics tools, such as WordSmith Tools and ConcGram. In the second part of this thesis we introduced the prototype of a computer-aided translation tool we devised, named ASSIM. Besides demonstrating how the system database was built, we highlight advantages and benefits of using sub-segments, such as the syntagmatic associations detected, in translation tools. Results obtained highlight the importance of studying and using elements of the specialized text that are different from terms and specialized phraseologies, specially in what concerns the development of support and reference material targeted at translators.

ÍNDICE DE FIGURAS

Figura 1	Tela do analisador da tabela de colocados	106
Figura 2	Texto a ser investigado (p-in-79.txt) e seleção da palavra de busca	124
Figura 3	Lista dos <i>congrams</i> de duas palavras	124
Figura 4	Lista dos <i>congrams</i> de três palavras	125
Figura 5	Lista dos <i>congrams</i> de quatro palavras	125
Figura 6	Lista dos <i>congrams</i> de cinco palavras	125
Figura 7	Linhas com os contextos do <i>congram</i> selecionado	126
Figura 8	Linhas de concordância do <i>congram</i> crianças/anos/idade	132
Figura 9	Linhas de concordância do <i>congram</i> crianças/arterial/pressão	132
Figura 10	Linhas de concordância do <i>congram</i> crianças/não/uso	133
Figura 11	Linhas de concordância do <i>congram</i> crianças/anos/entre	133
Figura 12	Linhas de concordância do <i>congram</i> crianças/anos/menores	134
Figura 13	Linhas de concordância do <i>congram</i> crianças/abaixo/idade	134
Figura 14	Linhas de concordância do <i>congram</i> crianças/anos/dois	135
Figura 15	Linhas de concordância do <i>congram</i> crianças/idade/meses	135
Figura 16	Linhas de concordância do <i>congram</i> estudo/foi/realizado	136
Figura 17	Linhas de concordância do <i>congram</i> estudo/foi/pelo	137
Figura 18	Linhas de concordância do <i>congram</i> estudo/aprovado/foi	138
Figura 19	Linhas de concordância do <i>congram</i> estudo/aprovado/foi	139
Figura 20	Linhas de concordância do <i>congram</i> estudo/foram/pacientes	140
Figura 21	Linhas de concordância do <i>congram</i> estudo/este/foi	141
Figura 22	Linhas de concordância do <i>congram</i> estudo/foi/transversal	141
Figura 23	Linhas de concordância do <i>congram</i> estudo/se/realizou	142
Figura 24	Linhas de concordância do <i>congram</i> pacientes/masculino/sexo	143
Figura 25	Linhas de concordância do <i>congram</i> pacientes/se/observou	144
Figura 26	Linhas de concordância do <i>congram</i> pacientes/anos/idade	144
Figura 27	Linhas de concordância do <i>congram</i> pacientes/foram/encaminhados	145
Figura 28	Linhas de concordância do <i>congram</i> pacientes/faixa/etária	145
Figura 29	Linhas de concordância do <i>congram</i> pacientes/foram/incluídos	146
Figura 30	Linhas de concordância do <i>congram</i> pacientes/alogênico/TMO	146
Figura 31	Linhas de concordância do <i>congram</i> pacientes/entre/p	147
Figura 32	Linhas de concordância do <i>congram</i> crianças/idade/anos	148
Figura 33	Linhas de concordância do <i>congram</i> crianças/anos/entre	149
Figura 34	Linhas de concordância do <i>congram</i> crianças/estudo/presente	149
Figura 35	Linhas de concordância do <i>congram</i> crianças/estudo/neste	150
Figura 36	Linhas de concordância do <i>congram</i> crianças/dieta/leite	150
Figura 37	Linhas de concordância do <i>congram</i> crianças/idade/meses	151

Figura 38	Linhas de concordância do conogram crianças/meses/vida	151
Figura 39	Linhas de concordância do conogram crianças/com/TVP	152
Figura 40	Tela em branco na página de acesso dos glossários do proz.com	193
Figura 41	Preenchimento da consulta	194
Figura 42	Resultado da busca	194
Figura 43	Comentário acerca do termo e sugestão de uso	195
Figura 44	Exemplo de consulta de uma estrutura fraseológica	196
Figura 45	Tela do Similis	205
Figura 46	Esquema com a visão do banco de dados	211
Figura 47	Visão geral do funcionamento do sistema	213
Figura 48	Tela do sistema ASSIM	214
Figura 49	Tela do sistema ASSIM com uma frase a ser pesquisada, no momento em que o usuário clica em Analisa	215
Figura 50	Tela do sistema ASSIM apresentando um segmento de exemplo ao consulente, no campo EN	216
Figura 51	Tela do sistema ASSIM apresentando uma segunda opção com as mesmas associações	216
Figura 52	Tela do sistema ASSIM apresentando uma terceira opção com as mesmas associações	217
Figura 53	Tela do sistema ASSIM apresentando uma quarta opção com as mesmas associações	217
Figura 54	Segmento encontrado pelo Trados na memória	219
Figura 55	Segmento encontrado pelo Trados na memória	220
Figura 56	Segmento encontrado pela ASSIM no banco	220
Figura 57	Segmento encontrado pela ASSIM no banco	221
Figura 58	Indicação de que o exemplo pertence à seção Métodos	221

ÍNDICE DE QUADROS

Quadro 1	Frequência comparada de ocorrência de seções em artigos de Linguística Computacional e Cardiologia	94
Quadro 2	Constituição do <i>córpus</i>	96
Quadro 3	Colocados da palavra <i>meses</i>	97
Quadro 4	Expressão de busca: <i>idade média de x meses</i>	98
Quadro 5	Expressão de busca: <i>com média de x meses</i>	98
Quadro 6	Colocados da palavra <i>transversal</i>	99
Quadro 7	Expressão de busca: <i>estudo transversal</i>	101
Quadro 8	Expressão de busca: <i>corte transversal</i>	101
Quadro 9	Exemplo do cálculo do Quadro de Colocados apresentando os valores de Mutual Information e T-score	106
Quadro 10	Cálculo do quadro de colocados para <i>transversal</i>	107
Quadro 11	Listas de palavras-chave de cada seção	110
Quadro 12	Quadro de colocados de crianças na <i>Introdução</i>	112
Quadro 13	Quadro de resultado do cálculo MI	113
Quadro 14	Quadro de colocados de crianças em <i>Métodos</i>	114
Quadro 15	Quadro de resultado do cálculo MI	115
Quadro 16	Quadro de colocados de crianças em <i>Resultados</i>	115
Quadro 17	Quadro de resultado do cálculo MI	116
Quadro 18	Quadro de colocados de crianças em <i>Discussão</i>	117
Quadro 19	Quadro de resultado do cálculo MI	118
Quadro 20	Número de palavras do <i>córpus</i> em português	127
Quadro 21	Número de palavras do <i>córpus</i> em inglês	
Quadro 22	Cinco primeiros substantivos mais frequentes de cada seção	130
Quadro 23	<i>Concgrams</i> de 3 palavras para a palavra-nodo <i>crianças</i>	131
Quadro 24	<i>Concgrams</i> de 3 palavras para a palavra-nodo <i>estudo</i>	135
Quadro 25	<i>Concgrams</i> de 3 palavras para a palavra-nodo <i>pacientes</i>	142
Quadro 26	<i>Concgrams</i> de 3 palavras para a palavra-nodo <i>crianças</i>	147
Quadro 27	<i>Concgrams</i> de palavras-nodo da seção <i>Introdução</i>	157
Quadro 28	<i>Concgrams</i> de palavras-nodo da seção <i>Métodos</i>	158
Quadro 29	<i>Concgrams</i> de palavras-nodo da seção <i>Resultados</i>	160
Quadro 30	<i>Concgrams</i> de palavras-nodo da seção <i>Resultados</i>	162
Quadro 31	Frequências absolutas e relativas de ocorrência no <i>córpus</i> dos substantivos mais frequentes	165
Quadro 32	Frequência relativa e absoluta dos <i>concgrams</i> de 3 palavras, utilizando a palavra-nodo <i>crianças</i> sem o uso de uma <i>stoplist</i>	166
Quadro 33	frequência relativa e absoluta dos <i>concgrams</i> de 3 palavras utilizando a	166

	palavra-nodo <i>crianças</i> com o uso de uma <i>stoplist</i>	
Quadro 34	<i>Concgrams</i> mais frequentes das palavras mais frequentes de cada cópús sem uma <i>stoplist</i> ativada	167
Quadro 35	<i>Concgrams</i> mais frequentes das palavras mais frequentes de cada cópús com uma <i>stoplist</i> ativada	167
Quadro 36	Número de <i>concgrams</i> de cada cópús, com a <i>stoplist</i> ativada	168
Quadro 37	Linhas de concordância e traduções de <i>concgrams</i> da Introdução	170
Quadro 38	Linhas de concordância e traduções de <i>concgrams</i> dos Métodos	171
Quadro 39	Linhas de concordância e traduções de <i>concgrams</i> dos Resultados	172
Quadro 40	Linhas de concordância e traduções de <i>concgrams</i> autores/outros/como	173
Quadro 41	Exemplo da formação dos <i>concgrams</i> de 2, 3, 4 e 5 palavras para <i>objetivo</i> em coocorrência com <i>avaliar</i>	208
Quadro 42	Palavras-nodo para as quais foram processados os <i>concgrams</i> do banco de dado	208
Quadro 43	Exemplo do formato do cópús alinhado	210

LISTA DE ABREVIATURAS

ASSIM	ASsociações Sintagmáticas em Medicina
CAT	<i>Computer-Aided Translation</i>
DBMT	<i>Dialogue-Based Machine Translation</i>
EBMT	<i>Example-Based Machine Transl</i>
ELU	<i>Extended Lexical Units</i>
FFLCH	Faculdade de Filosofia, Letras e Ciências Humanas
GST	Gramática Sistemico Funcional
IME	Instituto de Matemática e Estatística
JPED	Jornal de Pediatria
LA	Linguística Aplicada
LTM	<i>Lexeme-Based Translation Memory</i>
MSU	<i>Meaning Shifting Units</i>
NBMT	<i>Neural-Based Machine Translation</i>
NILC	Núcleo Interinstitucional de Linguística Computacional
PLN	Processamento da Linguagem Natural
SBMT	<i>Statistical-Based Machine Translation</i>
STM	<i>String-Based Translation Memory</i>
TA	Tradução Automática
TAP	<i>Think-aloud protocols</i>
UFE	Unidades Fraseológicas Especializadas
USE	Unidades de Significação Especializada
UTE	Unidade de Tradução Especializadas
TE	Texto de Entrada
TMs	<i>Translation Memories</i>

SUMÁRIO

INTRODUÇÃO	17
1. <i>Objetivos</i>	19
2. <i>Motivação</i>	20
3. <i>Questões de pesquisa e hipótese</i>	22
4. <i>Objeto de estudo e unidade de análise</i>	24
5. <i>Posicionamento do trabalho e embasamento teórico</i>	25
6. <i>Organização do trabalho</i>	30
PARTE I	31
ASSOCIAÇÕES SINTAGMÁTICAS COMO UNIDADES DE TRADUÇÃO ESPECIALIZADAS	31
CAPÍTULO I	32
ENFOCANDO AS UNIDADES DE ANÁLISE E CONCEITOS RELACIONADOS	32
1. <i>Colocações</i>	34
1.1. <i>Diferentes classificações para estudo e descrição das colocações</i>	39
1.2. <i>O princípio idiomático (idiom principle)</i>	42
1.3. <i>Lexical priming</i>	42
2. <i>Concgrams</i>	44
3. <i>Fraseologia</i>	47
4. <i>Definindo nossas unidades de análise</i>	51
CAPÍTULO II	54
O ENCADEAMENTO SINTAGMÁTICO NO PRODUTO E NO PROCESSO TRADUTÓRIO	54
1. <i>A relevância com foco no produto: risco e vantagem, erro e inadequação</i>	55
2. <i>A relevância com foco no processo: a teoria interpretativa e as unidades de tradução</i>	59
2.1. <i>O processo tradutório</i>	60
2.2. <i>Concepções sobre o processo tradutório</i>	62
2.2.1. <i>A Teoria Interpretativa da Tradução</i>	64
2.2.1.1. <i>Primeiro estágio: percepção do enunciado</i>	66
2.2.1.2. <i>Segundo estágio: desverbalização</i>	66
2.2.1.3. <i>Terceiro estágio: reverbalização ou reexpressão</i>	67
3. <i>Discussão</i>	68
CAPÍTULO III	72
O PROCESSO DE SINTAGMATIZAÇÃO E AS UNIDADES DE TRADUÇÃO	72
1. <i>As associações sintagmáticas e a tradução</i>	73
2. <i>As unidades de tradução</i>	79
CAPÍTULO IV	84
ABORDAGENS INICIAIS À QUESTÃO DE PESQUISA	84
1. <i>Primeiro estudo-piloto: abordagem via matrizes fraseológicas</i>	85
1.1. <i>Metodologia</i>	86
1.2. <i>Resultados</i>	88
1.2.1. <i>1º passo: identificação de sequências notáveis</i>	88

1.2.2.	2º passo: confirmação.....	89
2.	Segundo estudo-piloto: abordagem via seção dos artigos.....	92
2.1.	Descrição do <i>córpus</i>	94
2.2.	Metodologia.....	95
2.3.	Resultados.....	95
2.3.1.	Resultados dos processamentos para a palavra <i>meses</i>	95
2.3.2.	Resultados dos processamentos para a palavra <i>transversal</i>	98
2.4.	Discussão e sistematização dos resultados.....	100
3.	Terceiro estudo-piloto: abordagem via cálculo de Informação Mútua.....	103
3.1.	Metodologia.....	104
3.2.	Considerações sobre os resultados.....	106
4.	Quarto estudo-piloto: avaliando a mesma palavra em diferentes seções e seleção de palavras-nodo.....	107
4.1.	Detecção de palavras de busca.....	107
4.2.	Avaliando a mesma palavra em diferentes seções.....	111
4.2.1.	Em <i>Introdução</i>	111
4.2.2.	Em <i>Métodos</i>	113
4.2.3.	Em <i>Resultados</i>	114
4.2.4.	Em <i>Discussão</i>	116
4.3.	Resultados para a análise da ocorrência de crianças em todas as seções.....	118
5.	Conclusões.....	119
CAPÍTULO V.....		121
INVESTIGAÇÃO DA HIPÓTESE: MÉTODOS E RESULTADOS.....		121
1.	Descrição do aplicativo <i>ConcGram</i>	123
2.	Descrição dos <i>córpus</i>	126
2.1.	<i>Jornal de Pediatria</i>	127
2.1.1.	<i>Jornal de Pediatria</i> traduzido.....	127
2.2.	<i>Lácio-Web</i>	128
2.3.	<i>Córpus de contraste: Química Nova</i>	128
3.	<i>Etapa 1: concgrams e a flexibilidade das unidades de tradução</i>	129
3.1.	Metodologia.....	130
3.1.1.	Seleção das palavras-nodo.....	130
3.1.2.	Processamento do texto para obter concgrams.....	131
3.2.	Resultados da investigação da <i>Etapa 1</i>	132
3.2.1.	Crianças na seção <i>Introdução</i>	132
3.2.1.1.	crianças/anos/idade.....	132
3.2.1.2.	crianças/arterial/pressão.....	133
3.2.1.3.	crianças/não/uso.....	133
3.2.1.4.	crianças/anos/entre.....	134
3.2.1.5.	crianças/anos/menores.....	134
3.2.1.6.	crianças/abaixo/idade.....	134
3.2.1.7.	crianças/anos/dois.....	135
3.2.1.8.	crianças/idade/meses.....	135
3.2.2.	Estudo na seção <i>Métodos</i>	136
3.2.2.1.	estudo/foi/realizado.....	136
3.2.2.2.	estudo/foi/pelo.....	137
3.2.2.3.	estudo/aprovado/foi.....	139
3.2.2.4.	estudo/crianças/foram.....	139
3.2.2.5.	estudo/foram/pacientes.....	140
3.2.2.6.	estudo/este/foi.....	141
3.2.2.7.	estudo/foi/transversal.....	142
3.2.2.8.	estudo/se/realizou.....	142
3.2.3.	Pacientes na seção <i>Resultados</i>	143
3.2.3.1.	pacientes/masculino/sexo.....	143
3.2.3.2.	pacientes/se/observou.....	144
3.2.3.3.	pacientes/anos/idade.....	145
3.2.3.4.	pacientes/foram/encaminhados.....	145
3.2.3.5.	pacientes/etária/faixa.....	146
3.2.3.6.	pacientes/foram/incluídos.....	147

3.2.3.7.	<i>pacientes/alogênico/TMO</i>	147
3.2.3.8.	<i>pacientes/entre/p</i>	147
3.2.4.	<i>Crianças na seção Discussão</i>	148
3.2.4.1.	<i>crianças/anos/idade</i>	149
3.2.4.2.	<i>crianças/anos/entre</i>	149
3.2.4.3.	<i>crianças/estudo/presente</i>	150
3.2.4.4.	<i>crianças/estudo/neste</i>	150
3.2.4.5.	<i>crianças/dieta/leite</i>	151
3.2.4.6.	<i>crianças/idade/meses</i>	152
3.2.4.7.	<i>crianças/meses/vida</i>	152
3.2.4.8.	<i>crianças/adolescentes/TVP</i>	153
3.3.	<i>Discussão sobre os resultados da Etapa 1</i>	153
4.	<i>Etapa 2: associações semânticas e pragmáticas</i>	155
4.1.	<i>Metodologia</i>	156
4.1.1.	<i>Seleção das palavras-nodo</i>	156
4.1.2.	<i>Processamento para obter concgrams</i>	157
4.2.	<i>Resultados</i>	157
4.2.1.	<i>Seção Introdução</i>	158
4.2.2.	<i>Seção Métodos</i>	159
4.2.3.	<i>Seção Resultados</i>	161
4.2.4.	<i>Seção Discussão</i>	163
4.3.	<i>Discussão sobre os resultados da Etapa 2</i>	163
5.	<i>Etapa 3: Associações como particularidades do texto especializado</i>	165
5.1.	<i>Metodologia</i>	166
5.2.	<i>Resultados e discussão</i>	167
6.	<i>Etapa 4: As associações sintagmáticas como unidades de tradução especializadas</i>	170
6.1.	<i>Metodologia</i>	171
6.2.	<i>Resultados e discussão</i>	171
6.3.	<i>Discussão sobre os achados</i>	175
7.	<i>Considerações sobre este capítulo</i>	175
CAPÍTULO VI		177
CONSIDERAÇÕES FINAIS – PARTE I.....		177
PARTE II: FERRAMENTA DE APOIO AO TRADUTOR		185
CAPÍTULO VII		186
LINGUÍSTICA COMPUTACIONAL E A TECNOLOGIA EM TRADUÇÃO.....		186
1.	<i>Linguística Computacional</i>	186
1.1.	<i>Recuperação de informações</i>	187
1.1.1.	<i>Recuperação de informações multilíngues</i>	188
1.1.2.	<i>Sistemas de recomendação</i>	189
1.2.	<i>Alguns trabalhos aplicados em Linguística Computacional</i>	190
2.	<i>A tecnologia e a tradução</i>	191
2.1.	<i>Tradução automática</i>	196
2.1.1.	<i>A EBMT (Example Based Machine Translation, tradução automática baseada em exemplos)</i>	200
2.2.	<i>CAT Tools</i>	202
2.2.1.	<i>Memórias de Tradução</i>	202
2.2.1.1.	<i>Memórias de tradução de segunda geração</i>	204
2.2.1.2.	<i>Memórias de tradução de terceira geração</i>	206
CAPÍTULO VIII		207
A FERRAMENTA.....		207
1.	<i>Construção dos bancos de dados</i>	208
2.	<i>A composição do protótipo</i>	212
3.	<i>Funcionamento do sistema</i>	212
4.	<i>Vantagens e limitações</i>	218
4.1.	<i>Limitações</i>	221
CAPÍTULO IX		223
CONSIDERAÇÕES FINAIS – PARTE II.....		223

CAPÍTULO X.....	226
CONCLUSÕES.....	226
REFERÊNCIAS.....	231
ANEXO I.....	240
LISTA DE FREQUÊNCIAS DA SEÇÃO INTRODUÇÃO	240
LISTA DE FREQUÊNCIAS DA SEÇÃO MÉTODOS	242
LISTA DE FREQUÊNCIAS DA SEÇÃO RESULTADOS.....	244
LISTA DE FREQUÊNCIAS DA SEÇÃO DISCUSSÃO	246
ANEXO II.....	248
LISTA DE PALAVRAS EXCLUÍDAS DA ANÁLISE COM O CONCGRAM (<i>STOPWORDS</i>)	248
ANEXO III.....	250
LISTAS DE CONCGRAMS DE 3 PALAVRAS NA SEÇÃO INTRODUÇÃO PARA A PALAVRA <i>CRIANÇAS</i>	250
LISTAS DE CONCGRAMS DE 3 PALAVRAS NA SEÇÃO MÉTODOS PARA A PALAVRA <i>ESTUDO</i>	251
LISTAS DE CONCGRAMS DE 3 PALAVRAS NA SEÇÃO RESULTADOS PARA A PALAVRA <i>PACIENTES</i>	253
LISTAS DE CONCGRAMS DE 3 PALAVRAS NA SEÇÃO DISCUSSÃO PARA A PALAVRA <i>CRIANÇAS</i>	256
ANEXO IV	258
LISTAS DE CONCGRAMS PARA A PALAVRA <i>OBJETIVO</i>	258

*Tudo que existe existe talvez porque outra coisa existe.
Nada é, tudo coexiste: talvez assim seja certo.*

Fernando Pessoa

INTRODUÇÃO

O desenvolvimento de tecnologias e de recursos computacionais, sobretudo nos últimos 20 anos, trouxe viço, agitação e um certo ar de modernidade à área da tradução técnico-científica, muitas vezes vista como “mecânica, fria, unívoca, sem alternativa”, em oposição à tradução literária, “orgânica, quente, maleável, com alternativas”¹. Em se tratando de traduções que envolvem a língua inglesa, a informatização afetou não apenas o volume de conteúdo a ser traduzido (manuais, produtos de software, jogos, conteúdo de Internet), que se multiplicou exponencialmente, como também trouxe novos recursos de auxílio para o tradutor. Hoje, trabalha-se, em grande parte, com conteúdos apresentados em arquivos eletrônicos, o que possibilita aos profissionais da tradução contar com programas que armazenam traduções em memórias eletrônicas, que constituem uma área denominada de “tradução assistida por computador”. As fontes de consulta também se multiplicaram e se modificaram, e a “velha e boa” consulta ao dicionário de papel vem sendo substituída por pesquisas em textos publicados na Internet, em bancos de dados e em dicionários eletrônicos, além dos fóruns de tradutores em portais específicos.

A informatização também originou uma nova área de estudos, a da Tradução Automática, que está na origem do desenvolvimento de trabalhos em Processamento da Linguagem Natural. Esse novo contexto também permitiu novas maneiras de se pesquisar sobre o processo de traduzir e sobre a configuração dos textos traduzidos, com ferramentas que permitem processar automaticamente uma grande quantidade de textos e outras que ajudam a descrever os processos mentais e cognitivos que ocorrem durante a realização de uma tradução.

¹ Opinião do professor Elvio Funk publicada no jornal Zero Hora de 01/11/2003, Caderno de Cultura. Professor aposentado do PPG-Letras e do Departamento de Línguas Modernas do Instituto de Letras da UFRGS.

Mas é preciso dizer que, em meio a todo esse desenvolvimento da tecnologia em/para tradução, a contribuição do conhecimento humano é fundamental em cada uma dessas áreas mencionadas. Aqui nos referimos tanto às fases de desenvolvimento de sistemas automáticos para os fins citados, como ao trabalho ou à seleção do resultado de seus processamentos. Alguns sistemas de tradução automática, por exemplo, apresentam resultados satisfatórios para fins de leitura e de entendimento de textos. Já outras áreas que envolvem tradução e computador necessitam, em muito, a colaboração humana. Entre essas, citamos, por exemplo, os sistemas que extraem, automaticamente, de textos, palavras e sintagmas que podem vir a ser considerados termos técnicos, com vistas principalmente à construção de dicionários e glossários. As traduções automáticas, quando destinadas a fins mais refinados, como, por exemplo, quando os textos serão publicados, também precisam passar por revisões humanas.

É em meio a esse cenário, em que há uma associação entre tradução, processamento automático de textos e tecnologia, com a contribuição e a interferência humana, que desenvolvemos esta tese. Nela enfocamos, em um primeiro momento, uma faceta do texto especializado, mais especificamente de artigos da área médica: as associações sintagmáticas entre unidades lexicais que, por motivos que serão evidenciados, são relevantes para a tradução. O critério de relevância está embasado na crença de que o tradutor deve estar atento a essas unidades, uma vez que elas representam riscos, caso ele não lide com elas adequadamente, e apresentam vantagens, pois se ele as conhece, tem maiores chances de obter êxito em seu trabalho. O êxito consiste em elaborar um texto que cumpra o seu propósito e que corresponda às expectativas da audiência que os recebe. Em um segundo momento, voltados para uma aplicação prática, propomos um sistema de apoio e de referência para a tradução de artigos médicos, com base nas associações que são recorrentes nesses textos. Utilizamos um *cópus*² de artigos médicos de onde extraímos dados para comprovar nossa hipótese, nossos pressupostos e para embasar o desenvolvimento do referido sistema.

A seguir, prosseguimos com este capítulo introdutório, apresentando nossos objetivos, as razões que motivam a pesquisa, o posicionamento teórico do trabalho e as questões a serem investigadas.

² Optamos por essa grafia por simpatia à sugestão do Prof. Francisco da Silva Borba, que sugere que se use a mesma forma para o plural e o singular, seguindo a lógica usada para a palavra *lápiz*.

1. Objetivos

Conforme já mencionado nesta introdução, esta tese contempla objetivos de cunhos teóricos e aplicados. O objetivo aplicado se constitui como o resultado e a aplicação das crenças, dos pressupostos e da metodologia que delineamos, em termos mais teóricos, na primeira porção da tese.

Num primeiro momento, queremos demonstrar que determinadas associações frequentes entre as palavras³, com determinadas características, podem se configurar como Unidades de Tradução Especializadas. Essa crença tem origem na observação do texto especializado feita durante a prática da tradução. Por isso, também queremos demonstrar que a visão do texto especializado sob a ótica do tradutor pode trazer contribuições interessantes para os estudos de texto especializado e também para áreas que desenvolvem materiais de referência e de apoio para tradutores, como a Terminologia ou a Linguística Computacional.

Ao mesmo tempo em que construímos a porção teórica da tese, estaremos também construindo as bases para nossa aplicação prática. O objetivo do sistema pretendido é que ele apresente ao tradutor, a partir do texto em português a ser traduzido, exemplos de trechos de textos em inglês com associações sintagmáticas semelhantes às associações que ele deverá traduzir, as quais serão provenientes de um corpus paralelo. Com essa ferramenta, pretendemos prestar auxílio principalmente a tradutores iniciantes e aprendizes. Não queremos que essa ferramenta faça apenas uma comparação de similaridades de segmentos do texto-fonte e do texto-alvo, que é o que, muito genericamente, está na base de um sistema de Memória de Tradução. Nosso sistema terá uma base de dados composta pelas associações que identificamos e descrevemos na tese. Assim, entendemos que estaremos acrescentando ao sistema um “conhecimento” que um tradutor iniciante não tem, uma vez que não tem experiência suficiente para reconhecer quais associações são mais frequentes ou mais típicas em determinadas áreas. Saber fazer essas associações está relacionado, entre outros fatores, a saber associar, na cadeia sintagmática da língua de chegada, unidades lexicais⁴ que coocorrem nos textos de chegada. Pode-se considerar que esse fator estaria relacionado à competência tradutória. De tal modo, a ferramenta tem o intuito de oferecer ao tradutor um acesso mais

³ Cabe salientar que não estamos pensando em uma relação sequencial de contiguidade vizinha (como em sintagmas – terminológicos ou não – ou n-gramas, que são seqüências de n palavras pertencentes a um texto” e fraseologias)

⁴ Nos referimos a unidades lexicais, considerando que termos e palavras gramaticais estão incluídas neste conceito.

rápido e direcionado a um conhecimento que ele, provavelmente, adquiriria após algum bom tempo de experiência e/ou de pesquisa com textos de uma dada especialidade.

2. Motivação

O principal motivo que nos levou a realizar esta pesquisa foi o desejo de poder contribuir, ainda que com limitações, para a geração de materiais de consulta para o tradutor, compostos tais materiais por itens que realmente auxiliem a prática da tradução, principalmente a assistida por recursos computacionais. Além disso, outro grande ponto de interesse é a discussão sobre o que faz um texto técnico-científico ser diferente de um texto da linguagem cotidiana, não-científica.

O interesse pela pesquisa, no entanto, surgiu de uma experiência profissional de cerca de um ano, quando tivemos contato com a empresa que gerencia o processo de tradução e editoração dos artigos do *Jornal de Pediatria*. Como os artigos são, em sua maioria, traduzidos do português para o inglês, percebeu-se que a empresa acabava tendo a sua disposição um grande *cópus* paralelo de textos em português, traduzidos para o inglês, o que, costuma-se dizer, é uma “raridade”. Tínhamos a sensação de que, de alguma maneira, o esforço realizado pelos tradutores deveria ser aproveitado de alguma forma. Além disso, o contato com a tradução de artigos médicos também suscitava questões, como a que acabamos investigando, acerca de palavras que costumam “andar juntas” nesses textos e a necessidade de reconhecê-las e de saber como traduzi-las.

Conforme indica Coulthard (2005), uma das primeiras questões que tornam a tradução de textos médicos de e para o inglês muito particular é que a Medicina Ocidental é muito mais antiga do que a língua inglesa, com um vocabulário formado por palavras de origem grega e latina. Além do vocabulário, as formas de construção das frases são bastante particulares em cada uma das línguas em questão. Já Lee-Jahnke (2005) aponta que a tradução médica é de grande importância no âmbito mundial da tradução, pois um grande número de textos são traduzidos na área e, portanto, representam uma boa fatia do mercado para tradutores.

Também nos motivou o desejo de nos aproximarmos mais de textos de natureza acadêmica que precisam ser traduzidos no sentido português-inglês. Em geral, esses textos são produzidos por pesquisadores brasileiros que desejam ter sua produção intelectual divulgada na forma de artigos científicos, em sua maioria, em publicações do tipo revista acadêmica ou em anais de eventos, mas que muitas vezes não têm conhecimento necessário

em língua inglesa para alcançar o nível de exigência dessas revistas. Publicações em periódicos internacionais de qualidade são um trunfo para qualquer pesquisador, tanto pelo alcance da divulgação das pesquisas como pelo peso que têm em processos de avaliação, que podem ser decisivos para a continuidade dos estudos. Estudantes e pesquisadores, então, acabam encontrando como obstáculos para a realização profissional, não a qualidade de seus estudos, mas a dificuldade de escrever e publicar um texto, em sua língua materna e, para muitos ainda, em língua estrangeira, especialmente em língua inglesa.

Todo esse contexto favorece o mercado de trabalho para tradutores. O que acontece é que, muitas vezes, esses profissionais também não têm formação para trabalhar com o texto acadêmico-científico e nem encontram oportunidades de qualificação. Deduzimos que acabam aprendendo com a experiência, com os próprios erros ou, até mesmo, realizando trabalhos que ficam aquém da qualidade esperada, dadas as especificidades de cada trabalho, área, periódico, idioma.

Em um depoimento de um tradutor já experiente em um fórum de discussão de tradutores⁵, ele comenta

Foi por isso que parei de traduzir Medicina, apesar da insistência de um cliente. Eu cobrava o meu preço normal, mas perdia TANTO tempo em pesquisa, que não valia a pena. Além disso, eu tinha certeza de que o meu serviço dava um trabalhão ao revisor do cliente final, porque eu não sabia direito o que estava fazendo. J.H.L. 27/02/2008.

Acreditamos que esse depoimento representa o sentimento de muitos tradutores, mesmo o dos que têm experiência, como é o caso do tradutor do relato acima, de que a tradução do texto de Medicina exige muita pesquisa e que, ainda assim, o trabalho de tradução pode não ficar bom. Assim, para alguns seria melhor nem traduzir. Por outro lado, acreditamos que, se pudermos ajudar o tradutor, ainda em início de carreira ou em formação, com um sistema que lhe apresente sugestões de tradução para formações textuais típicas da linguagem médica, ele poderá ter mais chances de desenvolver um conhecimento necessário para traduzir textos nessa área. Poderá, posteriormente, atualizar-se, qualificar-se e tornar-se um bom tradutor de Medicina, ou ao menos ter tido a possibilidade de ter algum contato com essa área. Vale lembrar que cursos de especialização para tradutores (e mesmo de graduação), que os ajudem a se especializarem na tradução de áreas específicas de conhecimento (ex.:

⁵ A lista de discussões Trad-prt do Yahoo Groups.

Mecânica, Farmácia, Direito, Economia, Medicina, entre outras) são raríssimos, até onde sabemos.

Vale aqui também lembrar que uma das primeiras ideias que tivemos a respeito de nossa hipótese de estudo veio de uma reflexão sobre como é empregada a palavra “criança” nos textos de Pediatria. Seguiu-se um raciocínio já conhecido, o de que, nessa área, uma palavra que é da “língua de uso cotidiano” “atualiza-se” como um termo. Mas, ela não passa a adquirir um sentido totalmente diferente quando a pensamos no domínio da Pediatria, portanto não é a mudança de um conceito ou de um significado que a “atualiza” como “especial” dentro do contexto da Pediatria. Pensamos, então, que uma das maneiras de se acessar o contexto de seu uso é olhar para a frase em que a palavra se encontra e para o texto como um todo. Assim, notamos que, em Pediatria, “criança” se associa, frequentemente, a determinadas unidades do léxico, enquanto em outros contextos ela, provavelmente, se associa a outras ou a nenhuma. Acreditamos que, a partir das diferentes associações que vão se construindo dentro de uma área, vai se consolidando a linguagem do domínio especializado. O fato de “criança”, como veremos, estar associada a palavras como “idade”, “anos”, “faixa etária” aponta para um texto que descreve crianças e faz isso dentro de um padrão sintático que pode variar.

Unindo, portanto, os fatos de que i) há essa necessidade de profissionais do texto para trabalhar com artigos científicos; ii) os profissionais carecem de uma orientação especializada no início da carreira; iii) a Informática vem, cada vez mais, fazendo parte da vida do tradutor; e iv) os recursos computacionais são capazes de armazenar e de disponibilizar conhecimento especializado, sem precisarem estar vinculados a um padrão rígido de constituição (por ex.: incluindo apenas palavras associadas a conceitos), motivamo-nos a realizar esta tese de doutorado.

3. Questões de pesquisa e hipótese

Já antecipamos nos Objetivos que pensamos que a observação do texto especializado feita sob o ponto de vista do tradutor é diferenciada daquela feita por outros estudiosos da linguagem, tais como linguistas ou terminólogos, dadas as suas necessidades práticas. Acreditamos que essa perspectiva pode produzir resultados relevantes para aplicações práticas, bem como embasar estudos de natureza descritiva e teórica. Isso leva o tradutor a “recortar” unidades textuais relevantes de maneira peculiar. Pensamos que esse seu modo de

abordar o texto é motivado, entre outros fatores, por uma necessidade intrínseca ao processo de tradução, a “reverbalização do conteúdo”. Ela levaria o tradutor a considerar todas as unidades que compõem o texto, a princípio, em igual nível de importância, sejam elas palavras gramaticais, termos técnicos, números, símbolos, entre outros. Isso ocorre porque ao tradutor interessa um texto alvo bem construído em sua totalidade, e não há saída se não abordar cada um desses itens, encadeando-os no eixo sintagmático e fazendo as escolhas do eixo paradigmático, para que sejam fluentes e apropriados ao discurso em questão.

Outro pressuposto que permeia esse trabalho é o de que, assim como na linguagem cotidiana, também nos domínios especializados o significado do todo é construído a partir da relação sintagmática entre as palavras.

Essas crenças, nossos objetivos e o entendimento de que associações sintagmáticas são relevantes para a tradução dos textos devido a fatores como prototipicidade, risco e vantagem, levam-nos a ancorar o desenvolvimento da tese na seguinte hipótese:

Existem, nos textos especializados, associações no nível sintagmático que:

- a) são combinações mais ou menos fixas e frequentes de palavras;**
- b) não são necessariamente contínuas;**
- c) estão associadas no nível sintagmático por uma atração semântica e/ou pragmática e/ou gramatical;**
- d) não necessariamente contêm um termo em sua composição, sendo termo entendido aqui como uma expressão lexical associada a um conceito.**

Essas associações se configuram como Unidades de Tradução Especializadas porque são recorrentes, fazem parte dos “modos de dizer” do gêneros e do domínio em questão e apresentam riscos e vantagens ao tradutor.

Dada essa hipótese e seus desdobramentos, as questões de pesquisa que se colocam são:

- 1) É possível identificar e descrever, em larga escala e com base em cópulas, associações sintagmáticas com o perfil acima?

- 2) Uma vez identificadas tais associações, é possível aproveitá-las em uma ferramenta computacional que as armazene e as mostre para o tradutor na situação de uma tradução nova, trazendo também exemplos de como já foram traduzidas?

Essas duas questões orientam a organização desta tese em duas partes: a primeira será tratada na Parte I e a segunda na Parte II.

4. Objeto de estudo e unidade de análise

Como não poderia deixar de ser, citando a célebre frase de que “o ponto de vista cria o objeto”, como ensinou Saussure (1993 [1915]), nosso *objeto de estudo* é o texto especializado de artigos médicos visto pelo ponto de vista do tradutor. Entendemos também que, no objeto criado por esse ponto de vista, destacam-se componentes do texto especializado que adquirem relevância pela sua frequência, prototipicidade, risco e vantagem para a tradução. Esses critérios são inspirados no trabalho de Gouadec (1994) e serão devidamente explicitados na tese. Neste trabalho, consideramos como objeto de estudo o “todo” no qual se incluem as unidades de análise, por isso o texto especializado foi selecionado como objeto de estudo.

Dentre esses itens que essa abordagem do texto via tradutor apresenta, destacamos como nossa unidade de análise *associações sintagmáticas* que são recortadas como unidades de tradução. Cabe o destaque porque consideramos que elas são frequentes e típicas dos artigos médicos em Pediatria, tal como será demonstrado mais adiante. Desconhecê-las representa um risco para o tradutor e, conseqüentemente, conhecê-las é uma vantagem. Isso implica considerar que, em meio ao todo de associações sintagmáticas possíveis na língua ou em um texto a ser traduzido, por exemplo, algumas podem ser destacadas e caracterizadas a partir de uma perspectiva do tradutor.

Antes de prosseguir, abrindo um parêntese amplo, é preciso esclarecer ainda que consideramos que, ao termos em mente, neste trabalho, o *ponto de vista* do tradutor, não caberia destacar um elemento do texto em detrimento de outros, como, por exemplo, as associações sintagmáticas. O que entendemos, a esse respeito, é que as associações que enfocamos são uma parte desse texto que é nosso objeto de estudo e que é objeto de trabalho do tradutor. Os recortes são necessários por questões metodológicas e para isolar um dos elementos do texto especializado. Não se trata de menosprezar outros elementos. É preciso dizer também que o recorte que estamos fazendo privilegia muito a ocorrência natural desse

fenômeno das associações, sem com isso fazermos recortes e definirmos *a priori* que padrão (sintático, composicional ou estrutural) de associação se configuraria como nossa unidade de análise.

Cabe, ainda, fazer um outro esclarecimento antes de prosseguir. Este não é um estudo que se propõe a analisar e a descrever as características de textos traduzidos frente a seus originais, apesar de contarmos com um *cópus* de textos originais e suas traduções. Nosso interesse, antes disso, é contribuir para a realização de traduções e oferecer um recurso de referência que se pretende bastante diferente de dicionários e glossários usuais. Estamos, portanto, voltados para o produto tradução. Porém, a justificativa para o foco em associações sintagmáticas se dá por fatores intrínsecos ao processo e ao produto que se quer atingir, como explicitaremos no Capítulo II.

5. Posicionamento do trabalho e embasamento teórico

Esta tese pertence ao campo da Linguística Aplicada (LA) no sentido de que procuramos relacionar teorias da linguagem a outros campos com o propósito de trazer novos dados ao próprio campo da Linguística e a outras áreas – aplicadas – como, neste caso, a tradução. Conforme explicita Gledhill (2000):

O trabalho em linguística aplicada tende a abordar a linguagem contemporânea... e a possibilitar que modelos linguísticos surjam por meio da discussão dos dados e não por meio da apresentação de um modelo como o principal objeto de investigação⁶ (p.5).

Sendo um trabalho em e de LA, promovemos um encontro de áreas como: Linguística de *Cópus*, Tradução, Estudo das linguagens especializadas. O nosso trabalho tem grande influência das pesquisas em Linguística Aplicada realizadas em universidades britânicas, especialmente da escola de Birmingham. Segundo o autor supra citado, os interesses compartilhados por pesquisas em Linguística de *Cópus* são:

1. Um interesse pelo discurso (entendido como a linguagem em ação e em relação a seus usuários).
2. Uma ênfase na estreita relação entre vocabulário e gramática.
3. Uma preocupação com dados autênticos e não inventados.

⁶ Esta e outras citações originalmente em língua inglesa foram traduzidas por mim.

4. Uma preferência pelo uso de computadores para a análise de grandes arquivos com dados linguísticos.

Esses quatro interesses convergem, na maior parte das vezes, para estudos que se enquadram na Linguística de Córpus. Como o nome já indica, trata-se de uma área⁷ em que se compilam e se utilizam grandes coleções de textos (córpus), para evidenciar ou buscar características dos mesmos; seus padrões e frequências são identificados através de processamentos automáticos. Na Linguística de Córpus, diz-se que a linguagem é padronizada e que a padronização se evidencia pela recorrência “isto é, uma colocação, coligação ou estrutura que se repete significativamente mostra sinais de ser, na verdade, um padrão lexical ou léxico-gramatical” (Berber-Sardinha, 2004, p. 31). Nessa abordagem, a “linguagem forma padrões que apresentam regularidade (estáveis em momentos distintos...) e variação sistemática (correlacionam-se com variedades textuais, genéricas, dialetais etc.)” (Berber-Sardinha, 2004, p. 31). Esses padrões de associações regulares, o emprego de um grande número de expressões pré-fabricadas e a união dessas em sequências maiores já foram reconhecidos por serem os responsáveis por conferir fluência e naturalidade a falantes nativos. Sinclair (1987) concretizou essa ideia em um princípio de entendimento da linguagem, o *princípio idiomático*. Voltando a esta tese, portanto, se entendermos que, dadas as circunstâncias e objetivos, uma tradução deve ser fluente em uma segunda língua, certamente ela deverá satisfazer tais padrões e regularidades. Os métodos, os pressupostos e os resultados das pesquisas em Linguística de Córpus são potencialmente úteis para a tradução nesse sentido e, por isso, nossa metodologia se espelha em estudos e em métodos da referida área.

Segundo Sinclair (*op. cit*), léxico e sintaxe se encontram em um único espaço, pois a escolha de uma categoria gramatical ou de um item lexical reduz as possibilidades de escolha tanto gramaticais quanto lexicais. Na linguística sistêmico-funcional de Michael Halliday, que apresenta uma visão da linguagem muito parecida com a de Sinclair e de seus seguidores, esse espaço é chamado de léxico-gramática. A Gramática Sistêmico Funcional (GSF) descreve as estruturas gramaticais da língua com referência a seu significado ou funcionalidade comunicativa, por isso diz-se que é um teoria da língua tanto funcional como semântica. Neste trabalho, fazemos uso de algumas terminologias usadas na GSF. Esse modelo se preocupa em descrever padrões ou estruturas das combinações de palavras que ocorrem no contexto de funções ou propósitos comunicativos. Essas funções são referidas como

⁷ Há alguma controvérsia sobre o fato de a Linguística de Córpus ser uma área. Há linguistas que a consideram apenas como um conjunto de métodos, opinião da qual discordamos.

significados, usando este termo em um sentido mais geral do que o usado no senso comum. Há três tipos amplos de significado: ideacional, interpessoal e textual. Na Gramática Sistêmico Funcional, os conceitos de Processo, Participante e Circunstância são categorias semânticas que explicam de modo mais geral como fenômenos de nossa experiência do mundo são construídos na estrutura linguística (White, 2009).

Em termos gerais, a respeito de tradução, conforme classificação exposta em Hurtado Albir (2001, p. 94), estamos tratando de traduções:

- profissionais, no que concerne à natureza do processo tradutório em relação ao indivíduo;
- técnicas, no que concerne ao âmbito socioprofissional (em oposição à tradução administrativa, religiosa, literária etc.); e
- escritas, no que concerne à forma (feitas da língua fonte português para a língua alvo inglês).

Além disso, nosso trabalho enfoca especificamente a tradução da língua portuguesa para a língua inglesa, também conhecido como *versão*. Em termos de processo tradutório, não encontramos estudos que comprovem diferenças entre a tradução da língua materna para a língua estrangeira ou a direção contrária. Acreditamos que essa seja, no Brasil, uma distinção em termos de mercado e valores, já que geralmente a tradução para o inglês (ou outra língua estrangeira) é mais cara do que para o português.

Quanto ao processo de tradução, este estudo compartilha da representação “didática” do processo de tradução, apresentada pela Teoria Interpretativa da Tradução no trabalho de Danica Seleskovitch e Marianne Lederer (1994). Essa abordagem toma o processo de tradução como uma sucessão de três estágios:

1. Percepção de um enunciado linguístico que é portador de significado;
2. Abandono imediato e intencional das palavras e retenção da representação mental da mensagem (conceitos, ideias etc.); e
3. Produção de um novo enunciado na língua-alvo.

Nossa tese se volta para o momento de reexpressão ou reverbalização do conteúdo: quando são refeitas as associações sintagmáticas na língua de chegada.

Para justificar o critério de relevância de determinadas associações sintagmáticas, estamos embasados no trabalho de Daniel Gouadec (1994). A perspectiva desse autor ao tratar do tema “fraseologia” mostra-se como uma visão aberta, que não se limita a considerar cadeias de caracteres segundo critérios de tamanho, constituição sintática ou morfológica. O autor trabalha com uma ideia bastante ampla de cadeias de caracteres especializadas. Essas cadeias têm tamanho variável, “incluem palavras, grupos de palavras, termos, locuções, expressões, orações, segmentos de frases, frases, conjuntos de frases” (*op cit*). Tal visão parece ser a ideal para um estudo como este, que i) considera que associações lexicais sem um padrão fixo compõem um tipo de evento linguístico relevante para a tradução, e ii) visa a uma aplicação destinada a tradutores, que seriam incluídos no que o autor chama de *langagier* ou especialista em linguagem.

Neste trabalho também fazemos considerações acerca do caráter especializado das associações e dos textos. Nosso intuito é o de tratar dessas considerações na discussão ao final do trabalho. Não pretendemos elaborar uma revisão bibliográfica extensa sobre a literatura acerca do texto especializado, mas cabe colocar, desde o início, que entendemos o texto especializado como uma realização da língua, nos moldes do que considera Maciel (2001)

todos os usuários da língua, leigos ou especialistas, têm à disposição o mesmo repertório fonológico, morfológico ou sintático. Cada um faz a escolha que melhor serve a seus propósitos nas circunstâncias em que se realiza a comunicação, ainda que utilize o mesmo manancial linguístico (p.87).

Tais escolhas, que o usuário das linguagens técnicas ou científicas faz, acabam revelando tendências e similaridades, determinadas pelo contexto de realização que, por sua vez, determinam os gêneros, os tipos de textos, etc. Em um sentido de raciocínio mais “direto”, costuma-se pensar que o que faz um texto especializado constituir-se como tal é a presença de termos técnicos, mas, como explica Hoffmann (2004), desde os anos 80:

a especificidade das linguagens especializadas, em relação à linguagem comum e a outras sublinguagens, se expressa mais claramente pelo léxico, quer dizer, pelo vocabulário especializado ou pela terminologia, mas também pelo uso de determinadas categorias gramaticais, de construções sintáticas e de outras estruturas textuais (p. 81).

Em nossa trajetória de estudos, primeiro com o mestrado (Possamai, 2004), em que abordamos marcadores textuais e discursivos dos artigos científicos, e agora com o doutorado,

viemos, constantemente, abordando essas “outras estruturas textuais”, pois elas são importantíssimas para o tradutor e ainda parece muito difícil sistematizá-las e encontrá-las em materiais de referência. Acreditamos que enfatizar essas “outras estruturas” pode contribuir com o desenvolvimento de materiais de referência mais úteis para tradutores.

Além disso, pensamos que esse enfoque via tradução pode “jogar” uma luz diferente ao texto especializado, que refletirá mais a riqueza de sua constituição, olhando para itens que, a princípio, não parecem estar diretamente voltados ao caráter especializado do texto. Para dar um exemplo, voltamos ao nosso questionamento quanto à “especialidade” da palavra “crianças” que mencionamos anteriormente. Comparando frases como a) *Em um estudo transversal foram avaliadas 15 crianças de 0 a 3 meses de idade* e b) *No Clínicas são 21 crianças para 20 vagas*⁸, podemos notar que, mesmo com poucas palavras, são segmentos de diferentes textos, com graus de especialidade diferentes. Nesse exemplo hipotético, notamos que o tempo verbal usado no exemplo “b” (... *são 21 crianças*) não parece ser um tempo verbal que encontraríamos em um ambiente de artigo médico. Portanto, nota-se, nesse caso, que a presença de palavras mais técnicas, associadas a conceitos, nada tem a ver com o caráter mais “especializado” do texto, pois tanto *crianças* assume valores diferentes, como a construção da frase que dá “tons” diferentes ao texto. Nesse ponto, há também a questão da definição/construção de um significado, e aqui temos o apoio do ponto de vista desenvolvido por Firth (1957) que com muita ênfase buscou evidenciar a construção do significado por meio das relações sintagmáticas entre as palavras.

No que concerne ao desenho da aplicação prática, investigamos algumas áreas como Recuperação de Informações, Cross-Sectional Information Retrieval, Sistemas de Recomendação, Sistemas de Tradução Automática Baseados em Exemplos e Memórias de Tradução. No entanto, por não sermos linguistas computacionais, nossa contribuição foi o levantamento dos dados a serem inseridos na ferramenta que foi desenvolvida com base na ideia de como um sistema poderia fazer uso desses dados, mais ou menos assemelhando-se ao que existe nos sistemas de Memórias de Tradução.

Para alcançar nossos objetivos e discorrer sobre esses temas que mencionamos na introdução, nosso trabalho é dividido em duas partes principais, conforme explica a seção a seguir.

⁸ A primeira frase foi extraída do corpus do Jornal de Pediatria e a segunda do corpus do jornal popular Diário Gaúcho, cujas ferramentas de análise estão disponíveis em <http://www6.ufrgs.br/textecc>.

6. Organização do trabalho

Como dissemos, este trabalho está dividido em duas partes, que podem ser divididas conforme as duas questões de pesquisa que nos propusemos a investigar. A primeira investigando as associações sintagmáticas, caracterizando-as como unidades de tradução; a segunda aplicando essa hipótese a uma ferramenta de apoio ao tradutor.

Sendo assim, a parte I deste trabalho é composta de seis capítulos. O Capítulo I faz uma abordagem das áreas afins ao estudo das associações no nível sintagmático, o que nos ajuda a definir as unidades que analisaremos. O Capítulo II trata de aproximar os estudos sobre o processo de tradução ao nosso trabalho. Detém-se na teoria em que embasamos nossa discussão, a Teoria Interpretativa da Tradução, e delinea, então, como as unidades definidas no Capítulo anterior podem configurar-se como Unidades de Tradução Especializadas. O Capítulo III faz uma aproximação entre o tema das associações sintagmáticas e a tradução, mostrando também a relevância destas no texto traduzido. O capítulo IV descreve estudos-piloto desenvolvidos para responder a nossa primeira questão de pesquisa, faz questionamentos e apresenta dados importantes para a construção de nossa argumentação. No Capítulo V, descrevemos como investigamos nossa hipótese e pressupostos, bem como mostramos os resultados das análises. Já no Capítulo VI há um fechamento da Parte I e uma indicação de continuidade da pesquisa no desenvolvimento de uma aplicação prática. A parte II deste trabalho é composta de três capítulos: VII, VIII e IX. O Capítulo VII, que faz uma introdução à Linguística Computacional, em que citamos alguns trabalhos desenvolvidos na área e enfocamos as subáreas que têm uma interface com a tradução. O Capítulo VIII explica como foram selecionadas as associações que foram incluídas no aplicativo desenvolvido e como a ferramenta trabalha. Por fim, o Capítulo IX faz considerações finais sobre a parte II e o Capítulo X apresenta as conclusões do trabalho, em que discutimos as implicações do ponto de vista tomado, nossas conclusões gerais e perspectivas.

PARTE I

**ASSOCIAÇÕES SINTAGMÁTICAS COMO UNIDADES DE TRADUÇÃO
ESPECIALIZADAS**

CAPÍTULO I

ENFOCANDO AS UNIDADES DE ANÁLISE E CONCEITOS RELACIONADOS

Neste Capítulo, objetivamos revisar alguns conceitos relacionados ao tema das associações sintagmáticas. Trazemos quase que exclusivamente conceitos de tradição britânica, pois é com esses estudos que nossa pesquisa mais se identifica, embora eles sejam, geralmente, voltados para a chamada língua geral. Com essa revisão, procuramos buscar subsídios para identificar e descrever nossas unidades de análise, um tópico que discutimos ao final das revisões neste Capítulo.

Ao tomar como unidade de análise as associações no nível sintagmático, evocamos um dos eixos da linguagem estabelecidos no Curso de Linguística Geral (Saussure, 1993): o eixo das relações sequenciais em que cada item se define por oposição aos outros itens. É o eixo das relações presentes (*in presentia*) ou *eixo sintagmático*. Nosso foco recai sobre essas relações porque elas são fundamentais na construção de como o conteúdo de diferentes áreas de conhecimento é apresentado textualmente, levando em conta gêneros e audiências específicas. As relações *in presentia* são aquelas selecionadas pelo falante, por isso, por exemplo, uma linha de concordância⁹ que apresenta o contexto de uma determinada palavra representa as relações do eixo sintagmático. Considerando que devido a diversos fatores algumas associações acabam se tornando relativamente frequentes e muitas vezes apresentam algumas particularidades em sua construção sintática e léxica, tanto em uma língua fonte quanto quando as comparamos entre diferentes idiomas, temos em mente que o tradutor é

⁹ As linhas de concordância, em Linguística de Córpus, são resultados de buscas realizadas com as ferramentas de concordância, ou concordanciadores, que apresentam o contexto de uma determinada palavra nodo, que por sua vez aparece centralizada na linha, facilitando a observação de padrões e contextos imediatos.

beneficiado quando as conhece e/ou quando tem à sua disposição recursos ou exemplos que o auxiliem a traduzi-las.

Para orientar a leitura deste Capítulo e o entendimento do fenômeno linguístico que abordamos como nossa unidade de análise, introduzimos aqui três extratos de diferentes artigos científicos médicos de nosso *cópus*¹⁰ em que se percebe uma regularidade e uma frequência de coocorrência de determinadas palavras, como: *variáveis*, *comparadas/comparar/comparação*, *teste t de Student*, *amostras independentes*, *grupos*, *médios/média*.

Texto A

As **variáveis** quantitativas foram **comparadas** pelo **teste t de Student** para **amostras independentes** e, na situação de três ou mais **grupos**, pela análise de variância (ANOVA), com localização de diferenças pelo teste de post-hoc de Duncan.

Texto B

Utilizou-se o **teste t de Student** para **amostras independentes**, para **comparar** os resultados **médios** das **variáveis** analisadas.

Texto C

Foi realizada a **comparação** entre **grupos** pelo **teste t de Student** para **amostras independentes** com estimativa de diferença **média** e intervalo de confiança (IC) de 95%.

Além de percebermos essa regularidade nas coocorrências, vemos que também há uma certa flexibilidade entre essas associações, tanto em sua construção sintática como na ordem dos itens que compõem a sentença. Além disso, a distância entre uma palavra nodo¹¹ e seu(s) colocado(s) também é variável. Ainda, percebe-se que essas palavras destacadas estão em um mesmo “ambiente”, na mesma frase, mas elas não necessariamente estão ligadas como se fossem colocações. Por exemplo, podemos dizer que há uma coligação¹² entre o verbo *comparar* + *amostras independentes*, mas não se poderia dizer que há o mesmo tipo de relação entre *grupos* + *amostras independentes*.

É a partir desse cenário que parte nossa primeira questão de pesquisa:

¹⁰ Uma descrição completa do *cópus* é feita no Capítulo V.

¹¹ A palavra nodo é a que selecionamos para observação. No caso dos exemplos acima, qualquer uma das coloridas poderia ser considerada palavra nodo.

¹² Diferentes abordagens ao tema das colocações são abordadas na sequência do capítulo.

É possível identificar e descrever, em larga escala e com base em córpus, associações sintagmáticas com um determinado perfil e que se constituam em unidades de tradução?

Para responder essa questão, precisamos antes entender que tipo de associação é essa que opera no nível sintagmático. A leitura e a seleção dos conceitos que trazemos aqui não são feitas com o intuito de compor uma revisão bibliográfica extensa da literatura a respeito; apresentamos o que revisamos para chegar a um entendimento e definição das relações e associações que abordamos nesta tese. Dessa forma, acreditamos que é possível fazer uma revisão mais centrada nos aspectos mais pertinentes ao trabalho, em uma tentativa de não nos estendermos demais em classificações ou explicações a respeito das sobreposições dos conceitos, tão comuns nessa área de estudo.

Apresentamos, inicialmente, o conceito de *colocação* em diferentes perspectivas sob as quais esse fenômeno é abordado. A seguir, descrevemos outros conceitos relacionados, como o *princípio idiomático*, os *lexical primings*, as associações semânticas e as fraseologias. Também explicamos um conceito mais recente, o de *congrams*. Após esse panorama a respeito desses conceitos, definimos onde se situam nossas unidades de análise nesse cenário e como elas se conformam.

1. Colocações

Dentro do desenvolvimento histórico da linguística, W. Porzig é considerado o primeiro a chamar a atenção, em 1934, para o fato de que uma combinação sintagmática não tinha relação somente com características gramaticais, mas também com aspectos de significado. O autor considerava que “todas as palavras têm um campo magnético ao seu redor, no qual podem penetrar somente palavras de um tipo muito específico” (Porzig, 1970, p. 133 *apud* Beneduzzi, 2008).

Antes disso, Saussure, no Curso de Linguística Geral, vislumbrou o eixo sintagmático das associações, porém o enfoque de seu trabalho acabou privilegiando as relações do eixo paradigmático (Beneduzzi, 2008). Mas, foi apenas nos anos 50 e 60 que o conceito de *colocação* apareceu em outras abordagens teóricas. Hoje, as colocações continuam a despertar o interesse de pesquisadores, principalmente com o objetivo de elaborar novos modelos de material lexicográfico e terminográfico ou para compreender melhor os padrões das diferentes

línguas para fins de ensino de língua estrangeira e tradução, isso tudo com muito apoio dos métodos e ferramentas empregados pela Linguística de Córpus.

Manning e Schütze (1999), autores referência da área de Processamento da Linguagem Natural também especificam aplicações para os resultados de estudos sobre colocações, como a geração automática de eventos da língua natural (para que o resultado soe natural), a lexicografia computacional (para identificar automaticamente as colocações importantes de serem listadas em uma entrada de dicionário), o parseamento (para que seja dada preferência a análises com colocações naturais) e a pesquisa em Linguística de Córpus.

Para o tradutor, reconhecer as colocações e saber aplicá-las adequadamente ao seu trabalho resulta na produção de textos mais naturais e menos “ingênuos”¹³, diminuindo o estranhamento por parte do leitor. Por isso, estudos que reconheçam colocações e produzam resultados de aplicação prática são de auxílio para tradutores. Além disso, o desenvolvimento da fortuna crítica na área ajuda a despertar a consciência a respeito do tema, dentre os tradutores tanto experientes como aprendizes.

Os primeiros usos mais expressivos e sistematizados do termo *colocação* são creditados a Firth, em seus artigos *Modes of Meaning* (1957) e *A Synopsis of Linguistic Theory* (1968) em que ele não apresenta uma definição explícita para o termo colocação, mas exemplifica-a.

Firth apresenta o fenômeno como a caracterização de uma palavra de acordo com outras palavras que, tipicamente, ocorrem com ela, mostrando que o aspecto relevante do significado de uma palavra é o conjunto de todas as outras palavras que com ela se combinam. Sua preocupação inicial era com o estudo do significado, e é muito interessante observar como ele chega a ideia de colocação pelo viés do estudo do significado.

Inicialmente, ele considerava que, para falar sobre significado em termos linguísticos, devemos aceitar o evento *linguagem* como um todo e tratá-lo em diferentes níveis, por vezes no sentido descendente: do contexto social à sintaxe e ao vocabulário, à fonologia e à fonética e, por vezes, na ordem inversa (Firth, 1957, p. 192), cada nível tendo um *mode of meaning*¹⁴. Através da observação de padrões, Firth conclui que existem tipos de combinações que

¹³ O "tradutor ingênuo" é um termo cunhado por Tagnin (2002) com base no "falante ingênuo" de Fillmore (1979), que usou esse termo para se referir a um aprendiz de língua estrangeira que desconhece as convenções da língua.

¹⁴ Em tradução simples, modo de significação.

indicam distribuições características de colocabilidade que podem ser consideradas como um nível de significado dentro de um grupo em especial ou da linguagem de um indivíduo. Sendo assim, da mesma maneira como as formas na língua já bem estabelecidas e habituais nos níveis fonéticos, fonológicos e gramaticais de certos grupos dão uma base para uma expectativa mútua entre palavras e frases nesses níveis, também há o nível das colocações usuais para descrever e atribuir significados.

Segundo Firth, o significado via colocação é uma abstração no nível sintagmático e não está diretamente vinculado à abordagem conceitual ou de ideias quanto ao significado das palavras. No artigo *A synopsis of linguistic theory* (1968), as considerações sobre o fenômeno da colocação são mais amplas, sempre pautadas pelo desenvolvimento da teoria que o autor apresenta, de que o significado das palavras acontece em diferentes níveis, sendo as colocações um desses níveis. É nesse artigo que Firth cita a célebre frase *You shall know a word by the company it keeps*, enfatizando seu ponto de vista contextual ao significado¹⁵. No entanto, pontua que não se deve interpretar as colocações como contextos e exemplos/citações que lexicógrafos usam em dicionários para exemplificar os diferentes significados de uma palavra.

Esse olhar, na época inovador, de Firth para esse evento linguístico, concentrado em regularidades e coocorrências, é um ponto que se perpetua nos estudos feitos em Linguística de Córpus, e é inegável a contribuição que esses primeiros pontos de vista sobre padrões de coocorrência explicitados pelo autor tiveram para embasar os estudos feitos com grandes córpus e, conseqüentemente, para o desenvolvimento da Linguística de Córpus.

Depois de Firth, são conhecidos como neo-firthianos seus seguidores: Michael Halliday e John Sinclair. O trabalho de Firth está nas bases da Linguística Sistêmico-Funcional (fundada por Michael Halliday) e da Linguística de Córpus (impulsionada pelos interesses e estudos de John Sinclair).

Halliday (1961) postulou como conceito de colocação "a associação sintagmática de itens lexicais, quantificável textualmente, com a probabilidade de ocorrerem dentro de um intervalo n (distância de n itens lexicais) de um item x, aos itens a, b, c...". Segundo Berry-Rogghe (1973), em outras palavras, o objetivo que se vê na definição de Halliday é a

¹⁵ Danielsson, P. (2003) lembra que Firth cita esta frase ao parafrasear a frase de Wittgenstein "o significado de uma palavra está em seu uso"

compilação de uma lista de itens sintagmáticos (colocados) que coocorrem significativamente com um determinado item lexical (nodo) dentro de uma distância linear especificada.

Para Sinclair (1991), as palavras, em algumas ocasiões, são escolhidas em pares ou grupos, não necessariamente adjacentes. Ele consolida o uso do termo nodo para a palavra que está sendo estudada e o termo colocado para as palavras que ocorrem no ambiente especificado do nodo. De acordo com Sinclair (1991, p. 170), na colocação estatística, as linhas de concordância exemplificam uma palavra ou expressão, a qual é chamada de nodo, e as palavras a sua esquerda e direita são seus colocados. A definição do fenômeno que o autor faz no livro *Corpus, Concordance and Collocation* (Sinclair, 1991, p. 170) é a de que uma colocação é a "ocorrência de duas ou mais palavras dentro de um curto espaço no texto". Sinclair (*op. cit*) acrescenta que a medida usual de proximidade é um máximo de quatro palavras entre nodo e colocado¹⁶. Essas seriam as colocações que são interessantes por serem inesperadas. Porém, ele cita que haveria também um segundo tipo de colocação, as que são interessantes por conta de serem frequentemente repetidas. Essas são normalmente relacionadas a medidas de significância estatística e independentes de padrões gramaticais ou relações posicionais. No citado livro, Sinclair esclarece que, nos exemplos com que trabalhou, os padrões colocacionais são restringidos a pares de palavras, embora não haja restrição teórica quanto ao número de palavras envolvidas.

Stubbs (2009) ressalta que na trilha de desenvolvimento do trabalho de John Sinclair (1998), em estudos posteriores, ele fazia referência ao modelo das ELU (*Extended Lexical Units*, unidades lexicais estendidas). As ELU têm um núcleo obrigatório e uma prosódia semântica. Sua estrutura interna tem quatro parâmetros. Seriam eles:

(1) Colocação: nesse ponto, Sinclair define colocação como a relação de coocorrência entre um núcleo obrigatório/grupo de palavras (nodo) e o colocado individual. Estes são observáveis diretamente e podem ser contados nos textos.

¹⁶ Michael Hoey (2005) reproduz essa definição, mas ressalta que ela não reflete a visão que Sinclair tinha do assunto. Diz Hoey (*op cit*) "Essa definição (que devo acrescentar, não reflete o uso que o próprio Sinclair fazia do termo) não é útil e pode causar uma confusão com instâncias únicas de coocorrências com padrões repetidos de coocorrências".

(2) *Coligação*: relação de coocorrência entre o nodo e categorias gramaticais abstratas. Uma categoria tradicional como “negação” pode ser percebida gramaticalmente (*would not budge*) ou semanticamente (*would hardly budge, refused to budge*)¹⁷.

(3) *Preferência semântica*: relação de coocorrência entre uma unidade e palavras de campos semânticos característicos. Colocados recorrentes fornecem uma evidência observável do tópico característico do texto em seu entorno (ex: sujeitos ou objetos típicos de um verbo).

(4) *Prosódia semântica*: é a função da ELU. É uma generalização sobre o propósito comunicativo da unidade; é a razão para escolhê-la (está, portanto, relacionada ao conceito da força ilocucionária). A prosódia semântica diz respeito à associação entre os itens lexicais e uma conotação (negativa, positiva ou neutra). Um exemplo demonstrado por Louw (1993) seria a palavra em inglês *utterly* que, ao se juntar com adjetivos do tipo *grand* e *vulnerable*, resulta em um sentido irônico; ou o famoso *cause* que, em inglês, demonstra se associar a palavras de sentido negativo, como *accident* em *cause an accident*.

Stubbs (2009) faz uma releitura sobre esses parâmetros primeiramente delineados por Sinclair. Para ele, as relações entre colocação e coligação abordam como os signos lingüísticos se relacionam sintagmaticamente entre si. A preferência semântica trata de como os signos linguísticos se relacionam semanticamente com o tópico do texto. A prosódia semântica, por fim, trata de como os signos linguísticos se relacionam pragmaticamente com seus usuários. Nesse modelo, (1) e (2) são evidências de (3) e (4). As colocações e as coligações são parte de nosso comportamento linguístico inconsciente e podem ser identificadas automaticamente por *softwares*. Já os tópicos discursivos só podem ser identificados intuitivamente, e a prosódia semântica é uma hipótese sobre uma ação consciente e intencional.

Além do modelo das ELU exposto acima, encontramos também referência às MSU (*Meaning Shifting Units* - Unidades Alteradoras de Significado). Em manuscrito datado de 2007, Sinclair faz referência ao termo MSU para o que ele considerava estar chamando vagamente de colocação. O termo MSU deveria, então, enfatizar o principal critério do fenômeno, a alteração de significado que a seleção concomitante de constituintes provoca. Cada nova combinação de palavras gera uma alteração de significado, mesmo que seja sutil, em comparação com outras possíveis combinações, envolvendo uma ou mais das palavras em

¹⁷ Traduções literais para os exemplos seriam: *would not budge* (não afetaria); *would hardly budge* (dificilmente afetaria); *refused to budge* (recusou a afetar). Esses são exemplos usados por Sinclair.

questão. Esses trabalhos mais recentes elaborados por John Sinclair estão sob a forma de manuscritos, no Tuscan Word Centre, e tivemos acesso a eles apenas através de referências nos artigos de Stubbs (2009) e Cheng et al. (2008).

Após expor essas ideias fundamentais desses três expoentes britânicos sobre o estudo de colocações (Firth, Halliday e Sinclair), apresentamos também algumas classificações e sistematizações dos estudos sobre o tema, realizadas por outros pesquisadores, as quais ajudam a entender como o tema é/foi segmentado. Dessa forma, podemos também compreender semelhanças e diferenças entre os diferentes pontos de vista em um contexto maior, pois cada perspectiva é determinante para a definição do evento que enfoca.

1.1. Diferentes classificações para estudo e descrição das colocações

Segundo um amplo estudo realizado por Beneduzzi (2008), em se tratando de diferentes abordagens acerca das colocações, as abordagens se dividem naquelas com enfoque lexicológico e com enfoque lexicográfico. Os estudos com enfoque lexicográfico se preocupam com a forma com que as combinações devem/podem ser apresentadas nos dicionários, se a partir da base (ou núcleo) ou do colocado. São exemplos de autores que estudaram as colocações sob esse enfoque¹⁸: Cowie (1981, 1994); Haensch *et al* (1982); Benson (1985, 1986, 1997); Seco (1987a, 1987b); Hausmann (1979, 1985, 1989, 1998a). Os estudos com enfoque lexicológico se dividem em estudos voltados para aspectos semânticos (Bally, 1951; Porzig 1950; Corpas Pastor, 1996), estatísticos (Firth, 1957; Halliday, 1961, 1966; Sinclair 1991; Tagnin 2005) e lexicográficos (Cowie 1981, 1994; Hausmann 1979, 1985, 1989, 1998a, Benson 1985, 1986, 1997).

Já Michael Hoey faz uma classificação elaborada inicialmente por Partington (*apud* Hoey, 2005), que classifica as colocações nas que enfocam aspectos *textuais*, aspectos *estatísticos* e aspectos *psicológicos*.

A definição do ponto de vista *textual* compreende a ocorrência, em um texto, de duas ou mais palavras dentro de um curto espaço entre si. A definição *estatística* compreende a relação que um item lexical tem com itens que aparecem com probabilidade maior do que a aleatória em um contexto (textual). Para a definição psicológica, Hoey apresenta duas

¹⁸ As citações destes autores foram extraídas da dissertação de mestrado de Renata Beneduzzi, que fez esse importante levantamento (Beneduzzi, 2008)

definições. A primeira definição *psicológica* é a de Halliday e Hasan (1976): “tipo de coesão que resulta da co-ocorrência de itens lexicais que de alguma forma se associam tipicamente, porque tendem a ocorrer em ambientes semelhantes” (p.287). Hoey explica que Halliday e Hasan não estão se referindo a uma coocorrência regular com uma certa proximidade, mas a uma associação em que as palavras são regularmente associadas na mente devido a maneira como se encontram regularmente em contextos textuais semelhantes.¹⁹ A segunda definição *psicológica* seria a de Leech (1974, p. 20), que fala de “significado colocativo”, o qual consistiria de “associações que uma palavra adquire devido aos significados das palavras que tendem a co-ocorrer em seu ambiente”. Isso também implica que a palavra adquire conotações como resultado das palavras que a circulam. Hoey explica que essa posição também foi formulada por Louw (1993) e, posteriormente, foi empregada por Stubbs (1995).

Cristopher Gledhill (2000) também distingue o fenômeno *colocação* em três abordagens. Ele agrega a perspectiva estatística e textual em um grupo e as diferencia das definições semântico-sintáticas e discursivo-retóricas.

Na primeira perspectiva, a estatística/textual, o autor comenta que não há uma definição linguística de colocação que seja totalmente confiável e, por essa razão prática, geralmente as colocações têm sido definidas por fatores estatísticos. Essa perspectiva pode ser resumida na definição de Van Roey (*apud* Gledhill, 2000):

[Colocação é] o fenômeno linguístico no qual um dado item do vocabulário prefere a companhia de outro item em vez de seus "sinônimos" devido a restrições que não estão no nível da sintaxe ou do significado conceitual, mas no uso. (p. 46)

Gledhill (2000) acrescenta que, em alguns trabalhos da Linguística de Córpus, o que separa uma colocação de uma mera coocorrência é o nível estatístico, a partir do qual o pesquisador “tem prazer” em dizer que a coocorrência não é acidental. Essa definição enfatiza a colocação como coocorrência (palavras que frequentemente se combinam) e recorrência (combinações que ocorrem com frequência na língua). Com uma visão estatística, os linguistas de córpus podem identificar padrões que não seriam reconhecidos usando os métodos tradicionais.

Essa perspectiva é conhecida como textual porque ela se baseia somente na capacidade de sistemas computacionais de analisar grandes quantidades de textos. Segundo Gledhill

¹⁹ Posteriormente, Hasan abandonou esse conceito e usou relações semânticas mais específicas (hiponímia, meronímia etc.)

(2000), a colocação estatística também tem uma probabilidade de ter uma função textual específica ou de ocorrer em contextos mais específicos.

A segunda perspectiva é a semântico-sintática. A tradição da abordagem semântica e estatística, segundo esse autor, define o evento como uma relação mais abstrata entre as palavras, sem fazer referência à frequência, coocorrência ou probabilidade, chamando a atenção para a combinatória léxica. Nessa abordagem, concentram-se os trabalhos que fazem distinção entre níveis mais ou menos restritos de associações, posicionando o evento em um contínuo de colocações mais livres e expressões idiomáticas mais fixas. É uma abordagem que inclui também a associação entre um padrão gramatical específico e um significado em particular. Gledhill acredita que, dessa perspectiva, viu-se a noção de *colocação* passar de um sentido restrito de expressões idiomáticas para outras noções, menos convencionais, como *colocação gramatical*, *prosódia semântica*, *quadros colocacionais* e *coligações*.

A terceira perspectiva apresentada pelo autor é a da abordagem discursivo-retórica. Nesse ponto de vista, as expressões fixas são vistas como escolhas significativas no nível paradigmático, representando escolhas significativas na retórica de quem as fala/escreve. O foco não está no grau de cristalização ou na composicionalidade das expressões, mas, sim, em seu uso, que tem, principalmente, um efeito retórico.

A perspectiva semântico-sintática mostra a relação abstrata entre as palavras e a ideia de que uma expressão - ou colocação - existe como uma unidade de significado enquanto a perspectiva discursivo-retórica informa sobre a função da expressão dentro de um texto, sendo uma opção também relacionada a estilo.

Essas classificações são feitas conforme diferentes pontos de vista, o que, como vemos, acarreta diferentes classificações. Contudo, percebe-se que elas podem se sobrepor e não precisam ser excludentes. A abordagem estatística, por exemplo, é caracterizada por seu método de identificação (estatístico), mas é através desses levantamentos que se pode chegar a *insights* importantes a respeito dos níveis sintáticos, semânticos e pragmáticos das combinações.

No desenvolvimento desses estudos, encontramos uma série de denominações para caracterizar esses fenômenos de associações e coocorrências, tais como as associações semânticas, os *concgrams* e, em uma escala maior, as fraseologias, que veremos a seguir. Sinclair (1991, 2004) e Hoey (2005) também se preocuparam em tentar explicar, em termos

de descrição linguística, o que provocaria essas associações e coocorrências, e o fizeram através dos conceitos do *idiom principle* e do *lexical priming*, respectivamente.

Antes de passar para uma revisão dos três outros conceitos acima mencionados (associações semânticas, *congrams* e fraseologias), vamos explorar brevemente a descrição de dois fenômenos que são tratados como maneiras de explicar o que motiva a formação das referidas associações entre as palavras dentro de um idioma.

1.2. O princípio idiomático (*idiom principle*)

Em trabalhos mais antigos, Sinclair (1991) partia do pressuposto de que as palavras não ocorrem de maneira aleatória em um texto e de que o princípio da livre escolha (*open-choice principle*) não daria conta de restrições que podem ocorrer na seleção das palavras ou frases, além das restrições gramaticais. Por isso, estabelece também outro princípio organizador da língua, o *princípio idiomático (idiom principle)*, em que o falante tem à sua disposição um grande número de grupos de palavras pré-construídos que não seriam capturados pelo princípio da livre escolha.

O *princípio idiomático* daria conta da tendência fraseológica, em que as palavras são cosseleccionadas pelos usuários, dando lugar à formação de colocações e a outros recursos da idiomaticidade (Cheng *et al.*, 2006). Cheng (2008) aponta que Sinclair (1987) afirmava que os usuários da língua cosseleccionam as palavras que falam e escrevem para produzir unidades de significado, embora pareça que as palavras podem ser analisadas em segmentos.

Posteriormente Sinclair (1996) passou a referir esses fenômenos mais em termos de tendências: a *tendência fraseológica*, em que as palavras se agrupam e adquirem significado por suas combinações, e a *tendência terminológica*, em que as palavras têm um significado fixo em referência ao mundo. Por fim, o autor considerou que a maioria das expressões chamadas fixas, não são fixas, existem apenas expressões variáveis (Stubbs, 2009).

1.3. *Lexical priming*

A teoria do *Lexical priming* é apresentada por seu idealizador, Michael Hoey (2005), como uma nova teoria do léxico, que leva a uma nova teoria da linguagem. Para o autor, a teoria inverte os papéis do léxico e da gramática, no sentido de que o léxico é estruturado de modo complexo e sistemático e a gramática é o resultado da estrutura léxica.

O autor menciona que mesmo as teorias do léxico mais sofisticadas, como a linguística sistêmico-funcional, representam a relação entre gramática e léxico como se a escolha lexical precisa fosse a última a ser feita, sempre passando, primeiramente, por uma espécie de filtro gramatical. Ele cita ainda a *Construction Grammar*, que não separa sintaxe e léxico, mas que, ainda assim, lida com construções léxicas e sintáticas.

Hoey (*op. cit*) aponta que as teorias se voltam para o que é possível na linguagem, e não para o que é natural. A naturalidade tem sido muito discutida por meio dos estudos das colocações, que são um fator fundamental para atingi-la. Para o pesquisador, a colocação é, fundamentalmente, um conceito psicológico, em que as palavras integrantes são mentalmente “motivadas” (*primed*²⁰) para o uso de uma colocação.

O *priming* explicaria, além das colocações, outros recursos da língua, como as associações semânticas e outros. A seguir, reproduzimos as hipóteses a respeito de *lexical priming* apresentadas pelo autor:

1. todas as palavras são *primed* para ocorrer com determinadas palavras, formando colocados;
2. todas as palavras são *primed* para ocorrer com determinados ambientes semânticos, formando associações semânticas²¹;
3. todas as palavras são *primed* para ocorrer em associação com funções pragmáticas em particular, formando associações pragmáticas;
4. todas as palavras são *primed* para ocorrer com (ou evitar) certas posições e funções gramaticais, formando as coligações;

²⁰ O verbo *prime*, em inglês, tem o sentido de preparar, começar, pré-ativar. Acharmos que *motivar* seria uma tradução pertinente. No entanto, deixamos a palavra em inglês em alguns casos para que se pense no sentido que o autor quer dar à palavra.

²¹ A teoria dos Lexical Primings não faz menção à teoria dos sentidos primitivos de Anna Wierzbicka (2000), que segue uma direção mais cognitiva.

5. co-hipônimos e sinônimos diferem a respeito de suas colocações, associações semânticas e coligações;
6. quando uma palavra é polissêmica, as colocações, associações semânticas e coligações de um sentido da palavra diferem dos outros sentidos;
7. todas as palavras são *primed* para uso em uma ou mais funções gramaticais; essas seriam suas categorias gramaticais;
8. todas as palavras são *primed* para participar em, ou evitar, determinados tipos de relação coesiva em um discurso, essas são suas colocações textuais;
9. todas as palavras são *primed* para ocorrer em determinadas relações semânticas no discurso, essas são suas associações semânticas textuais; e
10. todas as palavras são *primed* para ocorrer ou evitar certas posições dentro do discurso, essas são suas coligações textuais.

E, muito importante, todas essas hipóteses são primeiramente restringidas por domínios ou gêneros. Ainda, o autor ressalta que as palavras não são ativadoras *per se*, elas são ativadoras ou ativadas por alguém. O que um córpus pode fazer, segundo Hoey, é indicar que alguns *primings*, provavelmente, sejam compartilhados por um amplo número de falantes, e, somente nesse sentido, o *priming* é independente do indivíduo. Os *primings* são considerados psicológicos, pois fazem parte das concordâncias que ocorrem na mente do falante.

2. Concgrams

Principalmente dos campos de PLN (Processamento da Linguagem Natural) e de linguística computacional, conhecemos o termo *n-gramas*, que são sequências de palavras contíguas, podendo ser compostos de duas (bigramas), três (trigramas) palavras e assim por diante. Algumas vezes, essas expressões podem ser denominadas *clusters*/agrupamentos de palavras, *clusters*/agrupamentos lexicais, lexical bundles (conjuntos/feixes lexicais), unidades/expressões multipalavras. Um *n-grama* seria, por exemplo, tanto a expressão “por outro lado” como a sequência “e então ele”.

Ainda no campo de PLN, foi criada a noção de *skip-gram*, que daria conta de preencher uma lacuna deixada pela busca de *n-gramas*, aquelas associações de palavras que não são contíguas. A ideia subjacente é que, embora haja uma regularidade na linguagem, que pode ser identificada em análises de cópua, há também o fato de o sistema linguístico ser composto de eventos raros, variados e complexos, e, por isso, nunca seria possível modelar todas as possibilidades existentes (Guthrie et al., 2006).

Guthrie *et al.* (2006), então, esclarecem que *skip-grams* correspondem a uma técnica inicialmente usada no campo de processamento da fala (em fonemas), em que se formam os *n-gramas*. No entanto, além de identificar palavras em sequência, permite que algumas delas sejam omitidas²². No exemplo fornecido em Guthrie *et al.* (*op cit*), na sequência *I hit the tennis ball*, haveria três trigramas: *I hit the*, *hit the tennis* e *the tennis ball*. Porém, um *trigrama* igualmente importante implicado por essa sentença e que não seria normalmente capturado seria *hit the * ball*.

Uma busca por *skip-grams* inclui também as sequências contíguas, ou seja, os *n-gramas* encontrados nessas sequências; portanto, segundo Greaves (2009), isso faz com que a denominação *skip-gram* seja potencialmente enganosa, pois pode levar a acreditar que são localizadas apenas associações não contíguas. As buscas por *skip-grams*, segundo Wilks (2005), também têm limitações. Elas buscam *skip-grams* de até 3 palavras e 4 "omissões". Assim, duas palavras associadas e que tenham uma distância de mais de 4 palavras entre si, não são detectadas e, além disso, as buscas requerem que uma fórmula seja inserida, o que pode ser um tanto confuso. Greaves (2009) acrescenta ainda que as pesquisas de *skip-grams*, da mesma forma que as de *n-gramas*, não dão conta de variações posicionais (por exemplo, quando as palavras associadas ocupam posições diferentes na frase: AB, BA)

Basicamente, é nesse ponto das variações que os pesquisadores que postularam a noção de *concgrams* (Greaves, 2009) e que desenvolveram uma ferramenta para identificação desses eventos, também denominada *ConcGram*, têm seu ponto mais forte de argumentação. Os *concgrams* são conjuntos de palavras que coocorrem independentemente de sua variação I) constitutiva, ou seja, se são contíguas (por exemplo AB e A * B); e II) posicional (por exemplo, AB e BA), se estão em sequência, ou ambas.

²² A palavra *skip*, em inglês, pode corresponder aos substantivos pulo, salto, omissão ou aos verbos pular, saltar, omitir.

A ferramenta *ConcGram*²³ foi desenvolvida para identificar fraseologias automaticamente e, em especial, a variação fraseológica em textos e *córpus*. O termo fraseologia, aqui, é usado em um sentido amplo, para descrever a tendência das palavras de serem cosseleccionadas pelos usuários da língua para produzirem significados. Os autores argumentam que, ao examinar os *concgrams*, nos aproximamos de um maior e melhor entendimento do princípio idiomático (Sinclair 1987), o qual embasa a ideia de que a fraseologia está no centro de todo o uso da linguagem. Nesse ponto, os autores destacam que, na descrição de fraseologia, é central a identificação das MSU, que seria uma “atualização” do que Sinclair havia anteriormente denominado “item lexical”²⁴.

Um ponto importante de se observar nos resultados do *ConcGram*, e apontado pelos autores, é que ele extrai todas as coocorrências de palavras, mas não determina se elas têm uma associação em termos de significado ou se são apenas coocorrências. O termo coocorrência diz respeito a qualquer palavra que ocorre dentro de um espaço de vizinhança definido pelo usuário em relação à outra palavra; uma palavra associada é aquela cuja coocorrência não é acidental, mas forma um colocado com a outra palavra.

```

1 next five years. We also estimate that operating expenditure will increase moderately, at a rate
2 in the Consolidated Account for 2008-09. Public expenditure as a proportion of GDP will increase
3 two and a half times the present population. The expenditure on the Old Age Allowance will increase
4 Since the planning of these projects takes time, expenditure on infrastructure is unlikely to increase
5 health care system were to remain unchanged, expenditure on public health care services would increase
6 and social development. We will increase expenditure on social services and welfare and return part
7 lead to a decrease in revenue and an increase in expenditure in 2008-09. I have also earmarked $50 billion to
8 revenue by $33.5 billion and increase operating expenditure by $41.5 billion in 2008-09.
9 This measure will increase government expenditure on CSSA payments when inflation rises. 145.
10 it is expected that the increase in overall expenditure on health care services will, on average, be two
11 be sustainable. If we increase recurrent public expenditure or reduce recurrent public revenue, we must
12 Government cannot increase public health care expenditure indefinitely, we hope that supplementary
13 we will increase the share of public health care expenditure to 17 per cent of government recurrent

```

No caso do exemplo, os resultados estão alinhados pela palavra *expenditure*, mas a coocorrência *increase* também aparece em negrito, evitando-se, assim, a falsa impressão de que a palavra nodo é a principal e as outras estão subordinadas a ela.

Atualmente, segundo Greaves (2009), três categorias de fraseologias estão sendo identificadas e descritas mais exhaustivamente graças ao *ConcGram*: as MSUs, os quadros colocacionais e os organizacionais. Também têm sido úteis na identificação de perfis fraseológicos de textos e *córpus*, os *aboutgrams*, que são associações de palavras específicas de um texto ou *córpus*.

²³ A ferramenta é comercializada pela editora John Benjamins: http://www.benjamins.com/cgi-bin/t_bookview.cgi?bookid=CLS%201

²⁴ “I think that lexical item is not a good choice for this because it is associated with dictionaries rather than structures” (Sinclair, 2007a, p. 3)

corpus are termed ‘aboutgrams’ (Sinclair, personal communication)

Os quadros colocacionais são cosseleções de palavras gramaticais. Estudos com *concgrams* em busca de quadros colocacionais (Greaves and Warren, 2008; Li and Warren, 2008) mostram que os cinco mais frequentes, em inglês, são ‘the ... of’, ‘of ... the’, ‘in ... the’, ‘a/an ... of’, e ‘the ... of ... the’ em uma amostra de 5 milhões de palavras do British National Corpus (Li and Warren, 2008).

Os quadros organizacionais são uma forma de fraseologia que denota as formas como elementos organizacionais no discurso, como conjunções, conectores e outros marcadores discursivos que podem ser cosseleccionados por usuários da língua, tanto na forma escrita como na falada. Seriam exemplos ‘either ... or’, ‘both ... and’ e ‘whether ... or’.

Os *aboutgrams* mencionados acima têm relação com o perfil fraseológico de um texto. O perfil fraseológico é composto de todas as associações de palavras de um texto ou cópuz; e o que se tem denominado *aboutness* de um texto ou cópuz pode ser determinado pelas associações de palavras que são específicas desses textos. Essas associações de palavras foram denominadas “*aboutgrams*” (Greaves, 2009).

3. Fraseologia

Para Berber-Sardinha (2004), “padrão” e “fraseologia” são referidos, algumas vezes, como sinônimos, como bem indicou Teixeira (2008). Há, no entanto, uma longa tradição de extensos estudos sobre fraseologia, muitas vezes tratada como uma área exclusiva de estudo, que pode abarcar tanto a linguagem cotidiana como as linguagens especializadas. Porém, nesta tese, estamos voltados para o sentido mais amplo e menos específico do termo fraseologia. Estamos voltados para fraseologia no que se refere aos padrões que podem ser observados de uma palavra em questão (Teixeira, 2008).

Cheng et al (2009) esclarece que, no estudo dos *Concgrams*, por exemplo, o termo fraseologia é usado em um sentido amplo para descrever a tendência de as palavras serem cosseleccionadas na linguagem oral ou escrita para que se produzam significados. Com sentido semelhante, Christopher Gledhill (2000) usa o termo fraseologia para se referir ao uso retórico e pragmático de uma expressão, definindo-o como “o jeito preferido de dizer as coisas dentro de um discurso em particular” (p. 1). Ele considera que a fraseologia é uma dimensão de uso

da língua na qual os padrões das construções (padrões léxico-gramaticais) codificam visões semânticas de mundo e, em um nível mais alto, as expressões idiomáticas e fixas têm funções retóricas e textuais dentro de um discurso específico. Para esse autor, a fraseologia é ao mesmo tempo uma dimensão pragmática da análise linguística e um sistema de organização que abrange outras relações lexicais: as colocações e a léxico-gramática. Assim, a análise da fraseologia de um texto não deve envolver somente a identificação de colocações e expressões específicas, mas também a correspondência entre a expressão e o discurso no qual foram produzidas.

Outro autor que tratou do tema de uma forma bem mais aberta foi Daniel Gouadec (1994). Ele define fraseologia como “sequência notável de caracteres” com variáveis e palavras-chave constituindo diferentes matrizes:

Não dissemos, no fundo, o que é um fraseologismo. Se tomo esta definição ampla, vocês vão me dizer que “qualquer cadeia de caracteres da língua responde à definição”. Se dou uma definição estrita, vocês vão me objeter que os “fraseologismos são, portanto, os idiomatismos”. Escolhi uma via pragmática, na qual o fraseologismo se situa entre o idiomatismo... e a combinatória...(p. 170)²⁵

E, acrescenta ele:

(...) dizer que a fraseologia – como a terminologia – concerne cadeias de caracteres (significativas, isso é óbvio) notáveis, porque especializadas, ou repetitivas (e, portanto, provavelmente especializadas e de risco), é colocar o debate no terreno do *langagère* mais do que no terreno do linguista.

Com a abordagem de matrizes, fica-se, no entanto, mais preso à noção de “cadeia” de caracteres, e o que queremos abordar são associações entre palavras e grupos de palavras que têm diferentes posições dentro de uma frase. Ainda assim, cabe descrevermos um pouco da maneira como Gouadec expôs a noção de fraseologia, pois ela apresenta uma visão aberta, que não se limita a considerar cadeias de caracteres segundo critérios de tamanho, constituição sintática ou morfológica. Tal visão poderia ser a ideal para um estudo como este, que não definiu *a priori* o tipo de evento linguístico que é considerado relevante para a tradução.

O autor trabalha com uma ideia bastante ampla de cadeias de caracteres especializadas. Essas cadeias têm tamanho variável, “incluem palavras, grupos de palavras,

²⁵ O texto de Gouadec em português foi obtido de uma tradução preliminar do texto original em francês, feita pelas professoras Patrícia Ramos e Cleci R. Bevilacqua da Universidade Federal do Rio Grande do Sul, integrantes do grupo TERMISUL.

termos, locuções, expressões, orações, segmentos de frases, frases, conjuntos de frases” (Gouadec, 1994, p. 167). Para estudos permeados por finalidades voltadas à tradução, uma abordagem que não mantém o foco exclusivamente em termos já indica mais chances de apresentar resultados mais úteis, uma vez que o profissional de tradução se vê em um processo que envolve uma construção de um todo, que é um texto, e não apenas a transposição de palavras isoladas.

Segundo o autor, o domínio de uso das cadeias de caracteres de usos especializados dá “indícios de competência técnica e linguística”, ou seja, “qualquer defeito de domínio da dita cadeia [...] seria interpretado como um indício de incompetência” (p.172). Uma cadeia de caracteres é notável devido a alguns fatores importantes de serem considerados no âmbito desse estudo: especialidade, recorrência, risco, vantagem.

Esses critérios são reiterados dentro do ponto de vista do autor e são para nós valiosos, no sentido de que buscamos associações mais relevantes para um tradutor. Os critérios apresentados por Gouadec nos dão respaldo para considerar as associações que identificamos como relevantes.

Ao observarmos as sequências a seguir, passamos a entender melhor ao que o autor se refere:

- (1) Loopback
- (2) Digital loopback
- (3) Remote digital loopback
- (4) Remote digital loopback test
- (5) Initiate a remote digital loopback test
- (6) Cause the modem to initiate a remote digital loopback test
- (7) After initiating the digital loopback test, the modem returns to direct or normal operation mode
- (8) Once the mode has initiated a remote digital loopback test, a message received from (x) will set up a response from (y)

O autor explica que a sequência 6 poderia se classificar entre os híbridos termos-fraseológicos, mas que ela se articula em torno de um termo e se indexa a partir desse termo: *digital loopback test*. O fraseologismo da cadeia 7, no entanto, não depende de termos incluídos; é notável por ele mesmo e sua matriz seria *after (x), the modem (y)*. Essa matriz inclui variáveis intercambiáveis e modificáveis. Já a cadeia 8 remete mais ainda a uma fraseologia de formulação, e não de designação. Gouadec diz ainda que, para que essa cadeia 8 seja objeto de uma observação especial, ela precisa ser: especializada, repetida – parcial ou integralmente – ou apresentar um risco ou apresentar algo que seja vantajoso tratá-la. Nessa cadeia, o caráter especializado se dá por meio dos termos incluídos e de uma provável repetição, segundo Gouadec.

A visão do autor sobre fraseologismo é ampla, tanto que ele se preocupa em dizer que, para que “justifique um tratamento particular, é preciso que apresente um caráter de estereotipia fortemente marcado” (p. 171). Essa estereotipia pode estar relacionada a domínio (área) ou tipo de discurso ou, ainda, à repetição. Também justifica-se pelos quesitos de anomalia ou ruptura e de frequência.

Para o autor, uma cadeia de caracteres significantes comporta (i) um pivô e uma variável ou (ii) uma matriz e variáveis, apresentando um caráter de estereotipia.

Gouadec (1994) tem uma visão de terminologia e fraseologia que bifurca, mas que também une as áreas de Terminologia e Fraseologia. Diz ele: “Fazer terminologia é definir o que os termos designam e como eles se comportam. Fazer fraseologia é explicar o que as expressões significam, para que elas servem e como se comportam” (p. 171). Segundo ele, a fraseologia ainda comportaria uma ramificação: fraseologismos com pivô terminológico e fraseologia de variáveis terminológicas (ou lexicais).

Outra observação importante no texto de Gouadec, para o nosso trabalho, é quanto a equivalências bilíngues de fraseologismos. Esse é um ponto valiosíssimo para fundamentar a tese que estamos desenvolvendo. Gouadec fala de chaves de acesso, ou seja, uma palavra que faria a correlação interlinguística entre uma língua e outra. Usando o exemplo do autor, entre *when a charge is deleted* e *qualquer supressão de taxa*²⁶, a correlação linguística entre os dois fraseologismos se daria pela chave de acesso *charge* (taxa). Esse aspecto também está relacionado com o que vínhamos discutindo no capítulo anterior sobre traduções linguísticas e

²⁶ Este exemplo está em francês no texto de Gouadec e foi por nós traduzido.

interpretativas, já que charge/taxa é um ponto de “união” entre as duas línguas, e o modo de dizer essa chave de acesso na frase é que sofre alterações sintáticas e até semânticas. Esse ponto de vista também não deixa de estar relacionado com as classes *Adhoc-block* e *Rest-block* presentes no modelo de Königs, a ser referido no Capítulo II.

4. Definindo nossas unidades de análise

As abordagens a associações sintagmáticas que são voltadas para a língua geral e que não se propõem motivadas por outros pontos de vista, como o da tradução, em especial as voltadas para as colocações, são, em geral, pautadas pela restrição/alteração de significado que os elementos operam uns sobre os outros. Em nosso ponto de vista, as associações sintagmáticas que se tornam relevantes para a tradução não necessariamente são enfatizadas por operarem restrições de significado. Elas podem demonstrar-se frequentes e recorrentes dentro de um determinado gênero ou domínio, formando associações motivadas por “forças” gramaticais, semânticas e pragmáticas, seguindo o modelo dos *lexical primings* apresentado acima.

Dentre as possíveis associações descritas acima, nas hipóteses apresentadas por Hoey (2007), temos o nível de motivação (*priming*) semântica, pragmática e gramatical, quando palavras atraem outras por estarem no mesmo ambiente semântico ou por estarem realizando uma função no texto e ainda por circunstâncias gramaticais. Mike Scott, no texto de ajuda do WordSmith Tools v.3²⁷, esclarece que a literatura nunca conseguiu diferenciar muito satisfatoriamente a noção de colocados como palavras associadas a uma palavra (por exemplo, *carta* e *selo*) e palavras coocorrentes (*carta, minha, esta, uma*). Segundo ele, os do primeiro tipo poderiam ser chamados de “colocados de coerência” e os do segundo “colocados vizinhos”. Acrescenta ainda que detectar colocados de coerência é complicado, pois, quando se começa a olhar além de um horizonte de 4 ou 5 palavras à direita ou à esquerda, a diversidade é tamanha que há mais ruído do que evidências no sistema.

Em nosso entendimento, o que Mike Scott chama de *colocados de coerência* está no nível das associações semânticas. Manning e Schütze (1999) esclarecem que alguns autores generalizaram a noção de colocação incluindo casos de palavras com uma forte relação entre si, mas que não necessariamente ocorrem como uma unidade gramatical e em uma

²⁷ O WordSmith Tools é um programa desenvolvido por Mike Scott e comercializado pela Oxford University Press, estando em sua quinta versão atualmente. <http://www.oup.com/elt/catalogue/isbn/6890?cc=gb>

determinada ordem, como médico-enfermeira e avião-aeroporto. Nesses casos, eles recomendam que seria melhor restringir o uso do termo colocações para o sentido mais estrito, o de uma associação vinculada gramaticalmente em uma determinada ordem, e usar o termo associação e coocorrência para o fenômeno mais geral de palavras que têm uma probabilidade de serem usadas no mesmo contexto (Manning e Schütze, 1999).

Também, parece que alguns autores usam os termos “prosódia semântica” e “associação semântica” como sinônimos, se referindo a certas palavras que estão em uma relação de colocação com grupos de palavras de fundo positivo ou negativo, ou com conjuntos semânticos de significado (Nelson, 2006). Sinclair foi quem primeiro notou o fato de que certas palavras pareciam ter como colocados outras palavras que eram positivas ou negativas, como, por exemplo, em inglês o verbo *happen* (acontecer) associado a coisas desagradáveis, como acidentes (Sinclair, 1991).

Posteriormente, Louw (1993) consolidou o uso do termo prosódia semântica e Stubbs (1995) o expandiu, incluindo além da possibilidade de as palavras se associarem com outras de cunho positivo ou negativo, também outros grupos de mesmos conjuntos semânticos. Um exemplo usado por Stubbs é a palavra *job* (emprego) que se mostrou positiva e negativa, enquanto *career* (carreira) foi apenas positiva e *unemployment* (desemprego) demonstrou se associar com o conjunto semântico da área da estatística (1995, p. 254).

As unidades que enfocamos neste trabalho poderiam, também, ser consideradas como fraseologias de um texto especializado, tomando fraseologia no sentido de tendência das palavras de serem cosseleccionadas pelos usuários da língua para produzirem significados (Gledhill, 2000). Porém, como há também definições mais específicas para Unidades Fraseológicas Especializadas, em que estas são entendidas como unidades sintagmáticas que incluem um termo entre seus elementos e que possuem um determinado grau de fixação e uma frequência relevante em um conjunto de textos ou em um âmbito especializado (Bevilacqua, 2004), achamos mais conveniente tratar nossas unidades de análise como uma categoria mais aberta.

Conforme constata Bevilacqua (2004), os textos especializados servem-se de termos e também de outras unidades que, com eles combinadas, podem originar distintas unidades portadoras de conhecimento especializado. Nossa investigação evidenciará, ao final, um desses outros tipos de unidades portadoras de conhecimento especializado, não exclusivamente terminológicas, que, conforme nossa hipótese: i) são compostas de itens não

contínuos, ii) apresentam variação posicional e composicional dos itens e iii) não necessariamente contém um termo em sua constituição. Além de assim serem, neste trabalho ressaltamos essas unidades porque elas se configuram como unidades de tradução especializadas, fator que observaremos no capítulo a seguir.

Dentre todas as definições pesquisadas, identificamos a ideia de associações motivadas por questões semânticas, pragmáticas e gramaticais como a que melhor traduz o nosso entendimento dessas associações. Como Hoey (2007) não enfoca, na teoria dos *lexical primings*, as variações posicionais e constitutivas, o conceito de *concgram* contempla justamente esse aspecto. Por isso, nossas unidades, em larga escala, serão identificadas usando o software *ConcGram* e serão descritas como associadas por uma “força” semântica, pragmática e/ou gramatical.

CAPÍTULO II

O ENCADEAMENTO SINTAGMÁTICO NO PRODUTO E NO PROCESSO TRADUTÓRIO

Como as unidades de análise se estabelecem em função da *tradução*, este capítulo abordará temas sobre a tradução como processo e como um produto e abordará em que medida as associações sintagmáticas estão relacionadas e são relevantes para cada um deles.

Entendemos que o encadeamento sintagmático do texto na língua de chegada é relevante para a tradução tanto quando pensamos no produto – pois o texto traduzido deve respeitar, entre outras coisas, as restrições de gênero, de domínio, de língua e de comunidade discursiva que recebe o texto – como quando pensamos no processo - pois, para atingir essa adequação, o tradutor reconstrói o significado não com base em palavras, mas em blocos maiores, ou em unidades de tradução²⁸, que, muitas vezes, podem ser constituídas de construções fraseológicas. Podemos, portanto, considerar como critérios de relevância para um produto e um processo satisfatórios não apenas a frequência e/ou a prototipicidade dessas associações, mas também i) o risco de não usá-las adequadamente no produto e ii) a vantagem de conhecê-las durante ou antes do processo de tradução. Esses dois critérios, de risco e vantagem, já foram usados por Gouadec (1994) para ajudar a definir o que torna uma sequência de caracteres “notável” do ponto de vista de quem trabalha com textos.

Enquanto o fator de prototipicidade pode ser comprovado por critérios de frequência, os fatores de risco e vantagem são mais difíceis de mensurar. No entanto, entendemos que eles estão associados com a ideia de erros e de inadequações na tradução. Embora a atividade

²⁸ Voltaremos a tratar de unidades de tradução neste capítulo.

de traduzir seja subjetiva e as possibilidades de interpretação sejam muitas, o que leva a se falar mais em inadequação de tradução do que em erro, em alguns ambientes e situações, as escolhas do tradutor podem, sim, ser classificadas como erros. Para ressaltar a importância das associações sintagmáticas na tradução, este capítulo inicia com uma apreciação sobre risco e vantagem, erro e inadequação. Posteriormente, o capítulo faz uma reflexão sobre o processo de sintagmatização na tradução. Para isso, buscou-se embasamento na Teoria Interpretativa da Tradução e na noção de unidades de tradução.

1. A relevância com foco no produto: risco e vantagem, erro e inadequação

Cabré (1999) sustenta que o tradutor de textos especializados deve conhecer alguns parâmetros para que sua tradução consiga atingir os índices mínimos de qualidade, ou seja, além de ser verdadeira, do ponto de vista do conteúdo, e correta, do ponto de vista gramatical, deve ser adequada e soar natural. Ao ferir esses índices, o tradutor pode ser julgado e as opções que realizou podem vir a ser consideradas erros ou inadequações.

Hurtado Albir (2002, p. 307) considera que a análise de erro não pode ser desvinculada dos mecanismos cognitivos que requer o processo tradutório, já que esses mecanismos são precisamente os que explicam as causas do erro. Para esta autora, o indivíduo pode cometer erros em diferentes fases: erros de compreensão, de ausência de desverbalização — que produz decalques — e erros de reexpressão, com uma seleção léxica ou morfosintaticamente equivocada, ou com a deficiência de mecanismos de coerência e coesão. Nós acrescentaríamos que, entre os equívocos de reexpressão, pode estar uma seleção equivocada das associações entre as palavras (como no exemplo que apresentaremos em seguida “via x rota de administração”). É preciso saber, por exemplo, se uma unidade como *to initialize the software*, utilizada na área da informática, se traduz, para o português, por *inicializar um programa* ou *iniciar um programa*. Assim, saber reconhecê-las é uma garantia a mais de que o texto será correto do ponto de vista linguístico e adequado do ponto de vista da temática tratada (Bevilacqua, 2005).

Observando pesquisas e métodos de avaliação de tradução, percebemos que, na avaliação da tradução profissional, conforme explicam Martinez Melis e Hurtado Albir (2001), além dos critérios de fidelidade e de qualidade, outros fatores, como eficácia e lucratividade entram em jogo. Nesse contexto, citamos também as planilhas usadas nas avaliações de traduções. No mercado das agências de tradução, essas planilhas, que visam

medir a qualidade e a adequação de um trabalho de tradução, são de grande importância. Em alguns casos, até mesmo descontos financeiros nos valores a serem pagos são aplicados às agências de tradução caso os tradutores não atinjam um determinado escore. Geralmente, essas planilhas são preenchidas por revisores ou tradutores mais experientes que fazem a revisão dos trabalhos.

Observando uma dessas planilhas de *feedback*, em que são apontadas correções e sugestões ao tradutor, podemos ter uma ideia do tipo de “armadilha” à qual o tradutor está sujeito. Entre os “erros” possíveis e avaliados em uma planilha, encontramos listadas falhas de diferentes naturezas. São problemas reconhecidos nesses instrumentos: erros de precisão (omissões, acréscimos, sentido incorreto, texto não-traduzido); de língua (pontuação, ortografia, erro de digitação, sintaxe); de terminologia (inconsistência, falta de compatibilidade com glossário fornecido); de estilo (geral, conformidade com guia do estilo do cliente); de leiaute (formato, texto oculto, tags, espaço); de regionalismos (padrões regionais/nacionais, adequação ao mercado local); de conformidade (instruções, glossários e guias de estilo).

Pensando em todos esses parâmetros, temos uma dimensão maior de como as possibilidades de insucesso em uma tradução podem ir muito além de equívocos ocasionados por uma lista estanque de falsos cognatos ou pela escolha errada de termos técnicos. Diante de uma relação de possíveis erros, como a que mencionamos, podemos entender melhor que o tradutor pode fazer escolhas “erradas” (ou inadequadas) ou deixar-se levar pelo que “parece mas não é”, tanto na seleção de um termo/palavra quanto nos planos sintático, semântico, pragmático, estilístico, cultural e formal.

Nos inspiramos no trabalho de Daniel Gouadec (1994) para trazer o critério de “risco”, associado também ao de “vantagem”, como elementos que perpassam o trabalho do tradutor e que justificam a relevância das associações sintagmáticas mesmo que elas não sejam altamente frequentes, pois uma tradução é um processo que envolve uma construção de um todo, que é um texto, e não apenas uma transposição de palavras isoladas. Em seu artigo *Natureza e tratamento das entidades fraseológicas*, Gouadec (1994) afirma que o domínio de uso das cadeias de caracteres de usos especializados dão indícios de competência técnica e linguística, ou seja, qualquer defeito de domínio da dita cadeia... seria interpretado como um indício de incompetência. Segundo o autor, uma cadeia de caracteres é notável devido a alguns fatores, entre eles o de “risco”.

Pym (2004) define risco como a possibilidade de o texto traduzido não atender aos seus propósitos e define que, no âmbito da tradução, existem elementos de alto e de baixo risco; a maioria deles fica entre os dois tipos. O autor cita como exemplo de alto risco um erro no nome dos pais ou na data de nascimento no caso da tradução de uma certidão de nascimento. O risco é entendido como o grau de probabilidade de uma opção em particular não levar à obtenção de uma condição de sucesso (Pym, 2003).

Para determinar esses riscos relativos, é necessário pensar na finalidade da tradução, no papel que ela desempenhará. Ele acrescenta ainda que os riscos não são necessariamente de natureza linguística. O autor defende a posição de que esses fatores são mais ou menos conhecidos pelos profissionais, mas aponta que as teorias de tradução não têm percebido o valor que a análise do risco pode lhes oferecer. Pym comenta que Gouadec (2002 *apud* Pym 2004), há mais tempo, elaborou uma visão em que os problemas de tradução são resolvidos, obtendo-se o máximo de informações possíveis antes do início da tarefa. Os esforços de investigação pré-tradução seriam uma maneira mais eficiente de gerenciar o risco do que a resolução dos problemas individualmente e a cada nova ocorrência. Pym acrescenta que as informações mais necessárias são aquelas relacionadas a opções de alto risco e que os tradutores deveriam obter o máximo possível de informações a respeito dessas opções antes de traduzir para eliminar o máximo de “adivinhação” possível e, portanto, reduzir o risco.

Conhecendo a linguagem do texto que deve traduzir, o que inclui saber como e quais palavras, geralmente, se associam, o tradutor, portanto, elimina alguns riscos e obtém algumas vantagens. O critério de vantagem é diretamente ligado ao de risco: tem mais vantagem no mercado de trabalho e êxito nas suas traduções os profissionais que conhecem e sabem como tratar elementos textuais que podem oferecer riscos de erros ou inadequações. Uma questão importante de ressaltar em nosso trabalho é que esse conhecimento da associação de palavras no nível sintagmático, de uma determinada área ou de um determinado gênero textual, o tradutor adquire com os anos de experiência. Assim, se pensarmos em tradutores iniciantes, entendemos que há uma necessidade de encontrar meios de propiciar referências que chamem a atenção para alguns desses elementos que apresentam riscos.

A relevância de associações que não se configuram especializadas, no sentido de não terem obrigatoriamente um termo em sua constituição, e que não são fraseológicas ou constituídas de *n-gramas*, no sentido de formarem uma sequência sem interrupções ou ordens invertidas (A*B, BA), por exemplo, se dá também pelo risco e pela vantagem de o tradutor

saber usá-las adequadamente no texto traduzido. O “estranhamento” e a falta de naturalidade podem, sim, ser considerados erros de estilo. Eles evidenciam uma falta de conhecimento a respeito do assunto, uma ingenuidade do tradutor e também um problema de adequação terminológica da área em questão. Tagnin (2002, p. 193) aponta que a ingenuidade do tradutor “se configura numa compreensão composicional do significado e numa falta de consciência do quanto uma língua é constituída dessas partes pré-fabricadas.” E, continua: “A ingenuidade do tradutor pode transparecer tanto na sua habilidade de compreensão, quanto na de produção.”

Além do que comentamos acima a respeito de erro e de inadequações em traduções em geral, no âmbito dos artigos médicos, é preciso lembrar que eles passam por critérios rígidos de avaliação antes de serem aceitos e publicados. As revistas especializadas mais conceituadas fazem julgamentos que levam em conta a conformidade dos artigos com os Requisitos Uniformes para Originais Submetidos a Revistas Biomédicas (ICMJE, 2008). Esses Requisitos começaram a ser elaborados em uma reunião de um grupo de editores de revistas da área médica, realizada em Vancouver, no Canadá, em 1978. Eles tinham o objetivo de estabelecer diretrizes para o formato dos originais submetidos a suas revistas. O grupo se expandiu até formar o Comitê Internacional de Editores de Revistas Médicas (International Committee of Medical Journal Editors – ICMJE), que se reúne anualmente. A cada reunião, os Requisitos podem ser atualizados. Neles são consensualmente definidos, para artigos de Pediatria e de outras áreas da Saúde, desde aspectos formais, como formatação de referências, espaçamento entre linhas, até as características relacionadas à composição do texto, como o tipo de informação que deve constar em cada seção, quem deve ser considerado autor do trabalho. A seguir, destacamos dois trechos dos *Requisitos*, a título de exemplificação, extraídos das instruções para redação da seção *Métodos*²⁹:

Descreva claramente a seleção dos sujeitos da observação ou experimentação (pacientes ou animais de laboratório, incluindo controles), incluindo os critérios de elegibilidade e exclusão e uma descrição da população de origem.

Defina os termos estatísticos, as abreviaturas e a maioria dos símbolos. Especifique o software empregado.

Todo esse “controle” sobre a produção dos textos acabou criando modos de apresentar as informações que acabaram determinando a linguagem dos artigos médicos. Nesse sentido,

²⁹ Uma versão em português dos Requisitos pode ser encontrada em: http://www.jped.com.br/port/normas/normas_07.asp

fazer traduções para o inglês que fujam a esse molde pode acabar até acarretando a recusa da publicação de um artigo.

Ao pensarmos nas associações sintagmáticas (ou nos outros conceitos que delineamos no Capítulo I) como fenômenos da língua em geral aos quais o tradutor deve estar atento para evitar os riscos e, conseqüentemente, os erros ou as inadequações, estamos voltados para o resultado concreto da tradução: uma realização linguística, um texto produzido por falantes de uma língua (que pode ser a materna ou não). No entanto, sob esse prisma não estamos enfocando o processo tradutório, nem a relevância ou a influência que ele tem sobre o texto/associações. Enfocamos apenas a produção textual que o tradutor realiza, independentemente de ela ser uma tradução ou não. Consideramos que, ao abordar as associações de nível sintagmático pela ótica do tradutor, aspectos relacionados ao processo tradutório devem ser incluídos. Só fazendo a reflexão sobre esse prisma é que podemos entender – e justificar – melhor a relevância das associações para a tradução.

2. A relevância com foco no processo: a teoria interpretativa e as unidades de tradução

Uma vez que queremos chamar a atenção para como a observação e a “leitura” do texto especializado e das associações, feitas sob o ponto de vista do tradutor, podem ser diferenciadas daquelas usualmente feitas por linguistas e terminólogos, cabe a nós trazer à discussão fatores relacionados ao processo tradutório. Nossa crença é de que essa “diferenciação de olhar para o texto” surge de um fator intrínseco a uma das etapas do processo, a reverbalização do sentido na língua de destino e a conseqüente resintagmatização das unidades lexicais.

É das necessidades que o tradutor tem para que possa cumprir a etapa de resintagmatização que pensamos que a maneira como ele aborda o texto — ou deveria abordar, em nosso entendimento — não faz recortes e nem elege *a priori* algumas unidades como mais “importantes” do que outras. Ao contrário, a necessidade de encadear as palavras no nível sintagmático desempenha um papel tão importante quanto saber escolher os itens adequados no eixo paradigmático (por exemplo, a escolha de um termo). Isso porque, além de conhecer e compreender os termos da área em que trabalha, o tradutor precisa conhecer as palavras da língua corrente, as “maneiras de dizer” nessa especialidade, ou seja, a *língua* e a *fala*. É da reunião desses dois que nasce a “linguagem”, corrente ou especializada, segundo o

uso que dela é feito, conforme bem explicitou Gémar (2000, p.23). Dessa forma, uma das maneiras de “recortar” uma unidade de tradução é pensando em seu encadeamento sintagmático.

2.1. O processo tradutório

Historicamente, os estudos em tradução revelaram diferentes perspectivas ao longo de seu desenvolvimento. Muito resumidamente, trilharam um caminho semelhante ao da Terminologia, da visão restrita de univocidade de termos, conceitos, equivalentes, a uma visão mais comunicativa, inserida no uso, voltada aos receptores e condicionada pelos seus enunciadores. Como bem ressalta Kilian (2007), dentro dessa evolução, os estudos de tradução “têm em comum discussões pautadas por dicotomias, como forma x conteúdo, fidelidade x criatividade, traduzível x intraduzível”.

Hurtado Albir (2001) apresenta de maneira muito clara as definições de tradução sob quatro perspectivas: da tradução como atividade entre línguas, como atividade textual, como ato de comunicação e como processo. Sem entrar em maiores detalhes, parece-nos claro que cada uma das definições têm sua porção de contribuição, o que ressalta ainda mais a complexidade que essa atividade encerra.

Por isso, o que nos interessa é nos posicionar neste contexto dentro daquelas correntes que consideram a tradução, principalmente a técnico-científica, como uma atividade de recriação de um texto fonte, intrinsecamente influenciada por fatores como interpretação e inserção do tradutor enquanto um sujeito do texto produzido.

Podemos falar de tradução como *produto* (o texto final, concreto) e como *processo* (a tradução enquanto atividade da qual resultará o produto).

De acordo com Azeredo (2007)

Como produto, a tradução focaliza o texto e a língua de trabalho do tradutor. Por outro lado, como processo, a tradução focaliza o tradutor como um produtor de texto, com uma interpretação particular do texto original, determinada por fatores externos que darão um sentido particular ao texto (p.59).

Considerando esses dois tipos de enfoque, podemos dizer que são condicionados pelas motivações, interesses e contextualização histórica de cada abordagem. Cada uma delas tem

dado a sua contribuição para o desenvolvimento do “fazer” e do “pensar” tradutórios. É importante ressaltar também que, embora se faça essa distinção entre abordagem de produto e de processo, ela acontece mais em termos de tendências de abordagem, ou até mesmo em termos didáticos, em termos de diferenciação das atividades envolvidas, uma vez que um não existe sem o outro. Portanto, considerações exclusivas acerca de um ou de outro são praticamente impossíveis.

Para atingir o fim a que nos propomos nesta tese, precisamos definir como entendemos que se configura o processo de tradução e por que ele condiciona i) o modo como o tradutor aborda o texto especializado e ii) o que o tradutor necessita para produzir um texto menos literal³⁰ e mais voltado para as condições de chegada. Nesse segundo item, quando pensamos em um tipo de material de consulta que atende às necessidades de tradutores, estamos referindo um material voltado para o texto e para as combinações entre as palavras “o quê se usa com quê” e “como se diz”. Em outras palavras, materiais que contemplem as convenções linguísticas e textuais que devem ser atendidas no texto traduzido.

Pensamos, também, que estudos empíricos em tradução que *declaradamente* produzem conteúdo de fundo teórico, considerando o ponto de vista da prática da tradução, não são muito frequentes. Parecem ser mais frequentes os trabalhos que se empenham em descrever estratégias de tradução ou características dos textos traduzidos, ou então que mencionam contribuições para a prática e o ensino da tradução, do que estudos que produzam contribuições teóricas para a Linguística ou para a Terminologia. Um bom exemplo de estudo que realmente leva um ponto de vista do tradutor ou profissional da linguagem para um desenvolvimento teórico acerca da linguagem em geral é o já citado trabalho de Daniel Gouadec (1994), intitulado *Natureza e tratamento das entidades fraseológicas*³¹.

Não se pode deixar de valorizar, no entanto, os muitos trabalhos que costumam ser considerados “com olhar de tradutor”, porque incluem em seu foco constituintes textuais diferentes dos chamados termos técnicos³², ou que se associam a esses termos. O interesse por estruturas textuais mais amplas do que os termos, como as fraseologias especializadas (Bevilacqua, 2004), e até mesmo por outros elementos textuais — como colocações

³⁰ Em se tratando do contexto que estamos enfocando neste trabalho, de traduções de textos especializados, mais especificamente de artigos médicos, a literalidade não se caracteriza como um fator positivo nos textos traduzidos.

³¹ Título original *Nature et traitement des entités phraséologiques*.

³² Aqui nos referimos a *termos* no sentido mais popularmente conhecido, o qual está mais em conformidade com a Teoria Geral da Terminologia, que considera que um *termo* é sempre uma associação a um conceito.

especializadas (Zílio, 2009), expressões anunciadoras de paráfrase (Azeredo, 2007) e marcadores textuais e discursivos (Possamai, 2004)³³, verbos modais (Sarmiento, 2009) parece ser uma constante em estudos que exibem alguma interface com a Tradução e o ensino de línguas.

Na seção que segue, apresentamos um panorama de trabalhos em tradução que enfocaram a atividade cognitiva do processo de tradução e, posteriormente, deteremo-nos em um deles, a Teoria Interpretativa da Tradução.

2.2. Concepções sobre o processo tradutório

Hurtado Albir (2001) considera que a tradução é um ato de comunicação, uma operação textual e uma atividade de um sujeito e, neste sentido, é preciso considerar o processo mental que esse sujeito desempenha durante a tradução. Quando falamos em estudos do processo tradutório, devemos entender estudos que se voltam para os caminhos cognitivos seguidos por esses sujeitos durante o processo, e não estudos que avaliam aspectos que cercam o ato comunicativo ou as operações realizadas durante a tradução (Hurtado Albir 2001, p. 312).

Segundo Bell (*apud* Hurtado Albir, 2001), há duas correntes de análise do processo tradutório. Uma corrente de estudos baseados no produto, que compara o texto traduzido e o original e que usa as diferenças encontradas para inferir, indiretamente, os processos mentais empregados, e uma outra corrente, que usa métodos como os TAPs (*think-aloud protocols*)³⁴, também conhecida como abordagem processual. Deteremo-nos, brevemente, nesta última, a título de contextualização.

O trabalho de Hans Krings, intitulado *Was in den Köpfen von Übersetzern vorgeht* [O que se passa na cabeça de tradutores], de 1986, é visto com unanimidade pelos pesquisadores da abordagem processual como o marco inicial que dá visibilidade aos trabalhos sobre o processo de tradução como objeto de estudo específico, nos estudos de fundamentação cognitiva.

³³ Todos trabalhos conduzidos e já defendidos na linha de pesquisa Terminologia e Lexicografia: relações textuais, do PPG-Letras/UFRGS.

³⁴ Os protocolos verbais são registros de observações e falas de usuários que servem como material de investigação de processos mentais.

Nesses trabalhos,

dados coletados por meio de estudos empíricos com sujeitos em situação de trabalho, por meio de protocolos verbais (*think-aloud protocols*) e outros tipos de protocolos verbais, registros de arquivos e outros métodos, são tomados como uma fonte de investigação de decisões e estratégias usadas por tradutores para solucionar problemas. (Alves *et al.*, 2002, p. 168)

Conforme explica Rodrigues (2002), um resultado consistente dessa abordagem é que “experiências lingüísticas e tradutórias que tradutores acumulam no exercício da profissão influenciam de maneira massiva a forma como eles traduzem”. Essa comprovação

reforça a tese de que os tradutores não são seres passivos durante o processo da tradução, como estava pressuposto em modelos mais antigos do processo da tradução. Muito pelo contrário, é a sua participação ativa através de suas atitudes, crenças, utilização de inúmeras estratégias e processos criativos diversos que contribui para a construção do processo da tradução. (Rodrigues, 2002, p. 25)

Este autor, em artigo que faz uma revisão dos trabalhos na área, acrescenta que muitos estudos foram feitos dentro da abordagem processual da tradução, mas apenas alguns deles elaboraram modelos do processo da tradução. O autor credita esse fato ao alto grau de individualidade dentro do processo de tradução. Mesmo assim, cinco modelos são referidos, dos quais destacamos dois: o de Königs (1986,1987) e o proposto por Alves (1995), que é uma ampliação do modelo de Königs. De acordo com Rodrigues (2002), este modelo descreve basicamente a interação entre duas categorias de processamento: *Adhoc-Block*, para processos automatizados e *Rest-Block*, para aqueles processos, na maioria das vezes estratégicos. O interessante dessa abordagem são justamente os enfoques nessas duas categorias. Na categoria *Adhoc-Block*, são processadas as unidades de tradução que têm uma correspondência semântica semelhante no outro idioma; e, quando não são processadas automaticamente, são transferidas para a categoria *Rest-Block* e são “processadas de acordo com algumas variáveis, como por exemplo fatores contextuais, problemas estilísticos, mensagem do autor, receptores da tradução, entre outras” (Rodrigues, 2002, p.42).

Alves (1995) propôs uma expansão do modelo da tradução desenvolvido em Königs, sugerindo uma “terceira categoria de processamento dentro da qual são processadas informações influenciadas pelo conceito da relevância” (Rodrigues, 2002, p.44). Para Alves (1995), o conceito da relevância no processo da tradução pode ser interpretado como uma característica do processamento relacionada com uma contextualização mais adequada de determinados processos comunicativos.

Hurtado Albir (2001) também faz uma revisão de trabalhos sobre processo de tradução, citando como principais os modelos interpretativo (Seleskovitch, 1978) e outros, como o modelo linguístico e psicolinguístico de Bell (1991); o modelo sociológico de Kiraly (1995); a tradução como comportamento cognitivo (Wilss, 1988); a aplicação da teoria da relevância de Gutt (1991); os modelos de esforços de Gile (1995) e o processo de compreensão segundo Dancette (1995).

O que nós buscamos, nesta tese, nos estudos de processo de tradução, não foi uma análise de nível tão específico como a dos estudos cognitivos; do que se passa na mente do tradutor a cada escolha. Embora existam todos esses trabalhos mencionados, optamos por uma teoria que, desde o primeiro momento, nos cativou pela clareza e pelo didatismo quanto ao processo da tradução além de, é claro, dar sentido a nossa preocupação com o momento de reexpressão do sentido. Encontramos respaldo, então, na descrição do processo elaborada na Teoria Interpretativa da Tradução, revisada a seguir. Selecionamos essa teoria de descrição do processo a fim de amparar nossa crença na existência de uma etapa de resintagmatização de unidades durante o processo de produção do texto na língua alvo.

2.2.1. A Teoria Interpretativa da Tradução

A Teoria Interpretativa da Tradução foi inicialmente desenvolvida pela pesquisadora e intérprete Danica Seleskovitch e por sua equipe. Seu foco, como diz o nome, deu-se primeiramente na atividade de interpretação, ou tradução oral. Posteriormente, foi aplicada à tradução escrita, no trabalho de Marianne Lederer (1994). Esta autora considera que as constatações inicialmente feitas pela Teoria Interpretativa acerca da atividade de tradução oral aplicam-se também à tradução escrita, e que a busca do sentido e sua reexpressão constituem um denominador comum a todas as traduções:

A teoria interpretativa (...) estabeleceu que o processo consistia em compreender o texto original, desverbalizar sua forma lingüística e expressar em outra língua as idéias compreendidas e os sentimentos experimentados. Essa constatação, feita inicialmente a respeito da tradução oral, ou interpretação de conferências, aplica-se também à tradução escrita. (Lederer, 1994, p. 11)

É importante que se chame a atenção aqui para a “coincidência” entre os conceitos de “interpretar” enquanto atividade de “tradução oral” (por exemplo, interpretação de conferências, de reuniões entre chefes de estado de países diferentes, etc.) e “interpretar” como dar um sentido próprio ou novo a um texto ou a um fato. A Teoria Interpretativa, em

sua origem, era voltada à interpretação no sentido de “tradução oral”. Já no início do século XIX, Schleiermacher (2001) abordou o fato de, na tradução escrita, também existir um tanto de interpretação (no segundo sentido acima exposto), fazendo uma oposição entre traduzir/interpretar (no sentido de interpretação de uma “mensagem”) e traduzir/transpor, que seria uma tradução mais “fria” e unívoca, sem “interferência” do tradutor.

Sobre interpretar em tradução, outro autor que vem tecendo considerações a respeito é Umberto Eco (2007). Embora esteja mais voltado para traduções literárias, há que se considerar que sempre há uma interpretação quando se traduz. Diz ele (2007)

Interpretar significa fazer uma aposta sobre o sentido de um texto. Esse sentido – que um tradutor pode decidir identificar – não está encerrado em algum mundo incorpóreo [...]. É apenas o resultado de uma série de inferências que podem ou não ser compartilhadas por outros leitores. (p.181)

Não é essa interpretação a que se referia a Teoria Interpretativa em seu início, embora se reconheça que o ato de traduzir sempre vem acompanhado de uma dose de interpretação, no sentido exposto acima por Eco.

No que diz respeito às etapas do processo, a teoria é pautada pelo entendimento de que o processo de tradução é uma sucessão de três estágios:

- i. Percepção de um enunciado linguístico que é portador de significado. Nesse estágio, é feita a apreensão da língua e a compreensão da mensagem por meio de um processo de análise;
- ii. Abandono imediato e intencional das palavras e retenção da representação mental da mensagem (conceitos, ideias, etc.). É o processo de desverbalização;
- iii. Produção de um novo enunciado na língua-alvo, que deve atender a dois requisitos: deve expressar a mensagem original completa e deve ser voltado para o destinatário. Este é o processo de reverbalização.

Resumindo, para a Teoria Interpretativa, traduzir é um processo que consiste em compreender o texto original, desverbalizar sua forma linguística (reter o sentido) e expressar numa outra língua as ideias compreendidas e os sentimentos sentidos.

Na primeira etapa, a percepção do enunciado pode ser feita tanto pela forma auditiva como pela leitura. Uma vez apreendida a mensagem por meio de sua forma linguística, ela será analisada e compreendida para que se chegue ao sentido, por meio de uma fusão do significado linguístico das palavras e frases com os “complementos cognitivos”. Esses complementos, segundo Lederer (1994, p. 56), são de três tipos básicos: contexto verbal, contexto situacional e contexto cognitivo.

Vamos tratar de cada uma das etapas da tradução para melhor explicitar o modelo descrito.

2.2.1.1. Primeiro estágio: percepção do enunciado

Esta é a etapa da compreensão. Segundo Hurtado Albir (2001), Seleskovitch e Lederer têm uma preocupação constante com o funcionamento da compreensão, acreditando que ela não depende somente do elemento linguístico em questão em uma tradução escrita ou oral, mas também da intervenção de uma série de conhecimentos, os complementos cognitivos. Através da leitura de um texto ou da audição de uma fala/conversa, um “saber” é acrescentado à bagagem cognitiva/conhecimento de mundo do tradutor. Todo esse “saber” se desverbaliza e permanece presente na memória, ajudando o tradutor a compreender o texto. A partir daí, segundo Hurtado Albir (2001), se produz uma interação entre a cadeia linguística e os complementos cognitivos. Para as autoras da teoria, a memória tem um papel de destaque no processo de compreensão, e elas consideram que todo processo de compreensão gera uma interpretação.

O processo da interpretação envolve a percepção de ideias, ou sentido, expressas no discurso. À medida que se percebe o sentido, as formas verbais utilizadas para transmiti-lo desaparecem, deixando apenas a consciência, a partir da qual o intérprete pode espontaneamente expressar o sentido, sem estar preso à forma da língua de partida (Seleskovitch e Lederer 1995 *apud* Pagura, 2003). Esse distanciamento da forma da língua ocorre na desverbalização, a etapa seguinte.

2.2.1.2. Segundo estágio: desverbalização

O resultado da compreensão realizada na etapa 1 tem um caráter não-verbal. Conforme esclarece Hurtado Albir (2001), a desverbalização é resultado da fase de compreensão e início

da fase de reexpressão. Esta etapa parece refletir bem a origem “oral” da teoria, em que o processo de interpretação como atividade profissional serviu de base. As autoras acreditam em uma dissociação entre as palavras e o que as palavras transmitem, e que o produto é resultado do sentido compreendido e não das palavras pronunciadas. O sentido é uma totalidade, em que há uma interdependência de elementos linguísticos e não-linguísticos. Porém, como há limite para as formas linguísticas e ausência de limite para os sentidos, o sentido, ao fim e ao cabo, é uma síntese não-verbal. O sentido resulta da desverbalização da cadeia sonora (ou gráfica) no momento em que os conhecimentos linguísticos e os complementos cognitivos se fundem. O sentido corresponde a um estado de consciência. (Lederer, 1994).

2.2.1.3. Terceiro estágio: reverbalização ou reexpressão

É o momento em que o tradutor e o intérprete dão uma nova feição à mensagem já compreendida. É a reexpressão do sentido com os recursos de outra língua. Como claramente dito por Seleskovitch (1978 *apud* Pagura 2003), o novo enunciado deverá atender a dois critérios básicos: a mensagem original deve ser completa, provida de todos os detalhes e deve refletir as características da língua de chegada. Ao se traduzir do inglês para o português, por exemplo, o texto alvo, quer seja ele escrito ou oral, deverá parecer ter sido produzido originalmente em português, sem traços que denotem sua origem em uma outra língua³⁵. A reexpressão consiste em conservar a identidade de conteúdo, escolhendo uma equivalência de forma, em um processo que vai do não-verbal (desverbalização) à verbalização em uma língua natural e que se assemelha ao processo de expressão na comunicação monolíngue (Hurtado Albir, 2001).

A Teoria ainda se posiciona quanto à *correspondência* e à *equivalência* em tradução, bem como deixa transparecer uma oposição entre *língua* e *discurso*. Comentaremos esses aspectos na próxima seção deste capítulo, a Discussão, em que evidenciamos como a Teoria Interpretativa dá respaldo a nosso trabalho. Nos pontos de insuficiência identificados e/ou dos quais discordamos, acrescentamos o nosso ponto de vista.

³⁵ Esse é um ponto de vista bastante delicado acerca do texto traduzido. Os textos traduzidos devem assim ser no contexto da tradução de artigos médicos, mas alguns autores discutem se o texto deva parecer uma tradução ou não.

3. Discussão

Os três estágios descritos pela Teoria Interpretativa da Tradução representam o modo como entendemos o processo de tradução. Existe um primeiro momento de compreensão e outro de reexpressão e, entre esses, um segundo estágio, de retenção do sentido. Notamos que esse segundo estágio parece ser bastante determinado pelo caráter de oralidade da Teoria Interpretativa em sua origem (voltada para a interpretação).

Pensamos que, na atividade de interpretação oral, a desverbalização é mais intensa pela natureza do processo envolvido, em que há um imediatismo e uma necessidade instantânea de se passar para o próximo passo (na interpretação as três fases ocorrem quase que ao mesmo tempo). Há, também, uma desvinculação com as representações materiais do texto escrito desde a primeira fase do processo. Podemos argumentar que toda a interpretação oral se dá em um plano muito mais abstrato do que o da tradução escrita. Mesmo assim, não negamos que a desverbalização também esteja presente na tradução escrita, como as próprias autoras da teoria consideram, e também entendemos que ela ocorre quando o tradutor tenta apreender as ideias ou sentidos. Nosso argumento, no entanto, é que não é possível eliminar totalmente a influência das formas textuais concretas do texto fonte no texto alvo³⁶.

Nesse ponto, enfatizamos dois aspectos que entendemos inerentes ao processo da tradução e que queremos destacar pois achamos que merecem atenção, ainda dentro da Teoria Interpretativa. O primeiro é esse que já mencionamos, de que a retenção do sentido/desverbalização, na tradução escrita, não consegue se distanciar da forma verbal/gráfica, ou do significante, porque o texto fonte está muito presente no processo, assim como está a necessidade de reexpressá-lo também de uma forma gráfica. Quando o tradutor se pergunta “O que isto quer dizer?” e processa apenas o sentido, a preocupação com a reexpressão, com o entender por que foram selecionados determinados itens léxicos na língua de partida, ou se eles carregam alguma marca de estilo, alguma rima ou trocadilho, não é possível fazer uma desverbalização completa. Devemos lembrar que traduzir não é criar um texto do zero. Traduzir é criar com base em uma sequência de caracteres/formas gráficas em uma sintaxe do texto e da frase que não se pode desconsiderar.

O segundo ponto que queremos ponderar é que a interferência do texto fonte no texto alvo é hoje algo que pode ser discutido. A etapa de reexpressão como ausente de traços que denotem sua origem em uma outra língua, conforme coloca a Teoria Interpretativa, pode ser

³⁶ Usamos essa designação, neste trabalho, porque é a usada no mercado da tradução, em geral. Provavelmente porque o software Trados utiliza a denominação *source* e *target text*.

questionada. Alguns autores (Venutti, 1995) pensam que a tradução não deve ser “transparente”, isto é, sendo uma tradução, ela não tem como não parecer uma tradução. Há autores que consideram que a tradução é um “terceiro código”, não sendo nem a língua de partida, nem uma ocorrência natural da língua de chegada. Compartilhamos do ponto de vista de que a tradução jamais será uma ocorrência tão natural como o texto produzido diretamente na língua de chegada, o que não significa desmerecê-la, pelo contrário, significa enaltecê-la enquanto tradução. No entanto, é preciso dizer também que entendemos que isso não “exime” o tradutor de ter que atender às expectativas da audiência do texto traduzido e isto, muitas vezes, significa chegar o mais próximo possível daquele texto produzido diretamente na língua alvo.

Ao discutir as etapas de desverbalização e reexpressão, estamos em uma fronteira com o tema da literalidade/não-literalidade. Afirmar que a desverbalização é menor equivaleria a dizer que o texto produzido na etapa seguinte, da reexpressão, seria mais literal. Partimos do princípio de que o texto fonte exerce grande influência sobre o texto alvo, portanto, se o que se deseja é que a mensagem original seja completamente transmitida e voltada para o destinatário, é preciso ser menos literal na etapa da reexpressão, mesmo que exista essa influência constante do texto fonte.

Consideramos também que podem haver alguns fatores que exerceriam uma influência na desverbalização, como as origens das línguas envolvidas, se pertencem a ramos próximos ou distantes (por exemplo, entre espanhol e português a desverbalização seria menor do que entre português e finlandês, por exemplo), a qualidade da redação do texto de origem, a dificuldade da matéria, o conhecimento que o tradutor tem do assunto, o prazo de entrega do trabalho. Aqui também há que se considerar que o tradutor é inicialmente e, antes de tudo, um leitor e que, esse estágio, de leitura, não é 100% infalível e, portanto, um erro de leitura afetaria todas as três etapas seguintes do processo.

Os pontos que discutimos sobre literalidade também se relacionam a outros conceitos abordados pela Teoria Interpretativa, os conceitos de equivalência e de correspondência. A teoria é pautada por uma separação entre língua x discurso, em que a linguagem é abordada por uma ótica mais discursiva do que linguística. Para Seleskovitch e Lederer, a tradução está no plano da comunicação, do sentido e, por isso, são enfatizados os aspectos comunicativos. Fazendo essa separação, as autoras estabelecem que a tradução interpretativa é feita por equivalência, e a tradução linguística, por correspondência.

A tradução por correspondência é aquela em que a relação é estabelecida entre os *elementos lingüísticos* (palavras, sintagmas, formas sintáticas) e a tradução por equivalência é aquela em que o sentido é o mesmo, mas as palavras raramente são correspondentes. Segundo Lederer (1994) “Toda tradução comporta evidentemente correspondências entre termos e vocábulos, mas se torna texto somente graças à criação de equivalências” (p.36).

Embora, para Hurtado Albir (2001, p. 41), “a tradução não se situa no plano da língua, mas no plano da fala” e para Lederer (1994) traduzir é um ato de comunicação e não linguístico³⁷, nós entendemos que é preciso iluminar também o elemento linguístico, que inegavelmente se faz presente (fato que a Teoria Interpretativa não nega, mas parece deixar de lado). E, não apenas está presente como é o objeto do trabalho do tradutor, que atua manuseando, a nosso ver, sentidos e palavras. Sendo assim, ao mesmo tempo que temos uma identificação muito grande com a Teoria Interpretativa, também queremos iluminar também o elemento concreto da tradução (o texto).

Nossa crença é a de que, na etapa de reverbalização do sentido na língua de destino, o tradutor refaz associações no plano sintagmático, as quais foram desfeitas no momento da desverbalização. Apesar dessa crença parecer estar relacionada somente com o plano da língua, entendemos que, se o novo enunciado deve atender a dois critérios básicos — a mensagem original deve ser completa, provida de todos os detalhes, e deve refletir as características da língua de chegada —, é no momento da reexpressão que ocorre uma preocupação com a sintagmatização na língua alvo, pois acreditamos que é então que se refazem as “características” da língua de chegada. Essas características, portanto, não vão ser atendidas se o tradutor se mantiver realizando uma tradução somente por correspondência de conceitos, mas, sim de modos de dizer, pois a equivalência se dá no nível do sentido. Por fim, destaca-se ainda, da Teoria Interpretativa, que o momento da compreensão é um momento que exige leitura. Consideramos que o processo de leitura na tradução é fundamental, pois é também nesse momento que se evita o erro de leitura, em que se evita o tradutor fazer uma leitura equivocada do texto fonte. Esse “erro” não deve ser confundido com uma segunda interpretação possível. Um exemplo de erro de leitura é apresentado a seguir. Neste exemplo, *through* foi traduzido como *por meio*, quando na verdade ele é parte da expressão *from x through y*, que significa *de x a y*.

³⁷ Talvez o elemento linguístico tenha sido deixado mais de lado na Teoria Interpretativa quando em referência aos modelos da linguística estrutural ou gerativa.

BPM must address the full process lifecycle, **from** discovery and design **through** execution and monitoring

O BPM deve cobrir o ciclo de vida completo do processo, **detecção e projeto por meio de execução e monitoramento**

O BPM deve cobrir o ciclo de vida completo do processo, **da detecção e projeto à execução e monitoramento**

Seleskovitch e Lederer (1995) destacam nessa fase o papel da memória, mais uma vez por estarem embasadas na tarefa de interpretação. Nós destacamos a leitura. Deslile (1988) assinala que a compreensão do texto original pela leitura também é um processo interpretativo de captação de sentido.

Sendo assim, entendemos que a Teoria Interpretativa nos oferece um modelo muito claro e didático para o que acontece no processo de tradução, além de privilegiar as etapas do processo, entre elas a de reverbalização, que nos interessa particularmente. Ao mesmo tempo, reconhecemos que ela é também muito abstrata. Por isso, damos luz para pontos diferentes na teoria: enquanto ela privilegia o discurso, nós privilegamos a língua (sem desconsiderar o discurso, mas entendendo que o discurso, na reverbalização, é materializado pela língua). E, enquanto ela privilegia a memória, nós privilegamos a leitura.

CAPÍTULO III

O PROCESSO DE SINTAGMATIZAÇÃO E AS UNIDADES DE TRADUÇÃO

A tradução que a Teoria de Seleskovitch e Lederer denomina *interpretativa* e a tradução *linguística* estão mais em uma condição de complementaridade do que de oposição. Entendemos a tradução linguística como fator inerente ao processo, isto é, negar a tradução no nível das palavras pode ser muito interessante em aspectos teóricos, mas, na prática, as formas gráficas são o ponto de partida do tradutor. Por outro lado, fazer uma tradução “somente linguística” também não é desejável, porque aspectos culturais e de incompatibilidade com a expectativa do público do texto alvo também podem ocorrer. Acrescentamos aqui um exemplo para ilustrar essa situação. Ao revisarmos uma tradução de um texto médico do inglês para o português, encontramos a frase *It includes medical, dental and vision coverage* traduzida como *Ele inclui cobertura médica, dentária e de visão*. Claramente, a escolha por *visão* foi influenciada pela palavra *vision* do inglês, mas nos parece que em uma frase como essa, em que o contexto tratava de um plano de saúde, *vision* poderia ter sido traduzida por *oftalmológica*, que nos parece a palavra mais adequada em português, nesse contexto.

Acreditamos que, na teoria, ambas podem ser vistas como processos com métodos distintos, mas, na prática, o tradutor faz uso das duas “técnicas”. Sabemos que, para realizar esses processos de maneira bem sucedida, é necessário mais do que o conhecimento da gramática e do vocabulário dos idiomas envolvidos. No entanto, é inegável que traduzir é uma operação que envolve um movimento entre dois sistemas linguísticos. A Teoria Interpretativa esquematiza seu modelo, considerando que, na etapa um, a de compreensão, há uma transformação das cadeias sonoras (ou gráficas) em sentido, e na etapa três, a de reexpressão,

há a transformação do sentido em cadeias sonoras (ou gráficas). Aqui é preciso mencionar uma crítica à noção de que o sentido está separado das cadeias sonoras, pois não há maneira de esquematizar o pensamento sem associar a imagem mental com a cadeia sonora. Mesmo assim, é possível entender que é uma etapa em que se procura, ao menos, abstrair a cadeia sonora.

Considerando, então, os aspectos linguísticos envolvidos no processo de tradução (desconsiderados pela Teoria Interpretativa), entendemos que, quando traduz, ou, especificando, quando reexpressa o sentido, o tradutor de textos escritos, invariavelmente precisa fazer escolhas no nível paradigmático e no sintagmático, como exemplificaremos nesta seção. Nesse processo de escolhas, há uma possibilidade de o tradutor “recortar”, como unidades de tradução, associações sintagmáticas entre as palavras quando busca a forma de reescrevê-las na língua de chegada. Como essas associações não se encaixam em uma definição de termo ou fraseologia especializada (por não conterem um termo) e como o nosso enfoque, como dissemos, é pautado pelo ponto de vista do tradutor, acreditamos que elas se constituem como Unidades de Tradução Especializadas, em contraponto com as unidades de tradução da língua geral.

1. As associações sintagmáticas e a tradução

Em Medicina, por exemplo, em uma tradução no sentido Inglês-Português, o tradutor pode traduzir *administration route* por *rota de administração* ou *via de administração*. Sob a perspectiva do nível paradigmático, ele pode fazer a escolha por *rota* ou *via*, já que ambas palavras carregam conteúdo com sentido de “meio”. Em uma perspectiva sintagmática, isto é, do encadeamento com as palavras vizinhas, em Português, o item *via*, por ser mais comum na linguagem médica, além de mais frequente, seria mais adequado. É mais adequado porque a convencionalidade entre os itens *via* e *administração* é maior do que entre *rota* e *administração*. Essa convencionalidade pode vir a receber outros rótulos, conforme vimos no Capítulo I, como o de uma colocação, por exemplo. Porém, consideramos aqui *convencionalidade* como definido por Tagnin (2005):

As convenções linguísticas são os “jeitos” aceitos pela comunidade que fala determinada língua. Assim podemos chamar de convencionalidade o aspecto que caracteriza a forma peculiar de expressão numa dada língua ou comunidade linguística.(p. 14)

É na busca pela “forma peculiar de expressão” que consideramos que o tradutor “desmonta” associações sintagmáticas e paradigmáticas na língua de partida para “montá-las” em um outro idioma. E, aqui, ressaltamos, novamente, que há uma fusão entre a tradução interpretativa e a tradução linguística: para expressar o que interpretou, o tradutor se vale do sistema gráfico, sendo seu ponto de partida as palavras da língua de partida. Essa noção é um pouco diferente da colocada pela Teoria Interpretativa, pois a Teoria considera que a etapa de desverbalização elimina totalmente o vínculo com o texto fonte, diferente de nossa opinião. É necessário lembrar que diversos fatores exercem sua influência e determinam o processo de reexpressão: a interpretação do texto por parte do tradutor, o objetivo do texto, as solicitações do cliente, os prazos, o formato do texto (em casos em que há espaços delimitados para a tradução), entre outros.

Quando falamos nessa etapa de *expressar em uma outra língua*, voltamos ao ponto das associações. Tomando-se as duas línguas envolvidas como dois sistemas diferentes e não-relacionados nem interdependentes (a não ser por características que as unem por questões de origem, i.e., latina, anglo-saxã etc.), cada um com suas devidas associações possíveis nos planos sintagmático e paradigmático, podemos dizer que o tradutor vai procurar reconstruir o sentido que depreendeu do texto fonte, expressando-o com os recursos linguísticos disponíveis da língua de chegada. Para nós, esse caráter se dá mais por uma relação de associações do que por itens isolados e “recortados” do texto. Esse trabalho de “degravação” ou “desconstrução” de um código composto por recursos de um idioma (recursos fonológicos, morfológicos, sintáticos, semânticos e pragmáticos), para que seja recriado em uma segunda língua (com toda a mesma gama de recursos acima mencionados), confere a todos os itens lexicais do texto, *a priori*, um nível de igualdade em sua “relevância” para a composição do texto alvo.

É preciso fazer uma ressalva, porém, quanto à forma que consideramos as línguas envolvidas na tradução como sistemas totalmente separados. Isso é verdadeiro apenas enquanto olhamos para elas em separado, como por exemplo, enquanto observamos produções na língua fonte e na língua alvo, independentes, com vistas a fazer pesquisas e a obter referências por meio de comparações. Porém, no momento em que as duas línguas são unidas pelo processo de tradução, quando a língua de chegada torna-se um produto do processo de traduzir, esse produto não pode mais ser visto de maneira ingênua, no sentido de que é praticamente impossível excluir dele as influências que a língua da partida exerce.

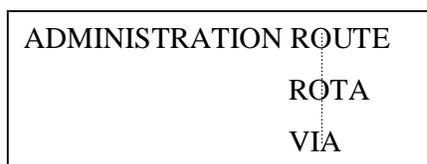
Lembramos que este estudo não está tratando de analisar o produto, isto é, a tradução como um “inevitável” terceiro código que surge a partir da consideração de um código matriz e outro alvo (Frawley, 1984, p. 168). No que diz respeito a essa questão, estamos tratando de apresentar uma forma de oferecer ao tradutor subsídios para realizar seu trabalho com base em realizações naturais da língua de chegada, contribuindo para que o tradutor desenvolva ainda mais sua competência tradutória na produção do texto traduzido. Ao atingir nosso objetivo prático, estaremos oferecendo exemplos de tradução que funcionarão como parâmetros de referência para uma tradução.

Sobre a noção de “terceiro código”, Batalha & Pontes (2005) esclarecem que o conceito

não aponta para “vícios de tradução” ou um “jargão de tradução” que revelariam uma conotação pejorativa, mas ... isto ocorre porque a tradução é uma forma específica de comunicação e que o tradutor se deixa, de certo modo, permeiar pelas estruturas da língua da qual traduz. (p. 40)

Esclarecidos esses pontos, voltamos às associações da cadeia sintagmática. Entendemos que elas parecem representar mais as escolhas que o tradutor faz para adequar os “modos de dizer” de uma determinada comunidade ou de um tipo de texto. Já no nível paradigmático, as escolhas estariam mais “coladas” no texto fonte, algo mais próximo da noção da tradução “palavra por palavra”, ou da tradução linguística.

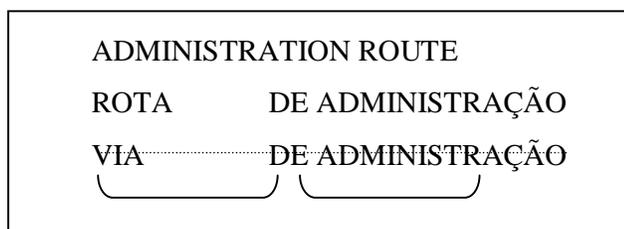
Exemplificando, como já dissemos, no contexto da tradução de um artigo médico, a combinação *via de administração* é mais apropriada do que *rota de administração*. A escolha que o tradutor faz ao traduzir *route* por rota ou via envolve uma seleção no nível paradigmático, isto é, uma escolha da palavra que preenche, em português, a ideia de *route*. Isso é o que busca ilustrar o esquema a seguir:



Esquema 1 – Possibilidades de tradução para a palavra *route* no eixo paradigmático

Neste momento, ele poderia escolher tanto *rota* como *via*. Já a decisão por uma das duas envolve uma seleção no plano sintagmático, um conhecimento da associação que mais

frequentemente está combinada à sequência de caracteres³⁸ *de administração*. No caso deste exemplo, seria mais adequado que o tradutor escolhesse *via*, dado o uso na área em questão.



Esquema 2 – Seleção, no eixo sintagmático, da associação mais adequada

No processo de escolhas de associações realizado no momento da tradução, poderíamos dizer que, no nível das escolhas paradigmáticas, são úteis os materiais de referência e de apoio como os dicionários e glossários. Porém, no nível das escolhas sintagmáticas, de associações, o tradutor precisa valer-se de outras referências. A sempre proclamada carência, por parte de tradutores, de materiais de referência satisfatórios nesse nível, é hoje suprida pela grande disponibilidade de textos na Internet e pelo próprio Google, que funciona como um comprovador de que determinada sequência de caracteres pode ser usada na língua de chegada. Além disso, notamos cada vez mais recursos com base em *córpus* sendo disponibilizados, como abordaremos na Parte II desta tese.

A seguir apresentamos mais alguns exemplos do que estamos considerando como escolhas no nível paradigmático e sintagmático em uma tradução de um resumo de um artigo médico.

Exemplo:

Verificar o comportamento dos valores da hemoglobina e a prevalência de anemia entre lactentes de termo de 3 a 6 meses de idade em aleitamento materno exclusivo.

To verify the behavior of hemoglobin levels and anemia prevalence in full term infants, aged 3 to 6 months and on exclusive breastfeeding.

Neste trecho, poderíamos destacar que a seleção de *levels* em vez de *values* para traduzir *valores* foi uma seleção de nível paradigmático, mas determinada pelo sintagmático,

³⁸ Estamos usando a definição *sequência de caracteres*, conforme utilizado por Gouadec (1994) e exposto no Capítulo II desta tese.

que levou em conta o uso do item mais adequado, conforme uso no contexto, para associar-se a *hemoglobin*. Destaca-se aqui que uma simples busca na ferramenta Google já nos apresenta pistas de que *levels* é a escolha mais adequada do que *values*. A combinação *hemoglobin values* retorna 118.000 *hits*, sendo que o terceiro e quarto *hits* são de ocorrência do sintagma em *sites* brasileiros. A combinação *hemoglobin levels* retorna 502.000 *hits*, sendo que o primeiro site brasileiro a ser apresentado está em décimo lugar. Mesmo que ambas combinações sejam possíveis, a pequena amostra gerada on-line sugere, tanto quantitativa como qualitativamente, que a segunda oferece menos risco de estar inadequada.

A escolha de *hemoglobin* para *hemoglobina*, por exemplo, é uma escolha que se dá completamente no nível paradigmático, enquanto *full term infants* para *lactentes de termo* pode ser considerada nos dois níveis: paradigmático, porque podemos considerar que há uma equivalência entre os sintagmas em português e inglês, e sintagmático, pois a escolha por *infants* como tradução de *lactentes* está condicionada por um conhecimento de que o sintagma é *lactentes de termo* em português, e *full term infants*, em inglês. Percebe-se nesse caso que não há uma equivalência entre as línguas do traço que envolve a amamentação (claramente percebido em português, pelo uso da palavra *lactente*). Um *infant*, por definição do dicionário Webster's New World Medical Dictionary, não necessariamente é alimentado por leite materno, como se pode ver:

A child up to 2 years (24 months) of age.

The word "infant" came from the Latin *infans* which was derived from *in-*, not + *Fari*, to speak = not to speak, speechless. The idea was that, since the ability to speak was thought to arrive at the age of two, younger children were infants.

[Uma criança de até 2 anos (24 meses) de idade.

A palavra "infant" tem origem no Latin *infans* que derivou de *in-*, não + *Fari*, falar = que não fala. A ideia foi de que como pensava-se que a habilidade de falar seria aos dois anos, crianças menores seriam infants.]

Como procuramos mostrar, o uso dos sintagmas *lactentes de termo/full term infants* é condicionado pela própria combinação de palavras, e não por uma equivalência no sentido. Se assim fosse, a tradução para o inglês deveria conter o traço "em amamentação", o que não ocorre. Além disso, em inglês, as outras possibilidades de ocorrência desse sintagma incluem *full term [children, baby]*, mas nenhuma explicitamente envolvendo esse traço mencionado.

Existe também, no trecho que estamos usando como exemplo, uma adaptação para a tradução de *3 a 6 meses de idade*, usando, em inglês, a construção *aged 3 to 6 months*. Aqui

percebemos que o sentido permanece, mas a maneira de reexpressão tem uma mudança do sintagma nominal *de X meses de idade* para um verbo *aged*.

Essas observações nos levam também a crer que a tendência a um texto traduzido ter mais ou menos traduções por equivalência (interpretativas) ou correspondência (linguísticas) depende também da área em questão. Na linguagem da Medicina nos artigos científicos os modos de dizer são bastante peculiares em cada idioma, talvez mais do que em artigos também científicos mas de outras áreas, como a Ciência da Computação, por exemplo. Assim, de acordo com a nossa perspectiva, essas outras áreas, aparentemente, têm menos restrições de combinações no nível sintagmático.

Nesta tese, estamos considerando níveis ainda mais amplos de associação. Como exemplo, no mesmo parágrafo citado anteriormente:

Exemplo:

Verificar o comportamento dos valores da hemoglobina e a prevalência de anemia entre lactentes de termo de 3 a 6 meses de idade em aleitamento materno exclusivo.

Além das associações *lactentes de termo, meses de idade, aleitamento materno exclusivo*, que poderíamos incluir em um conceito de colocação, ou de *n-gramas* e sintagmas, pensamos que há associações entre esses agrupamentos, no nível maior, da frase. Assim, entendemos que há uma associação de nível fundamentalmente semântico e pragmático entre *meses de idade* e *aleitamento materno exclusivo*. Dizendo de outra forma, há uma tendência de essas palavras andarem juntas e ocorrerem nas mesmas sentenças e nas mesmas seções dos artigos porque pertencem ao mesmo campo semântico e pragmático. Essa associação seria fator determinante para a construção do caráter especializado do texto.

Segundo Sinclair (2004, p. 140), o texto é essencialmente percebido como uma série de escolhas relativamente independentes de um item após o outro, e os padrões de combinação ficam seriamente subvalorizados: o significado parece ser criado pelas escolhas paradigmáticas. No texto *Trust the text* (2004), Sinclair apresenta a questão do significado em termos de uma hipótese, qual seja:

(...) o significado de um texto pode ser descrito por um modelo que reconcilia as dimensões paradigmáticas e sintagmáticas de escolha com cada ponto de escolha (p. 141)

Nesse modelo, cinco categorias de co-seleção são colocadas como componentes de um item lexical³⁹: duas obrigatórias e três opcionais. As obrigatórias seriam a palavra núcleo, invariável, e a *prosódia semântica*, que determinaria o significado do todo. As categorias opcionais realizariam escolhas coordenadas secundárias, fariam um ajuste fino do significado, dando coesão semântica ao texto como um todo. Elas relacionariam as palavras nas duas dimensões, sintagmática e paradigmática, e seriam: *colocação*, *coligação* e *preferência semântica*. Sinclair esclarece que os dois primeiros são termos postulados por Firth (1957, 1968).

2. As unidades de tradução

Considerando a leitura da seção acima, passamos agora a explicar o conceito de unidades de tradução, para que se possa entender por que as associações sintagmáticas, em especial essas que analisaremos nesta tese, podem ser tomadas como unidades de tradução especializadas. A relevância das mesmas para a tradução, esperamos, já ficou evidenciada, mas temos ainda outras considerações a fazer a respeito.

No momento que optamos por uma abordagem que deseja incluir o viés da tradução ou do tradutor, queremos salientar que as características de nossas unidades de análise são primeiramente destacadas pensando na sua relevância na abordagem de um texto a ser traduzido e, depois, avaliando o fenômeno linguístico *per se*. Tentando dizer de outra forma, essas unidades de análise, como veremos, são recortadas do texto porque podem apresentar alguma dificuldade ao tradutor ou porque podem ser uma vantagem que ele as conheça. Como vantagem, consideramos que, através delas, ou melhor, sabendo que determinadas palavras "andam" juntas, o tradutor conhece melhor o texto com que trabalha, economiza tempo e ganha confiança, ao saber como usá-las na língua de chegada. Em segundo, mas não menos importante, enfocamos seu caráter de especialidade, por elas fazerem parte do texto especializado e por contribuírem para a transmissão de um conhecimento da área médica.

Como se sabe, a tradução de um texto é feita por partes que não necessariamente – e pouco provavelmente, na verdade – são palavras isoladas. A ideia da tradução “palavra por

³⁹ Posteriormente, Sinclair fez uma alteração da terminologia em seu modelo. Em vez the item lexical, usou os termos ELU (Extended Lexical Unit) e, depois, MSU (Meaning Shift Units).

palavra” foi, com o passar do tempo, sendo substituída pela tradução mais preocupada em preservar o significado do que em preservar as palavras do texto fonte (Quah, 2006, p. 23). Na busca por equivalentes que possam melhor atender ao critério de adequação ao propósito e ao público-alvo, o tradutor trabalha, hoje, com unidades maiores do que a palavra. Conforme apontou Teixeira (2008), essas unidades "estão vinculadas a situações de comunicação específicas: uma para o texto de partida, outra para o texto alvo”(p.28).

Alves *et al.* (2000) definem unidade de tradução como “um segmento de texto de partida, independente de tamanho e forma específicos, para o qual, em um dado momento, se dirige o foco de atenção do tradutor” (p.38). Tognini-Bonelli (2001) faz uma distinção entre *unidade de significação* e *unidade de tradução*, a saber:

Enquanto unidades de significação são definidas contextualmente no nível linguístico – isto é, examinando o co-texto verbal da palavra ou grupo de palavras escolhidos e identificando os padrões de co-seleção – as unidades de tradução são definidas estrategicamente como resultado de decisões explícitas tomadas pelo tradutor para atingir o efeito ou o propósito equivalente do original.(p. 133)

No nosso entendimento, no entanto, mesmo na definição acima, o fato de uma unidade ser unidade de significação não impede que ela seja também uma unidade de tradução. Conforme esclarecem ainda Alves *et al.* (2000), na definição abaixo, a unidade de tradução é o ponto de partida para o trabalho processual do tradutor:

(...)Trata-se de um segmento em constante transformação que se modifica segundo as necessidades cognitivas e processuais do tradutor. A unidade de tradução pode ser considerada como a base cognitiva e o ponto de partida para todo o trabalho processual do tradutor. Suas características individuais de delimitação e sua extrema mutabilidade contribuem fundamentalmente para que os textos de chegada tenham formas individualizadas e diferenciadas. O foco de atenção e consciência é o fator direcionador e delimitador da unidade de tradução e é através dele que ela se torna momentaneamente perceptível. (p. 38).

Esse conceito, no entanto, não é unânime, pois depende de cada abordagem ao produto e ao processo tradutório. A unidade de tradução, dentro dos estudos teóricos, pode ser desde um “morfema, palavra, sintagma, frase, oração, parágrafo, texto, discurso e até mesmo cultura” (Teixeira, 2008. p. 56). Apesar de ser uma definição ampla, que não faz restrições quanto a estruturas sintáticas ou composicionais, acreditamos ser a única definição possível para o que é uma unidade de tradução. Sendo assim, ampla, podemos então começar a pensar em apontar e descrever diferentes tipos de unidades que podem vir a ser unidades de tradução.

Teixeira (2008), usou o termo Unidades de Tradução Especializadas como unidades que tipificam o uso da língua nos textos técnicos. A autora demarca que essas unidades vão além dos termos técnicos, mas, mais importante, prova que elas vão além do que a maioria dos dicionários técnicos costuma incluir em suas entradas, tais como advérbios e preposições.

No que diz respeito às associações sintagmáticas no texto médico, que serão estudadas no contexto de textos da área de Pediatria, pensamos primeiramente no fator da relevância do encadeamento sintagmático, como descrito na seção anterior, para defini-las como UTEs. Dada a dificuldade de se comprovar, metodologicamente, por que uma associação sintagmática pode ser uma UTE, pensamos que poderíamos demonstrar isso observando como elas foram traduzidas, o que faremos também com mais exemplos no Capítulo IV. A seguir, frases extraídas de nosso *cópus* de artigos em Pediatria e suas traduções:

<p><u>Foram avaliadas 26 crianças</u> com prescrição de dieta isenta de leite de vaca e derivados, <u>11 do sexo feminino e 15 do sexo masculino</u>, com <u>idade média de 19,1 meses</u>, durante a primeira consulta no Ambulatório da Disciplina de Gastroenterologia Pediátrica da Universidade Federal de São Paulo - Escola Paulista de Medicina, no período entre setembro de 2001 e abril de 2002.</p>	<p><u>A total of 26 children were evaluated</u>, <u>11 female</u>, <u>15 male</u> and with a <u>mean age of 19.1 months</u>. All of them had been prescribed a diet free from cow's milk and cow's milk by-products at their first consultation at the Pediatric Gastroenterology Clinic of the <i>Universidade Federal de São Paulo - Escola Paulista de Medicina</i>, during the period between September 2001 and April 2002.</p>
---	--

<p><u>As crianças foram avaliadas</u> com a <u>idade média</u> de 55,4 meses ($\pm 6,5$), <u>variando de 48 a 71 meses</u>.</p>	<p><u>Children were assessed</u> at the <u>average age</u> of 55.4 months (± 6.5), <u>range of 48 to 71 months</u>.</p>
--	--

Observando essas duas sentenças, em português, percebemos que elas se repetiram em dois trechos de nosso *cópus*. Ou seja, crianças, avaliadas, idade média, por exemplo, ocorreram juntas em trechos de diferentes textos. Na tradução, podemos ver que foram mantidas as associações, mas, da mesma forma que em português, há diferentes construções que se valem dessas palavras (só que em inglês). Inclusive, surpreendentemente, também são variáveis as traduções para *avaliadas* e *idade média*, para as quais foram selecionados diferentes correspondentes em inglês: *assessed/evaluated* e *mean/average*. Para nós, essa diferença ressalta como pode ser produtivo tratar dessas associações e apresentá-las ao tradutor, já que as mesmas expressões em português podem ser traduzidas de diferentes formas em inglês. Outro ponto que podemos destacar nos exemplos acima é que a opção de tradução para *variando de x a x meses* não manteve o verbo, usando no lugar um substantivo, em uma outra forma de expressão em inglês *range of x to x months*. Pensamos que, se o

tradutor tiver que traduzi-la, é interessante que conheça, por exemplo, essa possibilidade de alcançar uma “equivalência”⁴⁰.

Outro fator importante a considerar é que, nas associações que avaliaremos e que consideraremos como UTEs, não há uma mera coocorrência ocasional de itens. Esses itens coocorrem em uma mesma frase mas com uma finalidade ou significado ideacional, no sentido da GSF. O significado ideacional, segundo Halliday e a GSF, se relaciona à representação de mundo, isto é, ao acontecimento, aos participantes e às circunstâncias. Sendo assim, as realizações textuais podem variar, mas o significado ideacional permanece. Por exemplo, na associação entre os itens *idade/anos/crianças*, podem ocorrer estruturas como *foram avaliadas crianças com idade inferior a 5 anos e crianças de 3 a 9 anos de idade participaram do estudo*. São estruturas diferentes, mas que cumprem um mesmo papel de descrever a idade dos participantes do estudo.

Ao avaliarmos, brevemente, a maneira como se apresentam as associações no texto fonte e no texto alvo das traduções especializadas, acima, observamos que o modo de introduzir as palavras na frase é relevante para o tradutor, uma vez que a associação que existe em português, no nível de palavra (significante), é desconstituída para ser refeita em inglês, em que novas associações, com os significantes de outra língua, devem ser identificadas. Vemos que palavras que podem ser consideradas chave são mantidas e apenas o modo como são apresentadas se altera. Essas palavras centrais seriam os elos de ligação entre texto-fonte e texto-alvo; um elo no nível do significado. Poderíamos pensar que existe aqui uma estratégia de tradução, em que, no momento da apreensão do sentido, o tradutor mantém alguns itens que são mais fixos, no nível do conteúdo, e joga com o modo de inserir tais itens na frase na língua alvo.

No que tange à denominação “especializadas” em estudos em Terminologia, encontramos a denominação Unidades de Significação Especializada (USE), introduzida por Estopà (2000), que considera as USE como unidades que veiculam conhecimento especializado e que abarcam diversos tipos de unidades sígnicas, podendo ser linguísticas ou não-linguísticas. As linguísticas podem ser léxicas (termo/substantivo, adjetivos, verbos...) ou não léxicas (sintáticas). Dentro das não-léxicas estão as Unidades Fraseológicas

⁴⁰ Temos conhecimento das críticas, principalmente da escola desconstrutivista, ao conceito de equivalência. Por isso, utilizamos a palavra *equivalência* entre aspas. Referimo-nos a equivalentes, aqui, pensando nas funções que as palavras, as expressões ou os textos desempenham, pois pensamos na equivalência de funções entre um texto fonte e um texto alvo.

Especializadas (UFE) e as combinações nominais recorrentes. Visto que as USE dependem da atividade, um texto especializado pode ter unidades do ponto de vista temático ou da finalidade profissional. Exemplos de USE conforme Estopà (2000): *anemia moderada, estudo histopatológico*. Nesse exemplo, *estudo* e *moderada* não são considerados USE, mas a composição em que se inserem são USE de caráter terminológico, segundo a autora.

Estopà afirma também que “Combinações especializadas recorrentes são parte das USE” e que essas são “unidades que adquirem valor especializado em uma área de conhecimento, representando e transmitindo esse conhecimento” (Estopà, 2000, p. 26). Os exemplos de combinações recorrentes citados, no entanto, não se assemelham ao mesmo conceito de combinações que investigamos nesta tese. São combinações recorrentes para Estopà: *radiografia do tórax, diagnóstico de toxoplasmose, biópsia de pele*. O que percebemos é que é um conceito de combinação bem restrito a sintagmas nominais e atrelados à significação especializada no sentido de corresponderem a um referencial, a um conceito. As unidades que enfocamos são mais amplas e o conceito de combinação realmente abrange possibilidades de combinação, ou de variação, entre os itens que as compõem, que não necessariamente são portadores de um significado específico, terminológico, embora, a nosso ver, são fundamentais para a viabilização do conteúdo especializado.

Teixeira (2008), em sua tese de doutorado, esclarece que as Unidades de Tradução Especializadas podem compreender categorias até o momento desprezadas por dicionários e podem não conter termo algum e ainda serem da maior importância para garantir a naturalidade na tradução. Em nosso trabalho, investigaremos associações sintagmáticas que são UTEs e que atendem a esses quesitos. Porém, enfocamos na hipótese de que há, nos textos especializados, unidades de tradução com um determinado perfil, dentre tantos que são possíveis, e que, além de não conterem termo algum, são associações com combinações mais ou menos fixas e frequentes de palavras, não necessariamente contínuas, associadas no nível sintagmático por uma atração semântica e/ou pragmática e/ou gramatical. Para investigar tal hipótese, desenvolvemos alguns estudos iniciais, relatados no Capítulo IV a seguir, que permitiram chegar a um método de investigação e de análise, o qual será relatado no Capítulo VI.

CAPÍTULO IV

ABORDAGENS INICIAIS À QUESTÃO DE PESQUISA

Este capítulo demonstra as principais abordagens e metodologias empregadas em nosso percurso para investigar nossa hipótese e a primeira questão de pesquisa:

É possível identificar e descrever, com base em corpus, associações sintagmáticas relevantes para a tradução de artigos médicos em Pediatria com um determinado perfil?

Essa questão não encerra um fim em si mesma. A finalidade de encontrar um meio de identificar tais associações está relacionada aos nossos objetivos: i) demonstrar que essas associações podem ser consideradas Unidades de Tradução Especializadas e delinear um perfil para as mesmas; e ii) elaborar um banco de dados com as associações mais frequentes, para que ele possa ser inserido em uma ferramenta de apoio ao tradutor.

Desde o início da pesquisa sabíamos que a identificação de *n-gramas* ou de agrupamentos recorrentes de palavras é um procedimento mais simples e que, hoje, computacionalmente, não apresenta maiores dificuldades. No entanto, buscar palavras associadas que poderiam estar em diferentes posições dentro de uma mesma frase (variação posicional) sem saber quantas outras estariam entre elas (variação constitutiva) apresentou-se como um desafio de pesquisa. O desafio, assim, torna-se identificá-las em sua descontinuidade

Muito do que se realiza em termos de identificação/extração automática de fraseologias, colocações e terminologias é bastante baseado em buscas que levam em conta a formação morfológica e sintática desses eventos. Mas, conforme já constatou Bevilacqua

(2003), muitas vezes esse tipo de extração resulta em um alto nível de ruído, sendo necessário levar em conta também outros critérios, como “fatores pragmático-discursivos”.

Nosso intuito, no desenvolvimento da tese, portanto, foi encontrar um outro caminho que não o de padrões de formação sintática de expressões para identificá-las. Melhor dizendo, não bastaria apenas etiquetar o *cópus* e acionar buscas que fossem, por exemplo, uma seqüência do tipo substantivo + adjetivo + preposição em determinada função sintática. Durante nosso percurso, nos estudos-piloto realizados, encontramos dados que consideramos interessante mencionar no contexto desta tese, além dos métodos que utilizamos, que podem ajudar a entender a complexidade da busca e os caminhos encontrados. O objetivo da busca por um método apropriado para identificar as associações sintagmáticas frequentes é o de poder comprovar nossa hipótese com o respaldo de um *cópus*, usando recursos e princípios da Linguística de *Cópus*.

Para conseguir responder a questão de pesquisa acima citada, exploramos diferentes abordagens em estudos-piloto e experimentos. Em alguns desses estudos-piloto, trabalhamos com amostras menores do *cópus* para melhor manejo e melhor entendimento de como tratar a questão de pesquisa. Embora esses estudos tenham sido considerados *piloto*, eles também incluíram observações extensivas, por isso apresentamos aqui neste capítulo relatos resumidos do que foi realizado. Procuramos discutir os resultados parciais obtidos ao término da descrição de cada um deles e revelar no que puderam contribuir para o prosseguimento da pesquisa.

Relatamos, a seguir, esses estudos. O primeiro seguiu uma abordagem de matrizes fraseológicas, conforme definição dada por Gouadec (1994). Depois, fizemos experimentos levando em consideração a ocorrência de conjunto de elementos em determinadas seções dos artigos. O passo seguinte envolveu o estudo dos dados com base nas seções dos artigos e incluiu utilização de algumas ferramentas de cálculo estatístico.

1. Primeiro estudo-piloto: abordagem via matrizes fraseológicas

Trazemos aqui a explanação das fases da metodologia empregada, a descrição dos resultados e discussão. Essa abordagem via matrizes fraseológicas, como já reiteramos, procurou seguir o modelo apresentado por Gouadec (1994) para avaliar seu rendimento em meio às nossas necessidades. Conforme exposto no capítulo anterior, Gouadec segue a noção de “matrizes” para as entidades fraseológicas.

Ressaltamos novamente que, embora esse autor use os termos fraseológico/fraseologismo, sua noção de fraseologia não é respaldada por um critério de cristalização dessas expressões; é, antes disso, uma percepção mais ampla do fenômeno, denominado genericamente de *fraseologismo*. A proposta do autor é que “em um fraseologismo, a matriz é o elemento invariante no interior do qual permutam variáveis” (Gouadec, 1994). O exemplo que o autor apresenta é:

As cláusulas e condições de utilização do produto abaixo constituem um compromisso concluído entre V.Sa na qualidade de usuário final e o construtor

Nesse segmento de texto, *constituem um compromisso concluído* seria a matriz e as outras partes as variáveis.

No primeiro estudo piloto, nosso objetivo foi reconhecer matrizes fraseológicas em uma pequena escala, em artigos científicos de Pediatria. A partir dessas matrizes, fizemos uma observação das palavras-chave que as compunham. Nossa suposição era de que existiriam, nas matrizes, determinadas palavras, que chamamos de palavras-chave, que aglutinariam em torno de si estruturas e informações relevantes para a tradução. Dizendo de outra forma, palavras que se associariam a outras no nível sintagmático, resultando dessa associação uma espécie de “modo de dizer” da área e do tipo de texto em questão.

Para alcançar os objetivos acima delineados, utilizamos um resumo de um artigo do qual extraímos, manualmente, possíveis matrizes, e um cópulo composto de 231 artigos em Pediatria, nos quais projetamos as matrizes encontradas no resumo, utilizando o software WordSmith Tools. Essa projeção pretendeu validar as matrizes no cópulo, isto é, confirmar se a matriz selecionada no resumo aparecia em outros textos do cópulo. Escolhemos partir do resumo porque, em tese, ele aglutina os pontos temáticos e lexicais mais importantes do todo do texto.

1.1. Metodologia

O resumo do qual selecionamos as matrizes é parte integrante do artigo *Prevalência de rinite alérgica em adolescentes do Distrito Federal: comparação entre as fases I e III do ISAAC*, publicado no periódico *Jornal de Pediatria*, volume 82(2), no ano de 2006. O artigo foi publicado em português e inglês, nos meios eletrônicos e papel, e foi uma escolha

aleatória, já que o estudo visava apenas a uma aproximação inicial ao nosso objeto de estudo e a uma metodologia.

Já o *cópus* utilizado no segundo passo constitui-se de 231 artigos do mesmo *Jornal de Pediatria*. Este *cópus* foi compilado para a realização de um estudo de mestrado elaborado por Robert Coulthard (Coulthard, 2005) e é composto por 496.145 palavras.

Para atingir nossos objetivos de identificação e observação de matrizes, a investigação foi elaborada em duas fases, ou passos, descritos a seguir.

1º passo: identificação de sequências notáveis

Como meio de abordar as matrizes fraseológicas em artigos de *Pediatria*, em escala reduzida, contando apenas com o ponto de partida de resumo de 1 artigo, fizemos uma extração manual de sequências de caracteres que poderiam ser consideradas matrizes no resumo do artigo.

Para fazer a seleção, procuramos por sequências que se encaixariam no critério de “notabilidade”, mencionado por Gouadec. Conforme exposto, os critérios apresentados por Gouadec como fundamentais para que uma cadeia seja considerada “notável” são: caráter especializado, repetição, risco inerente a sua manipulação e vantagem que se pode ter ao dominá-las. Além disso, é preciso ressaltar que nos valem da nossa própria experiência como tradutores desses artigos para captar a notabilidade ou relevância dessas sequências.

Após o reconhecimento de oito sequências de caracteres no resumo e da elaboração de uma possível matriz, sublinhamos o que consideramos as palavras-chave da sequência. Esse procedimento tem fundamento em uma noção também explicitada por Gouadec, de que “é preciso (...) que as chaves sejam aquelas que espontaneamente os usuários mobilizam”(p. 171). Selecionamos palavras que têm potencial de agregarem diferentes tipos de construções ao seu redor, sem diferenciar categoria gramatical, nem diferenciar termos de palavras da linguagem cotidiana.

Os resultados da extração manual podem ser conferidos na seção a seguir.

2º passo: confirmação

No intuito de demonstrar que as sequências selecionadas tinham um caráter matricial, isto é, que suas variáveis podiam ser preenchidas por outras palavras ou sequências,

utilizamos o corp us constitu do de 231 textos do Jornal de Pediatria, brevemente caracterizado anteriormente.

Com o software *WordSmith Tools* e sua ferramenta *Concord*, fizemos uma busca nos textos visando encontrar outras constru es com a mesma palavra-chave das sequ ncias selecionadas, ou, dizendo de outra forma, fizemos uma busca das concord ncias das palavras-chave sublinhadas. Na se o Resultados, a seguir, apresentamos apenas os dados mais relevantes com rela o   matriz que definimos, ou seja, as constru es com a mesma sequ ncia de palavras fixas selecionadas no 1  passo e que preenchem as vari veis selecionadas.

1.2. Resultados

Vejamos os resultados seguindo a sequ ncia de passos da metodologia. Como resultados, apresentamos o produto da identifica o manual (1  passo) e o produto do processamento autom tico (2  passo). No decorrer desta se o, tamb m h  alguns coment rios com rela o a regularidades encontradas e sobre a sua pertin ncia para o desenvolvimento da tese.

1.2.1. 1  passo: identifica o de sequ ncias not veis

A seguir, temos as sequ ncias selecionadas manualmente e, logo abaixo, em negrito, uma poss vel matriz para cada uma. A por o sublinhada corresponde   palavra ou sequ ncia de caracteres usada como palavra-chave, ou express o de busca. A rela o apresentada configura-se nos dados brutos da sele o. Ap s uma observa o mais detalhada das sequ ncias e das palavras-chave (2  passo), definimos o potencial dessas sequ ncias se constitu rem matrizes trazendo, nesta tese, apenas exemplos da primeira matriz, embora no estudo piloto tenhamos estudado todas as sete abaixo. Usamos nota o semelhante a que Gouadec (1994) utiliza para identificar matrizes, palavras-nodo e vari veis.

Os itens entre colchetes representam a parte vari vel da matriz. A parte em negrito   a parte fixa. As letras X, Y, Z representam as vari veis.

1. **Determinar a preval ncia de** [rinite al rgica] **em** [um grupo aleat rio de escolares de Bras lia (DF)], [com idade entre 13-14 anos...]

Determinar a prevalência de [X] em [Y] [Z].

2. [Dois] estudos de corte transversal [foram realizados]

[X] estudo* de corte transversal [Y]

3. ...**comparar** [as taxas de prevalência] **entre** [diferentes grupos socioeconômicos]

...comparar [X] **entre** [Y]

4. ...[39 escolas] **foram** [escolhidas] **aleatoriamente**

[X] foram [Y] aleatoriamente

5. **Foram obtidos** [3.009] **questionários, sendo** [53,5%] **do sexo** [feminino]

Foram obtidos [X] questionários, sendo [Y] **do sexo** [Z]

6. ... **houve aumento significativo da** [prevalência de rinite diagnosticada]

... **houve** aumento significativo de [X]

7. **A prevalência de** [rinite alérgica] **aumentou de maneira significativa**

A prevalência de [X] **aumentou de maneira significativa**

Esse é, então, o conjunto das matrizes que selecionamos apenas no resumo. A seguir, apresentamos os resultados da busca por outras ocorrências de caracteres, no córpus, que atendem a matriz identificada. Como já mencionamos, este próximo passo visa verificar se a matriz é produtiva, ou seja, se ela se repete no córpus e, portanto, atende ao critério de repetição. Queremos, porém, deixar claro que não é só o critério de repetição que torna uma matriz uma sequência notável, como já explicamos.

1.2.2. 2º passo: confirmação

Neste passo, utilizamos as palavras-chave das matrizes (i.e. a porção sublinhada) para fazer buscas de ocorrências no córpus. Apresentamos apenas as ocorrências que seguem a matriz selecionada, ou seja, que preenchem as variáveis X, Y e Z acima identificadas. Em alguns casos, traremos outros exemplos e observações, bem como outras tentativas de busca

por outras matrizes, motivados pelas evidências apresentadas. Consideramos também que nem todas as variáveis X, Y e Z precisariam ser preenchidas. Os resultados são apresentados para cada sequência selecionada no 1º passo, seguindo a mesma ordem, de 1 a 8.

1. Determinar a prevalência de [X] em [Y], [Z]

X – evento (sintoma, doença)

Y – população/sujeitos

Z – complemento/especificação de Y, especialmente um dado de idade

Sete foram os resultados de concordâncias com a expressão *determinar a prevalência*. Houve ainda três ocorrências com a flexão do verbo *determinar* na terceira pessoa do plural: *determinaram*.

A matriz se mostrou produtiva, uma vez que as variáveis foram preenchidas por elementos condizentes com o identificado no texto fonte (X, Y, Z; sintoma/doença, população/sujeitos e complemento) em todas as sete ocorrências.

Chama a atenção, nesta análise, o fato dessa matriz ocorrer com a palavra *objetivo* em cinco das sete ocorrências. Com esse dado, entendemos que poderíamos ter também como matriz as sequências *o objetivo do estudo é determinar a prevalência* e *o estudo teve como objetivo determinar a prevalência*. Além disso, já projetando esse resultado para nosso objetivo de comprovar a existência de associações sintagmáticas entre palavras específicas, poderíamos dizer que *objetivo + determinar + prevalência* seria uma associação típica e relevante, independentemente da formulação da frase.

É essa “flexibilidade” de abordagem que queremos alcançar em nossa investigação. Por isso, estamos usando a noção de associações e não de fraseologismos. Além disso, há a diferença de constituição das variáveis, que, mesmo preenchendo o mesmo campo funcional (população, evento, sujeitos etc), constituem-se de elementos textuais muito distintos. Por exemplo, preenchendo a categoria *sujeitos*, temos:

- gestantes;
- crianças que nunca tiveram diarreia;

- crianças da rede de ensino público e privado da zona urbana de Feira de Santana-BA;

Não menos importante é a revelação de que a variável X, acima demonstrada, configura-se uma sequência de caracteres com um caráter mais aproximado ao de um termo canônico⁴¹, sendo, evidentemente, preenchida por conceitos que assumem um caráter terminológico no contexto em que se encontram.

Nesta etapa de validação desta matriz, concluímos que:

- a matriz selecionada é válida, pelo critério de repetição no *cópus*;
- a matriz pode ser expandida;
- a coocorrência *determinar a prevalência* poderia constar como aglutinadora de informações relevantes, uma vez que o que acontece em torno dela pode ser considerado notável pelo critério de repetição e de frequência; e
- é uma expressão, que poderia vir a ser parte de uma regra, em um sistema automático, que definiria que a variável X constituir-se-ia de um termo/expressão especializada e, se cruzada com as palavras do título, ajudaria a identificar o assunto tratado no texto.

Neste estudo-piloto, avaliamos todas as 7 matrizes. Porém, visto que não pretendemos desviar o foco de nossas questões principais, relatamos apenas este resultado, a título de ilustração.

Nesse intuito de continuar observando matrizes, fizemos uma observação de quais seriam as coocorrências da palavra *prevalência*, pois essa se mostrou muito frequente. Após utilizar a ferramenta *Concord* do *WordSmith Tools* para verificar a frequência de ocorrência da palavra, utilizamos um outro recurso do *Concord*, a opção *Show Collocates*. Constatamos que palavras que indicam índices elevados: maior, aumento, alta e elevada, estão entre os 20 primeiros coocorrentes. Assim, confirmamos que esse tipo de associação realmente existe no nosso *cópus*; restava encontrar uma maneira mais automática de encontrá-la.

Observando os resultados que obtivemos, pudemos ver que, quando fizemos as buscas pelas palavras-chave, o leque de possibilidades de ocorrências para essas palavras se abre. Do

⁴¹ Nos referimos aqui ao entendimento geral de que termo é aquela palavra ou locução que denota um conceito de uma determinada área técnica ou científica.

ponto de vista do potencial de estruturas que se formam em torno dessas palavras-chave, como *prevalência*, por exemplo, pensamos na riqueza de coocorrências formadas e, portanto, na validade de selecioná-la como palavra-chave. Por outro lado, nos questionamos quanto à melhor forma de abordar essa gama de resultados (afinal, há 533 ocorrências de *prevalência*).

Após todas essas observações feitas e depois do experimento com matrizes, percebemos, mais uma vez, como é difícil enquadrar esse tipo de evento que resolvemos estudar. É complexo colocá-lo em classes, como também é penoso tirar conclusões mais sistemáticas sobre sua configuração no texto. O que se pode verificar é que a reflexão teórica elaborada por Gouadec (1994) harmoniza-se muito bem com os resultados que obtivemos. Entendemos que as palavras-chave que selecionamos surtiram resultados positivos, pois levaram a outras expressões e confirmaram-se como parte de sequências que ocorrem em um universo maior de textos.

Nesse estudo-piloto, lembramos, o foco de observação esteve voltado para diversos fatores, pois procuramos extrair o máximo de informações. No entanto, nosso propósito na tese é outro, e precisávamos obter mais rendimento em termos de detectar essas associações. Queríamos encontrar uma maneira de tornar o método de identificação das cadeias sintagmáticas, em seus núcleos e em seus elementos satélites, menos manual e mais automático. Era preciso obter um procedimento em que pudéssemos não apenas comprovar a existência de uma associação frequente entre determinadas palavras, mas também demonstrar que/quais palavras são essas que se associam. Em busca desse objetivo, realizamos ainda outro estudo piloto, trabalhando com o *córpus* separado em seções, conforme as divisões que compõem os *papers*. Isso é que relatamos a seguir.

2. Segundo estudo-piloto: abordagem via seção dos artigos

Conforme já comprovado (Teuffel, 1999), os artigos médicos são, em sua grande maioria, bem estruturados em seções, característica que os distingue do usual encontrado em artigos de outras áreas. No estudo de Teuffel (1999), fez-se uma comparação entre artigos de Linguística Computacional e de Medicina quanto à presença de subtítulos prototípicos de artigos. O resultado aponta que 74% dos subtítulos dos artigos de Linguística Computacional não são prototípicos ou fixos, em contraste com o *córpus* de Medicina, em que os únicos cabeçalhos que não correspondiam a elementos prototípicos estavam dentro de outros

prototípicos (especificações de métodos, por exemplo). Para ilustrar um pouco melhor, apresentamos mais alguns números levantados pela referida pesquisa.

Quadro 1 – Frequência comparada de ocorrência de seções em artigos de Linguística Computacional e Cardiologia

Linguística Computacional (80 artigos)			Cardiologia (103 artigos)		
Título	Frequência		Título	Frequência	
Introdução	63	79%	Introdução	103	100%
Conclusão	34	43%	Resultados	97	94%
Discussão	13	16%	Discussão	97	94%
Conclusões	13	16%	Métodos	95	92%
Agradecimentos	12	15%	Tabelas	81	79%
Resultados	8	10%	Estatística	41	40%
Resultados experimentais	8	10%	Pacientes	30	29%
Avaliação	7	9%	Limitações	29	28%
Histórico	7	9%	Conclusões	26	25%
Implementação	6	8%	Análise estatística	23	22%
Exemplo	6	8%	Conclusão	18	17%
Agradecimentos	6	8%	Características dos pacientes	9	9%

Como se pode perceber, as seções do artigo em Medicina seguem, em geral, o formato IMRAD, composto de **I**ntrodução (a razão pela qual os autores decidiram fazer o estudo), **M**étodos (como fizeram o estudo e como analisaram os resultados), **R**esultados (o que encontraram) e **D**iscussão (o que os resultados significam) (BMJ, 1997).

Outra regularidade em artigos médicos é a classificação em categorias de artigos, conforme seus propósitos. De acordo com publicação do BMJ (British Medical Journal), essas categorias têm uma hierarquia de presença de evidências. Isso quer dizer que, para o médico usar um artigo para tomar decisões sobre intervenções clínicas, ele encontrará mais evidências em artigos do tipo revisões sistemáticas do que em relatos de caso. A ordem de importância é a seguinte:

- (1) Revisões sistemáticas e meta-análises
- (2) Ensaio randomizado controlado com resultados definitivos
- (3) Ensaio randomizado controlado com resultados não-definitivos
- (4) Estudos de coorte
- (5) Estudos caso-controlados

(6) Estudos transversais

(7) Relatos de caso

Toda essa estruturação, reconhecidamente bem definida dos artigos médicos, permite que trabalhem no *cópus* de Pediatria com um nível de precisão que é mais difícil de encontrar em artigos de outras áreas. As definições de seções já consagradas, internalizadas pelos autores e impostas pelos órgãos que publicam os artigos (editoras, associações médicas), possibilitam que tenhamos a segurança de levantar características comuns aos artigos na área de Pediatria, contando que poderão ser replicadas para outras áreas da Medicina. No *Jornal de Pediatria*, a classificação dos artigos é um pouco diferente, pois os artigos são classificados como artigos originais e artigos de revisão. Há ainda as comunicações breves, as cartas ao editor e os artigos especiais, que não se enquadram nas categorias anteriores, mas que são considerados relevantes pelos editores da revista. Os artigos originais do *Jornal de Pediatria* incluem os estudos dos tipos 2, 3, 4, 5 e 6 da classificação acima.

O que buscamos no estudo-piloto que relatamos a seguir é demonstrar que as co-ocorrências de certas palavras são determinadas pela seção do artigo em que se encontram. Esse pressuposto é embasado na crença de que certas características coocorrem consistentemente e de que há um padrão textual e um propósito comunicativo embasando seu uso (Biber, 1988). Entendemos que não apenas o artigo como um todo tem um propósito comunicativo, mas também cada uma das seções que o compõem. Assim, cada seção funciona como um pequeno texto em meio ao todo do artigo. O estudo relatado a seguir permitiu simular, em menor escala, como poderiam ser analisados os textos de nosso *cópus* completo. Realizamos, para isso, as etapas de descrição do *cópus* e da metodologia empregada, apresentação e discussão de resultados.

2.1. Descrição do *cópus*

No estudo piloto antes relatado, em que observamos as matrizes fraseológicas, trabalhamos com o *cópus* “bruto” do *Jornal de Pediatria*, isto é, da maneira em que nos foi fornecido pelo pesquisador que o compilou (Coulthard, 2005). Este *cópus*, lembramos, é composto por 231 textos (496.145 palavras).

Observamos, porém, que dentre esses 231 textos, nem todos eram artigos originais, pois haviam sido incluídos os artigos de revisão e as cartas ao editor. Foram excluídos esses textos, ficando nosso córpus constituído apenas por artigos originais. Ficou-se, então, com 120 artigos. Em seguida, elaboramos quatro subcórpus, contendo cada um apenas os textos correspondentes à seção em que se encontravam. O córpus ficou assim constituído:

Quadro 2 – Constituição do córpus do Jornal de Pediatria por seção

	Introdução	Métodos	Resultados	Discussão
Palavras	51.971	83.367	83.076	130.448
Total de palavras				348.862

2.2. Metodologia

Com base na leitura e na observação da composição lexical das seções de resumos de 4 artigos do Jornal de Pediatria⁴² e na observação da lista de frequências dessa pequena amostra, selecionamos duas palavras que são bastante comuns na seção Métodos e que, frequentemente, se associam a outras, formando associações significativas. São as palavras *meses* e *transversal*. Fizemos, então, uma busca dos colocados de cada uma delas em nosso córpus da seção Métodos, utilizando a ferramenta *Concord* do *WordSmith*. Em seguida, utilizamos o recurso de *mostrar colocados* desse mesmo software para fazermos uma avaliação de quais seriam as palavras que mais frequentemente se associaram a cada uma das selecionadas (*meses* e *transversal*). A partir daí, procuramos sistematizar possíveis associações entre elas e seus colocados, usando para isso novamente a ferramenta *Concord*.

2.3. Resultados

As seções a seguir apresentam os resultados obtidos com os procedimentos acima descritos para cada uma das duas palavras selecionadas, além de uma breve discussão dos dados obtidos.

2.3.1. Resultados dos processamentos para a palavra *meses*

⁴² Os resumos do Jornal são chamados de “estruturados”, pois são também divididos em seções.

Vejam os 15 primeiros resultados dos colocados da palavra *meses*, no quadro originado pelo *WordSmith*, a partir do *cópus* composto pela seção *Métodos*. Este quadro apresenta a frequência de coocorrência dos colocados até a posição de cinco palavras à esquerda e cinco palavras à direita da palavra de busca (nesse caso *meses*). É importante mencionar que utilizamos uma lista de *stopwords* nesta etapa.

Quadro 3 – Colocados da palavra meses

N	WORD	TOTAL	LEFT	RIGHT	L5	L4	L3	L2	L1	*	R1	R2	R3	R4	R5
1	MESES	181	14	12	4	8	2	0	0	155	0	0	2	8	2
2	ANOS	29	18	11	5	0	8	5	0	0	0	0	10	0	1
3	IDADE	29	10	19	1	5	4	0	0	0	0	18	0	0	1
4	COM	23	15	8	7	6	0	2	0	0	0	4	2	2	0
5	AOS	22	21	1	1	1	0	19	0	0	0	0	0	0	1
6	QUATRO	22	20	2	0	0	2	0	18	0	0	0	2	0	0
7	SEIS	20	19	1	0	0	1	0	18	0	0	0	0	0	1
8	ENTRE	16	15	1	1	7	0	7	0	0	1	0	0	0	0
9	APÓS	15	5	10	2	1	0	2	0	0	8	1	1	0	0
10	VIDA	13	0	13	0	0	0	0	0	0	0	12	0	0	1
11	CRIANÇAS	12	11	1	6	1	4	0	0	0	0	0	0	1	0
12	MAIS	12	7	5	0	5	2	0	0	0	0	2	0	1	2
13	NOS	12	10	2	0	0	7	3	0	0	0	0	2	0	0
14	FOI	11	8	3	1	2	5	0	0	0	0	1	0	1	1
15	MÉDIA	11	6	5	1	0	4	1	0	0	0	0	1	1	3

Neste quadro, na coluna marcada com um asterisco, encontra-se a palavra de busca. A seguir, deve-se olhar, então, para a primeira palavra na coluna *Word* depois da palavra *meses*. Os números indicam a frequência com que a palavra de busca ocorreu com esse colocado e em qual posição. *Idade*, por exemplo, ocorreu mais vezes com *meses* na posição 2 à direita. Essa colocação reflete a expressão *meses de idade*, por exemplo.

Optamos por investigar a coocorrência da palavra *média* com *meses* no subcórpus de *Métodos*, já que a maioria das outras coocorrências, na tabela, eram mais facilmente identificáveis e previsíveis (*meses de idade*, *meses de vida*, *seis meses*, *quatro meses* etc.).

Usamos então a expressão de busca *média * * meses* – que corresponde à combinação entre *média*, duas palavras quaisquer e *meses* - na ferramenta *Concord* do *WordSmith*, que apresentou os seguintes resultados:

1 - O tempo entre as aplicações variou de 1 mês a 2 anos, com média de 3,5 meses entre as aplicações

2 - O tempo de seguimento após a última aplicação de OK-432 variou de 2 meses a 40 meses, com média de 15,8 meses.

3 - Foram avaliadas 26 crianças com prescrição de dieta isenta de leite de vaca e derivados, 11 do sexo feminino e 15 do sexo masculino, com idade média de 19,1 meses, durante a primeira consulta no ambulatório

4 - O grupo controle foi composto por 30 crianças, sendo 13 do sexo feminino e 17 do sexo masculino, com idade média de 16,8 meses e dieta normal,

Identificamos que houve uma regularidade de dois tipos de construção:

idade média de X meses

com média de X meses

Considerando que nosso objetivo não é o de identificar colocações imediatamente vizinhas ou contíguas, optamos por investigar que outras palavras se associam com essas duas expressões nos outros corpúsculos (as seções: Introdução, Resultados, Discussão).

Quadro 4 – Expressão de busca: idade média de x meses

Seção (Nro. ocorr.)	Contexto de ocorrência (colocados)
Introdução (0)	∅
Resultados (3)	<ul style="list-style-type: none"> - As crianças foram avaliadas com a <u>idade média de 55,4 meses</u> ($\pm 6,5$), variando de 48 a 71 meses. - As 40 crianças do grupo seguimento Rotavirus A (+) apresentavam <u>idade média de 19,2 meses</u>, sendo 19 pacientes do sexo masculino e 21 do sexo feminino. - O grupo controle, constituído também por 40 crianças, apresentavam <u>idade média de 19,7 meses</u>, sendo 20 crianças do sexo masculino e 20 do sexo feminino.
Discussão (0)	∅

Para a sequência *com media de x meses*, coletamos os seguintes contextos:

Quadro 5 – Expressão de busca: com media de x meses

Seção/Nro. ocorr.	Contexto de ocorrência (colocados)
Introdução (0)	∅
Resultados (4)	<ul style="list-style-type: none"> - A idade variou de 2 meses a 12 anos, <u>com média de 25 meses</u> (desvio padrão = 19,5 meses) e mediana de 22,2 meses. - O tempo de queixa variou de 3 dias a 48 meses, <u>com média de 5,8 meses</u> (desvio padrão = 6,6 meses) e mediana de 3 meses. - O período entre a primeira avaliação médica adequada e o tratamento cirúrgico foi maior que 6 meses em 9/13 pacientes, <u>com média de 13,8 meses</u> e mediana

	de 21. - A idade referida de início dos sintomas variou desde o nascimento até 20 anos, com média de 16 meses e mediana de 3 meses (Figura 1).
Discussão (0)	Ø

Nossas considerações acerca das observações dessas expressões são relatadas após a apresentação dos resultados dos quadros de colocados de nossa próxima palavra selecionada, a palavra *transversal*.

2.3.2. Resultados dos processamentos para a palavra *transversal*

A seguir, temos os resultados dos colocados da palavra *transversal*, no quadro originado pelo *WordSmith*, no cópuz da seção *Métodos*. Esta tabela apresenta a frequência de coocorrência dos colocados até a posição de cinco palavras à esquerda e cinco palavras à direita da palavra de busca (nesse caso, a palavra *transversal*).

Quadro 6 – Colocados da palavra transversal

N	WORD	TOTAL	LEFT	RIGHT	L5	L4	L3	L2	L1	*	R1	R2	R3	R4	R5
1	TRANSVERSAL	28	0	0	0	0	0	0	0	28	0	0	0	0	0
2	ESTUDO	24	23	1	2	4	3	0	14	0	0	0	0	0	1
3	FOI	12	10	2	0	5	5	0	0	0	0	0	0	1	1
4	REALIZADO	9	7	2	1	0	5	1	0	0	0	2	0	0	0
5	CORTE	8	8	0	0	0	0	0	8	0	0	0	0	0	0
6	TIPO	7	7	0	0	0	0	2	5	0	0	0	0	0	0
7	VOL	7	7	0	6	1	0	0	0	0	0	0	0	0	0
8	FORAM	5	0	5	0	0	0	0	0	0	0	0	2	3	0
9	OBSERVACIONAL	5	2	3	0	0	2	0	0	0	0	2	0	1	0

Optamos por investigar a primeira coocorrência mais frequente: entre as palavras *transversal* e *estudo*.

Usamos, então, a expressão de busca *estudo transversal*, na ferramenta *Concord* do *WordSmith*, na seção *Métodos*, que apresentou os seguintes resultados:

1 - Trata-se de um estudo transversal, no qual foram comparados dois grupos: um com dieta isenta de leite de vaca e seus derivados e outro com dieta normal, isto é, sem recomendação médica ou terapêutica de exclusão de algum alimento.

2 - Tratando-se de um estudo transversal com variáveis dicotômicas, foram utilizadas as regras descritas por Browner et al. para o cálculo do n necessário para um a bicaudal de 0,05 e um β de 0,10.

3 - Foi um estudo transversal em que se avaliaram as variáveis socioeconômicas (anos de escolaridade da mãe e renda familiar), atividade física e hábito de fumar, e foram considerados, ainda, peso, altura e pressão arterial.

4 - Estudo transversal, no qual foram avaliados 30 pacientes, dos quais, 17 crianças e 13 adolescentes, sendo 17 do sexo feminino, com diagnóstico de DC de acordo com as recomendações da Sociedade Europeia de Gastroenterologia Pediátrica, Hepatologia e Nutrição...

5 - Trata-se de um estudo transversal, realizado entre novembro de 1999 e julho de 2001, nas sete unidades de terapia intensiva existentes em Belém (PA) e nos 14 berçários das maternidades do mesmo município.

6- Foi realizado um estudo transversal nas UTIPs e UTINs do Hospital São Lucas da PUCRS e da Irmandade Santa Casa de Misericórdia de Porto Alegre (ISCOMPA).

7 - Foi realizado um estudo transversal durante um período de dois anos (junho de 1997 a junho de 1999), no Hospital de Clínicas de Porto Alegre.

8- Este é um estudo transversal tendo como população-alvo 56.520 alunos do ensino médio (2º grau) regularmente matriculados nas escolas públicas (municipal, estadual e federal) e particulares do município de Porto Alegre, de acordo com a listagem fornecida pela Secretaria de Edu

9 - Foi realizado estudo transversal, observacional, baseado na avaliação das prescrições dos pacientes admitidos na UTIP do Hospital de Clínicas de Porto Alegre, no período de seis semanas consecutivas, considerando um dia diferente em cada semana avaliada.

10 - Foi realizado um estudo transversal com 91 pacientes, de ambos os sexos, entre 5 e 17 anos de idade, portadores de enurese noturna monossintomática ou associada a alterações diurnas do padrão miccional, encaminhados ao Ambulatório de Urodinâmica Pediátrica do Instituto Fernand

11 - Realizou-se um estudo transversal, observacional, contemporâneo, com dados subindividuais (orelhas), e o fator em estudo foi constituído pelas efusões observadas na OME.

12 - Foi realizado um estudo transversal controlado, cujo fator em estudo foi colestase crônica, e o desfecho é o nível plasmático de vitamina D.

13 - Foi realizado um estudo transversal, de caráter observacional, envolvendo todos os recém-nascidos internados na unidade de tratamento intensivo neonatal do Hospital de

Clínicas de Porto Alegre, que nasceram com peso menor ou igual a 1.500g, no período de 1º de setembro de 2000 a 31 d

14 - Este é um estudo transversal no qual foram avaliadas crianças de duas instituições de ensino da cidade de Belo Horizonte (MG).

No quadro a seguir, é possível verificar as outras ocorrências de *estudo transversal*.

Quadro 7 – Expressão de busca: estudo transversal

Seção/Nro. ocorr.	Contexto de ocorrência (colocados)
Introdução (3)	- Este <u>estudo transversal</u> tem como objetivo explorar preliminarmente possíveis fatores associados com níveis elevados da pressão arterial em crianças de - Diante dessa situação peculiar da cidade de Feira de Santana (precárias condições de vida e elevadas taxas de aleitamento), foi realizado um <u>estudo transversal</u> com o objetivo de avaliar o grau de proteção do aleitamento materno contra a diarreia aguda nas crianças menores de um ano. - No Brasil, em um <u>estudo transversal</u> , realizado em Belo Horizonte, capital de Minas Gerais, verificou-se uma soroprevalência de 34,1% em crianças (6).
Resultados (0)	0
Discussão (1)	- Concluindo, embora nosso <u>estudo transversal</u> tenha limitações, chama a atenção sobre a importância do pediatra para a monitoração da pressão art

Quadro 8 – Expressão de busca: corte transversal

Seção/Nro. ocorr.	Contexto de ocorrência (colocados)
Introdução	0
Resultados	0
Discussão	0

2.4. Discussão e sistematização dos resultados

Na primeira análise, investigamos a coocorrência da palavra *média* com *meses*, detectada na tabela de colocados obtida com o *WordSmith*. Ao buscar contextos para essa associação, na seção *Métodos*, chegamos a duas expressões mais ou menos fixas: *idade média de x meses* e *com média de x meses*.

A associação *idade média de x meses*, no *cópus* da seção *Métodos*, teve apenas duas ocorrências, nas quais se associou com palavras como *crianças*, *sexo feminino*, *sexo masculino* e os verbos na voz passiva *avaliar* e *compor*. No entanto, diferente do esperado, a

associação teve quase a mesma frequência na seção *Resultados*, demonstrando que é uma associação que se faz presente tanto numa quanto noutra seção.

Na seção *Resultados*, confirma-se a associação com *crianças*, *sexo feminino*, *sexo masculino* e também com *grupo controle* e *grupo seguimento* e com a expressão *apresentar idade*. Percebemos, então, que algumas associações se mantêm (com *crianças*, *sexo feminino*, *sexo masculino*). O que diferiu foi forma verbal: *avaliar* e *compor*, na seção *Métodos*, e *apresentar idade*, nos *Resultados*.

Essa constatação corrobora a ideia que também perpassa nossa tese, a de que o caráter especializado de um texto se constrói também pelas associações entre as palavras. A especialidade não está na palavra, mas no texto, no conjunto. Isso se dá porque, como se percebe, itens da seção *Métodos* estão presentes na seção *Resultados*, mas o que faz uma ser diferente da outra é a relação com outros itens diferenciadores, como, por exemplo, os verbos *avaliar* e *compor*, que são mais descritivos e, portanto, entende-se porque constituem a seção *Métodos*. Assim, o que faz essa seção ser a seção *Métodos*, entre outras coisas, não é a presença das palavras *crianças*, *sexo*, *grupo*, mas, sim, a associação destas com os verbos *avaliar* e *compor*. Continuando, percebe-se que algumas palavras são comuns a mais de uma seção dos artigos, e outras são específicas de determinadas partes do texto, construindo as diferenciações entre as partes. É nesse sentido também que acreditamos que o caráter especializado vai se construindo no texto, pelas relações de oposição. Se tomarmos *crianças* apenas, não conseguimos determinar seu nível de especialidade em meio ao texto. Mas, já quando olhamos *crianças* em relação a *sexo feminino*, e a *apresentar idade*, o item *crianças* já assume um outro *status*.

Continuando com nossa análise dos resultados, vemos que o modo como as palavras estão distribuídas nas sentenças não é fixo, no sentido de seguir uma matriz fraseológica ou até mesmo de se comportar como uma fraseologia estereotipada. Apenas percebe-se que tais palavras “acontecem” juntas em uma mesma sentença, com uma certa regularidade. Esse é o caso, por exemplo, das associações entre *foram avaliadas*, *crianças*, *idade média de x meses*:

Outra observação que podemos fazer, ainda para as expressões que levam as associações entre *média* e *meses*, é que ambas (*idade média de x meses* e *com média de x meses*) ocorreram nas duas mesmas seções dos artigos: *Métodos* e *Resultados*. Apresentaram ocorrência nula nas seções *Introdução* e *Discussão*. Com essa evidência, percebemos que há

uma tendência de que descrições da população estudadas não sejam apresentadas nessas seções, sendo mais características do desenvolvimento da metodologia e da apresentação dos resultados.

Ainda na expressão que observamos entre a associação *meses* e *média*, em *com média de x meses*, notamos que, nas coocorrências da seção *Métodos*, as palavras que mais se associaram foram *tempo* e o verbo *variar*. Já na seção *Resultados*, a associação foi com as palavras *mediana*, *desvio padrão*, *idade*, *tempo/ período* e com o verbo *variar*. Também aqui queremos chamar atenção para as palavras bem características da seção *Métodos* que se associaram à expressão de busca, como *mediana* e *desvio padrão*.

Na observação dos colocados de *transversal*, também identificamos duas sequências de caracteres que se mostraram mais fixas: *estudo transversal* e *corte transversal*.

Para *estudo transversal*, a primeira observação é que as ocorrências são em sua maioria na seção *Métodos*, caracterizando esse item como substancialmente típico dessa seção. Observamos que se associa com as realizações verbais *foi realizado* e *trata-se de*, e com uma outra caracterização do caráter do estudo: *observacional*. Identificamos que há uma regularidade da construção *foi realizado um estudo transversal* e *este é* ou *trata-se de um estudo transversal*.

Para *corte transversal*, observamos a associação com as palavras *estudo* e *tipo*, e também com os dois verbos que apareceram com a expressão acima, *realizar* e *tratar*, além do verbo *ser*. Também, para *corte transversal*, observa-se uma associação com *observacional* e ainda outros tipos de estudos como: *retrospectivo*, *descritivo*, *de base populacional*, *epidemiológico*. Percebe-se, então, que a caracterização do estudo geralmente é composta das diferentes qualidades desse estudo: *retrospectivo*, *descritivo* etc.

Outro dado que consideramos muito relevante é que, para a expressão *estudo transversal*, foi mais frequente a forma verbal *foi realizado*, enquanto para a expressão *corte transversal* foi mais frequente a forma verbal *realizou-se*. Para nós, tal fator demonstra que existe uma “atração” entre as palavras, que a princípio não tem uma explicação sintática ou semântica evidente, é apenas o modo como a língua naturalmente se expressa no nível do significante, resultado dos usos que seus falantes fazem dela e, neste caso, falantes de uma comunidade discursiva (Swales, 1990; Bakhtin, 1997) de um âmbito de especialidade.

Com este segundo estudo-piloto, evidenciamos o tipo de associações que gostaríamos de detectar com mais precisão e sistematicidade para que, depois, elas possam ser introduzidas como uma base de dados em uma ferramenta. No entanto, ainda trabalhamos com um método muito manual, o que ainda tornava de certa forma inviável fazer uma cobertura extensa das associações.

Esse fator poderia não nos preocupar demasiadamente porque chegamos a pensar, em um primeiro momento, que, para o sistema automático de auxílio ao tradutor que queremos desenvolver, talvez não fosse necessário incluir essa espécie de *dicionário de associações* que estamos buscando. Se trabalhássemos com um sistema do tipo memória de tradução baseada em exemplos, os exemplos poderiam ser evidenciados conforme o percentual de semelhanças entre o texto a ser traduzido e o exemplo, independentemente das palavras ou associações que os compõem. Em um segundo momento, porém, pensamos que seria importante, sim, que o sistema reconhecesse essas associações, pois será esse fator que diferenciará a ferramenta que objetivamos em relação aos sistemas usuais de memórias de tradução. O reconhecimento de tais associações faz parte da competência de um tradutor experiente, uma vez que elas representam um conhecimento necessário para que o texto traduzido torne-se fluente ou legítimo na língua de chegada. Por esse motivo, decidimos avançar no método para detecção de associações.

Como dissemos, os métodos seguidos até aqui demandam muitas análises e trabalho manual. Visando ao avanço da pesquisa, nosso próximo passo foi utilizar ferramentas que contabilizassem as associações entre as palavras a partir do cálculo de Informação Mútua⁴³. A Informação Mútua é uma medida de estatística lexical que representa “quanto uma palavra pode saber sobre a seguinte”. Os resultados e métodos utilizados em mais um estudo-piloto, nossa terceira incursão de investigação, estão relatados na seção a seguir.

3. Terceiro estudo-piloto: abordagem via cálculo de Informação Mútua

Com o intuito de avaliar que outras ferramentas e cálculos, normalmente usados em Linguística de Córpus, poderiam nos ajudar a detectar essas associações entre as palavras, fizemos alguns testes com o cálculo de Informação Mútua.

⁴³ Não esperávamos que o resultado do cálculo de MI fornecesse diretamente as associações, mas que ele fornecesse um respaldo estatístico para que determinados itens lexicais sejam considerados como tendo uma associação significativa e que, a partir deles, pudéssemos identificar outras possibilidades de associação

Conforme esclarecido na página da ferramenta que calcula essas estatísticas que utilizamos (*Analisador de tabela de colocados do WordSmith Tools (cálculo de Mutual Information e T-Score)*), disponível em <http://www2.lael.pucsp.br/corpora/index.htm>):

- As estatísticas de associação informam se duas palavras coocorreram mais vezes do que o esperado por acaso.
- Quando o valor da estatística de associação para um par de palavras (o nóculo e o colocado) iguala ou ultrapassa o valor mínimo exigido, então pode se considerar que o par de palavras forma uma colocação dentro do córpous ou texto em questão.

São apresentadas três maneiras (razão observado/esperado; informação mútua e escore T) estatísticas de calcular se a associação entre itens lexicais pode ser considerada não aleatória a ponto de poder ser considerada uma colocação; em outras palavras, são demonstradas maneiras de se saber se a associação é mais comum do que o esperado.

Nosso objetivo com este estudo foi o de verificar se os resultados desses cálculos poderiam ajudar na identificação das associações. Este terceiro estudo-piloto é aqui apresentado pela descrição do método empregado e pelo quadro dos resultados obtidos.

3.1. Metodologia

Utilizamos os quadros de colocados produzidas no piloto descrito anteriormente para fazer as análises no recurso *Analisador de tabela de colocados do WordSmith Tools (cálculo de Mutual Information e T-Score)*, disponível em <http://www2.lael.pucsp.br/corpora/index.htm>.

Neste recurso, primeiramente é necessário fazer o *upload* da lista de palavras do córpous, o qual recebe um código. Nosso córpous ficou assim classificado:

Introdução	– código 720382
Métodos	– código 501401
Resultado	– código 391160
Conclusão	– código 118888

O próximo passo consistiu em colar, em uma área determinada da página do recurso, a lista de colocados da palavra de busca que procuramos e outras informações, como mostra a figura a seguir:

Ajustes

Código da lista de palavras **em formato txt** enviada (seis dígitos):

Nóculo:

Frequência do nóculo:

Colar lista de colocados, **com todas as colunas**, no formato WordSmith Tools 3, ordenada de qualquer maneira no espaço abaixo (ao colar, as colunas ficarão desalinhadas, mas não tem problema).

N	WORD	TOTAL	LEFT	RIGHT	L5	L4	L3	L2	L1
	*	R1	R2	R3	R4	R5			
1	MESES	181	14	12	4	8	2	0	0
	155	0	0	2	8	2			
2	ANOS	29	18	11	5	0	8	5	0
	0	0	0	10	0	1			
3	IDADE	29	10	19	1	5	4	0	0
	0	0	18	0	0	1			
4	COM	23	15	8	7	6	0	2	0
	0	0	4	2	2	0			

Horizonte (igual para esquerda e direita):

Ordenar resultados por:

Figura 1 – Tela do analisador da tabela de colocados

Ao clicarmos no botão *Fazer*, os dados são calculados e os resultados apresentados da seguinte forma:

Quadro 9 – Exemplo do cálculo do Quadro de Colocados apresentando os valores de Mutual Information e T-score

#	Colocado	MI	T
0001	COM	12.625	4.796
0002	FOI	12.493	3.316
0003	MESES	9.325	13.433
0004	SEIS	7.809	4.452
0005	QUATRO	7.333	4.661
0006	ANOS	6.444	5.323
0007	VIDA	6.326	3.561
0008	MAIS	5.878	3.405
0009	MÉDIA	5.874	3.260
0010	IDADE	5.775	5.287
0011	APÓS	5.446	3.785
0012	ENTRE	4.906	3.867
0013	CRIANÇAS	4.206	3.276
0014	NOS		
0015	AOS		

Para Informação Mútua são considerados resultados estatisticamente significantes os valores acima de três (3).

O cálculo do quadro de colocados para *transversal* é apresentado a seguir.

Quadro 10 – Cálculo do quadro de colocados para transversal

#	Colocado	MI	T
0001	FOI	15.087	3.464
0002	FORAM	13.784	2.236
0003	OBSERVACIONAL	10.722	2.235
0004	CORTE	10.046	2.826
0005	TIPO	8.377	2.638
0006	REALIZADO	8.233	2.990
0007	ESTUDO	7.411	4.871
0008	VOL		

3.2. Considerações sobre os resultados

No estudo-piloto anterior, da análise dos colocados (seção 2 deste capítulo), identificamos duas associações para a palavra meses: *idade média de x meses* e *com média de x meses* e itens lexicais que se associam a essas estruturas. Pela análise manual, concluímos que *idade média de x meses* se associaria com itens como *crianças*, *sexo feminino*, *sexo masculino* e os verbos na voz passiva *avaliar* e *compor*, na seção *Métodos*. Já, *com média de x meses* se associaria com *tempo* e com o verbo *variar*. Conforme a análise de Informação Mútua acima, *meses*, *média*, *crianças* constituem-se em associações significativas.

Para *estudo transversal*, havíamos identificado, manualmente, que *foi realizado* e *trata-se de* estariam associados ao referido sintagma, além de a uma palavra relacionada à caracterização do estudo: *observacional*. Para *corte transversal*, observamos a associação com as palavras *estudo* e *tipo*, com os verbos *realizar* e *tratar* e, ainda, com *retrospectivo*, *descritivo*, *de base populacional*, *epidemiológico*. Pelo cálculo de Informação Mútua, teríamos como associações significativas, do ponto de vista estatístico: *observacional*, *corte*, *tipo*, *realizado*, *estudo*.

O que esse procedimento de cálculo de Informação Mútua nos mostra é que algumas das palavras que havíamos identificado, observando as associações manualmente foram detectadas também estatisticamente e tiveram um índice de MI (Mutual Information) significativo. Além disso, podemos chegar naquelas associações que não aparecem como significativas no resultado do cálculo, como por exemplo, *sexo feminino* com a palavra de

busca *idade média*, quando estamos avaliando os contextos de ocorrência de outras associações que têm MI significativo.

4. Quarto estudo-piloto: avaliando a mesma palavra em diferentes seções e seleção de palavras-nodo

O objetivo deste experimento foi detectar as diferentes associações de nível sintagmático com a palavra *crianças* nas diferentes seções do artigo. Essa palavra, apesar de nos acompanhar desde as primeiras reflexões desta tese, não foi, neste estudo-piloto, selecionada aleatoriamente. É uma palavra que se mostrou presente em todas as seções dos artigos do Jornal de Pediatria no método que seguimos para detectar palavras de busca. Essas palavras de busca serão depois utilizadas na Etapa 2 da investigação da hipótese, no Capítulo a seguir. Elas servem como ponto de partida para detectar as palavras que a elas se associam. Apresentamos o método de detecção de palavras de busca a seguir e, posteriormente, a análise da palavra *crianças* nas quatro seções dos artigos.

4.1. Detecção de palavras de busca

A seguir, os passos seguidos para fazer a detecção de palavras de busca. Foram detectadas palavras que são exclusivas de cada seção, dentro de um limite de corte, como ficará explicado, e também as palavras que estão presentes nas quatro seções dos artigos.

a) *Wordlist* de cada seção com recurso de *stopwords* ativado

Utilizamos nessa etapa o *WordSmith Tools* v. 3.0 e o configuramos para que não considerasse em seu processamento as palavras constantes em uma lista de *stopwords*⁴⁴. As listas foram feitas para os quatro subcorpúis em separado: Introdução, Métodos, Resultados e Discussão. Fizemos também uma *wordlist* com o recurso de *stopwords* ativado para um Corpúis de língua geral, o Lácio-Web⁴⁵.

b) Comparação de *Wordlists*

⁴⁴ *Stopwords* são palavras com alta frequência e que podem ser eliminadas de processamentos de corpúis, dependendo dos objetivos almejados. A lista das *stopwords* utilizadas nesta etapa pode ser encontrada no Anexo II. Esta lista foi refinada a partir de uma lista maior, que nos foi fornecida pela Profa. Viviane Orengo, do Instituto de Informática da UFRGS.

⁴⁵ Para uma descrição do corpúis, ver seção 2.2, Capítulo V.

Utilizamos a ferramenta *Compare 2 wordlists* do *WordSmith* para compararmos a lista de frequência de cada seção com a *Wordlist* do *Cópus Lácio-Web* a fim de obtermos as *keywords*. O objetivo desta etapa foi selecionar palavras, em cada seção, que se sobressaíam como mais prováveis da linguagem dos artigos do que dos textos de jornal. Com esse procedimento, detectamos o fator de *keyness*⁴⁶. Quando o valor de *p* para uma palavra é muito baixo e o de *keyness* muito alto, pode-se dizer que esta palavra é uma *keyword*, ou palavra-chave, conforme terminologia do aplicativo usado. Neste procedimento, obtivemos as *keywords* para cada seção. No entanto, as listas ainda ficaram muito extensas (Introdução – 283 palavras; Métodos – 462; Resultados – 523; Discussão – 516). Por isso, fizemos um recorte metodológico, estabelecendo o limite de 50 *keywords* para cada seção. Percebemos, então, que algumas palavras ocorriam em mais de uma seção, por isso realizamos o procedimento descrito a seguir.

c) Comparação de *keyness*

Com o intuito de selecionar palavras de busca para posterior análise de associações, fizemos um quadro comparativo, que pode ser visto a seguir. Com isso, conseguimos selecionar as palavras que: i) são estatisticamente significativas no *cópus* de artigos médicos (na etapa descrita em *b*) e ii) são exclusivas de cada seção (etapa *d*). Também selecionamos as que eram comuns a todas as seções. Neste quadro, as listas de 50 *keywords* foram organizadas em ordem alfabética para podermos melhor fazer as comparações.

⁴⁶ O fator de *keyness*, segundo Mike Scott, autor do pacote de ferramentas *WordSmith*, quando indicado positivamente (que foi o que consideramos), representa que uma palavra ocorre com mais frequência do que se consideraria aleatoriamente em relação ao *cópus* de referência.

Quadro 11 – Listas de palavras-chave de cada seção

Introdução	Métodos	Resultados	Discussão	
adolescentes	adolescentes	aleitamento	achados	Exclusivas da seção Introdução
AI	amostra	apresentaram	adolescentes	Exclusivas da seção Métodos
aleitamento	amostras	apresentavam	al	Exclusivas da seção Resultados
alterações	análise	característica+	aleitamento	Exclusivas da seção Discussão
amamentação	após	casos	amostra	Presentes em todas as seções
asma	através	crianças	asma	Presentes só na Introdução e Discussão
aumento	avaliação	curva	associação	Outras
avaliar	c	diagnóstico	autores	
clínica	coleta	diferença	avaliação	
clínicos	comitê	distribuição	casos	
criança	criança	entre	crianças	
crianças	crianças	escore	dados	
desenvolvimento+	critérios	estatisticamente	diagnóstico	
diagnóstico	dados	estudo	dieta	
doença	diagnóstico	etária	doença	
estudo	escore	feminino	entre	
estudos	estudo	figura	estudo	
Et	ética	foi	estudos	
etária	exame	foram	et	
fatores	foi	freqüência	etária	
freqüência	foram	gestacional	fatores	
idade	g	grau	freqüência	
II	gestacional	grupo	grupo	
incidência	grupo	grupos	grupos	
infância	grupos	houve	idade	
infecção	idade	i	incidência	
Leite	kg	idade	infecção	
lesão	leite	mães	leite	
materno	mães	masculino	literatura	
mortalidade	materno	média	mães	
nascidos	ml	mediana	maior	
neonatal	nascidos	médias	materno	
nutricional	paciente	meninas	menor	
obesidade	pacientes	meses	nascidos	
objetivo	período	n	neste	
óssea	peso	nascidos	níveis	
pacientes	pesquisa	nascimento	pacientes	
Peso	protocolo	p	peso	
Pré	quadrado	paciente	pode	
precoce	qui	pacientes	precoce	
prevalência	realizada	peso	presente	
pulmonar	realizado	recém	prevalência	
recém	recém	relação	pulmonar	
relação	sexo	respectivamente+	recém	
Risco	significância	sexo	relação	
saúde	teste	significativa	resultados	

terapia	uso	tabela	risco
termo	utilizado	uso	significativa
tratamento	utilizando	valores	tratamento
Uso	variáveis	variáveis	uso

d) Palavras exclusivas de cada seção

Consolidando os dados do quadro acima, no que concerne à ocorrência das palavras exclusivas de cada seção, e também as comuns a todas, temos:

- Palavras que só ocorrem na seção *Introdução*:

alterações, amamentação, aumento, avaliar, clínica, clínicos, desenvolvimento, il, infância, lesão, mortalidade, neonatal, nutricional, obesidade, objetivo, óssea, pré, saúde, terapia, termo

- Palavras que só ocorrem na seção *Métodos*:

amostras, análise, após, através, c, coleta, comitê, critérios, ética, exame, g, kg, período, pesquisa, protocolo, quadrado, qui, realizada, realizado, significância, teste, utilizado, utilizando

- Palavras que só ocorrem na seção *Resultados*:

apresentaram, apresentavam, características, curva, diferença, distribuição, estatisticamente, feminino, figura, grau, houve, i, masculino, média, mediana, médias, meninas, meses, n, nascimento, p, respectivamente, tabela, valores

- Palavras que só ocorrem na seção *Discussão*:

achados, associação, autores, dieta, literatura, maior, menor, neste, níveis, pode, presente, resultados

- Palavras comuns a todas as seções:

crianças, diagnóstico, estudo, idade, nascidos, pacientes, peso, recém, uso

Dessa forma, neste estudo-piloto, conseguimos estabelecer um recorte metodológico para estudar uma palavra que ocorre em todas as seções e, também, para usarmos as palavras

detectadas como ponto de partida para encontrar outras associações sintagmáticas, conforme veremos no Capítulo V.

4.2. Avaliando a mesma palavra em diferentes seções

Com o objetivo de deixar mais clara a visualização dos dados, apresentamos a seguir apenas o quadro de colocados e o resultado do cálculo de MI, já descrito no terceiro estudo-piloto, para a mesma palavra em cada seção do *córpus* do *Jornal de Pediatria*.

4.2.1. Em *Introdução*

Quadro 12 – Quadro de colocados de crianças na *Introdução*

N	WORD	L5	L4	L3	L2	L1	*	R1	R2	R3	R4	R5
1	CRIANÇAS	0	0	0	0	0	294	0	0	0	0	0
2	ADOLESCENTES	0	0	0	0	0	0	0	38	0	0	0
3	ANOS	0	0	1	0	0	0	0	0	0	11	7
4	NÃO	1	0	4	0	0	0	3	2	3	5	1
5	ENTRE	2	0	0	4	8	0	1	0	1	1	1
6	MENORES	0	0	0	0	0	0	14	1	0	0	0
7	NAS	0	2	1	0	9	0	2	0	0	0	0
8	SÃO	1	1	0	0	0	0	3	3	0	2	1
9	ADULTOS	0	0	1	1	0	0	0	3	2	1	2
10	MAIS	1	1	2	0	0	0	4	1	0	1	0
11	AMAMENTADAS	0	0	0	0	0	0	3	3	2	1	0
12	ARTERIAL	0	0	2	6	0	0	0	0	1	0	0
13	COMO	1	0	0	0	0	0	0	1	1	4	2
14	DOIS	0	0	2	0	0	0	0	1	4	2	0
15	PRESSÃO	0	2	6	0	0	0	0	1	0	0	0
16	ANO	0	1	0	0	0	0	0	0	0	6	1
17	IDADE	1	1	2	0	0	0	0	3	0	0	1
18	INFECCÃO	3	2	0	1	0	0	0	1	1	0	0
19	MAIOR	2	0	0	1	0	0	0	2	0	2	1
20	PESO	1	1	0	3	0	0	0	0	1	0	2
21	PREVALÊNCIA	2	2	0	1	1	0	0	0	0	1	1
22	USO	1	0	0	6	0	0	0	0	0	0	1
23	AINDA	0	1	0	0	0	0	2	0	1	1	2
24	AVALIAÇÃO	5	0	0	1	0	0	0	0	1	0	0
25	BRASIL	1	2	0	1	0	0	0	1	0	0	2
26	CONTRA	5	0	0	0	0	0	1	1	0	0	0
27	DESSAS	0	0	0	0	7	0	0	0	0	0	0
28	ESTUDO	1	0	0	3	0	0	0	1	0	1	1
29	MESES	1	0	0	0	0	0	0	0	1	4	1
30	TAMBÉM	1	1	3	0	0	0	0	1	0	0	1
31	TÊM	0	0	1	0	0	0	1	1	1	2	1

32	TRATAMENTO	2	1	0	3	0	0	0	0	0	1	0
33	ALTO	1	1	0	0	0	0	0	3	1	0	0
34	CRÔNICA	1	1	1	0	0	0	0	0	2	0	1
35	DOENÇA	3	0	1	0	0	0	0	0	1	1	0
36	NUTRICIONAL	0	0	2	4	0	0	0	0	0	0	0
37	OBESIDADE	1	0	0	3	0	0	0	0	0	1	1
38	PAÍSES	1	0	0	0	0	0	0	3	1	0	1
39	SOBRE	1	4	0	0	0	0	0	0	0	1	0
40	TEM	1	2	0	0	0	0	1	2	0	0	0
41	ABAIXO	0	0	0	0	0	0	5	0	0	0	0
42	AL	2	1	2	0	0	0	0	0	0	0	0
43	AUMENTO	0	4	0	0	0	0	0	0	0	0	1
44	BRASILEIRAS	0	0	0	0	0	0	4	0	0	1	0
45	CLÍNICAS	1	0	1	1	0	0	0	0	0	2	0
46	CRESCIMENTO	1	1	2	1	0	0	0	0	0	0	0
47	ESSAS	0	0	0	0	2	0	0	0	1	0	2
48	ESTUDOS	1	1	0	0	0	0	0	0	1	2	0
49	EXCLUSIVAMENTE	0	0	0	0	0	0	0	4	1	0	0
50	MAIORES	0	0	0	0	0	0	3	1	0	0	1

Quadro 13 – Quadro de resultados do cálculo de MI

	Colocado	MI	T
0001	CRIANÇAS	5.150	16.664
0002	AMAMENTADAS	4.997	2.906
0003	MENORES	4.734	3.729
0004	BRASILEIRAS	4.472	2.135
0005	PORTADORAS	4.472	2.135
0006	ADOLESCENTES	4.420	5.877
0007	EXCLUSIVAMENTE	4.302	2.124
0008	SAUDÁVEIS	4.012	2.099
0009	ABAIXO	3.771	2.072
0010	TMO	3.664	2.061
0011	MAIORES	3.564	2.048
0012	CONTRA	3.313	2.381
0013	ALTO	3.274	2.198
0014	DESSAS	3.256	2.370
0015	ARTERIAL	3.231	2.681
0016	ADULTOS	3.224	2.825
0017	CRÔNICA	3.149	2.174
0018	DOIS	3.149	2.663
0019	PRESSÃO	3.071	2.645
0020	ANO	3.021	2.480

4.2.2. Em Métodos

Quadro 14 – Quadro de colocados de crianças em Métodos

N	WORD	L5	L4	L3	L2	L1	*	R1	R2	R3	R4	R5
1	CRIANÇAS	4	1	0	1	0	357	1	0	1	0	2
2	FORAM	8	8	12	6	0	0	29	4	5	4	3
3	ESTUDO	8	7	7	8	0	0	0	2	3	4	3
4	FOI	6	9	8	0	0	0	0	2	2	3	0
5	ADOLESCENTES	0	0	0	0	0	0	0	28	1	0	0
6	GRUPO	4	6	1	3	0	0	1	5	3	0	3
7	TODAS	0	0	0	24	1	0	0	0	1	0	0
8	IDADE	0	1	0	5	0	0	0	7	2	3	4
9	ANOS	1	0	0	2	0	0	0	0	1	9	8
10	INCLUÍDAS	6	2	3	1	3	0	2	1	0	0	1
11	ENTRE	0	1	1	2	1	0	2	3	4	0	3
12	NÃO	5	1	2	1	0	0	0	2	2	0	3
13	PESO	1	1	2	5	0	0	0	0	2	0	5
14	CONTROLE	2	1	6	3	0	0	0	0	1	1	0
15	ESTUDADAS	0	0	1	4	0	0	5	2	0	0	0
16	ETÁRIA	0	0	0	1	0	0	0	0	5	3	3
17	FAIXA	0	0	1	0	0	0	0	5	3	3	0
18	MÃES	1	1	0	2	0	0	0	3	4	0	1
19	COMO	2	0	2	1	0	0	0	1	3	1	1
20	DOIS	1	0	0	0	0	0	0	3	2	4	1
21	MESES	0	0	1	0	0	0	0	0	4	1	5
22	MÉDIA	0	5	0	0	0	0	0	0	3	1	2
23	SELECIONADAS	0	1	1	0	1	0	2	5	0	0	1
24	AVALIAÇÃO	0	5	2	0	0	0	0	0	0	2	1
25	SUBMETIDAS	1	0	0	0	0	0	3	4	0	1	1
26	USO	0	1	0	6	0	0	0	1	0	1	1
27	CINCO	3	0	0	0	0	0	0	0	1	4	1
28	COMPRIMENTO	0	0	0	6	0	0	0	0	1	1	1
29	GRUPOS	0	1	1	0	1	0	0	2	2	0	2
30	TOTAL	0	0	7	1	0	0	0	0	0	0	1
31	ALEITAMENTO	2	2	2	0	0	0	0	0	0	2	0
32	MATERNAL	0	2	3	1	0	0	0	0	0	1	1
33	NASCIMENTO	2	0	1	1	0	0	0	3	0	1	0
34	PAIS	2	3	0	2	0	0	0	0	0	0	1
35	PESQUISA	3	1	1	1	0	0	0	0	2	0	0
36	POPULAÇÃO	2	3	0	1	0	0	0	1	1	0	0
37	RESPONSÁVEIS	0	1	2	5	0	0	0	0	0	0	0
38	SEIS	0	0	0	0	1	0	0	2	2	2	1
39	TRATAMENTO	0	1	1	0	0	0	0	0	2	3	1
40	AMOSTRA	3	2	2	0	0	0	0	0	0	0	0
41	EXCLUÍDAS	0	1	1	3	1	0	0	1	0	0	0
42	MENORES	0	0	0	0	0	0	7	0	0	0	0
43	SEM	0	1	0	0	0	0	1	2	1	1	1
44	TRÊS	0	1	1	0	2	0	0	1	0	1	1
45	AMAMENTADAS	0	0	0	1	0	0	1	2	1	1	0

46	ATÉ	0	0	0	0	0	0	2	2	0	1	1
47	AVALIADAS	0	0	0	2	1	0	1	2	0	0	0
48	COLETA	2	1	0	2	0	0	0	0	1	0	0
49	DOENÇA	0	0	0	0	0	0	0	2	4	0	0
50	LEITE	0	1	0	0	0	0	0	0	1	0	4

Quadro 15 – Quadro de resultado do cálculo MI

	Colocado	MI	T
0001	FORAM	11.733	8.886
0002	FOI	10.375	5.473
0003	CRIANÇAS	5.576	18.756
0004	AMAMENTADAS	5.121	2.380
0005	CONSTITUÍDA	5.050	2.169
0006	NASCIDAS	5.050	2.169
0007	RECEBIAM	4.799	2.362
0008	INCLUÍDAS	4.783	4.202
0009	SUBMETIDAS	4.535	3.026
0010	MENORES	4.436	2.525
0011	PORTADORAS	4.420	2.337
0012	ACOMPANHADAS	4.273	2.122
0013	ADOLESCENTES	4.204	5.093
0014	EXCLUÍDAS	4.173	2.500
0015	ESTUDADAS	3.991	3.247
0016	TODAS	3.951	4.770
0017	ETÁRIA	3.800	3.215
0018	SELECIONADAS	3.787	3.077
0019	PARTICIPAÇÃO	3.688	2.063
0020	COORTE	3.688	2.063
0021	CONTROLE	3.671	3.449
0022	AVALIADAS	3.661	2.258
0023	INICIALMENTE	3.611	2.055
0024	FAIXA	3.536	3.166
0025	ESSAS	3.536	2.043
0026	COMPRIMENTO	3.420	2.721
0027	QUAIS	3.420	2.223
0028	EXCLUSIVO	3.398	2.025
0029	CINCO	3.214	2.678
0030	ESTAVAM	3.103	1.976

4.2.3. Em Resultados

Quadro 16 – Quadro de colocados de crianças em Resultados

N	WORD	L5	L4	L3	L2	L1	*	R1	R2	R3	R4	R5
1	CRIANÇAS	3	5	4	2	0	565	0	0	4	3	3
2	ENTRE	2	2	12	23	4	0	2	1	4	3	4
3	NAS	3	1	3	8	34	0	3	2	1	2	0
4	NÃO	5	2	1	0	0	0	13	14	6	2	6
5	FORAM	8	1	7	0	0	0	12	6	7	2	4
6	FOI	11	13	4	0	0	0	4	2	3	2	4
7	GRUPO	3	2	3	10	0	0	0	14	4	0	6

8	ADOLESCENTES	0	0	1	1	0	0	0	30	0	0	0
9	ESTUDO	1	2	5	3	0	0	0	4	6	6	2
10	MAIOR	2	2	11	6	0	0	0	2	0	3	3
11	TABELA	7	9	0	0	0	0	0	2	1	6	4
12	SEXO	0	1	2	3	0	0	0	12	3	2	4
13	APRESENTARAM	0	1	1	0	0	0	6	6	5	3	3
14	MENORES	0	0	2	1	0	0	17	0	1	2	1
15	IDADE	7	1	1	5	0	0	0	3	1	3	2
16	MÃES	0	0	2	5	0	0	0	10	2	0	4
17	MESES	2	4	6	3	0	0	0	0	1	3	3
18	DISTRIBUIÇÃO	0	4	4	11	0	0	0	0	0	0	2
19	SOBREPESO	3	3	0	0	0	0	0	4	1	0	9
20	ANOS	2	2	3	3	0	0	0	0	0	9	0
21	ALEITAMENTO	3	2	1	0	0	0	0	9	0	0	3
22	SEIS	0	1	3	0	6	0	1	0	4	0	3
23	TOTAL	1	2	4	9	0	0	0	2	0	0	0
24	VOL	4	5	7	0	0	0	0	2	0	0	0
25	MAIS	2	2	5	2	1	0	2	0	2	0	1
26	PESO	0	2	1	1	0	0	0	2	6	3	2
27	RELAÇÃO	2	8	0	2	0	0	0	0	0	3	1
28	AMAMENTADAS	1	0	0	1	0	0	4	6	2	1	0
29	P	3	2	2	0	0	0	3	1	1	2	1
30	DIAS	4	2	1	0	0	0	0	0	1	6	0
31	GRUPOS	4	0	0	6	0	0	0	0	0	4	0
32	MATERNO	1	3	2	1	0	0	0	0	7	0	0
33	OBESIDADE	0	1	2	6	0	0	0	0	0	4	1
34	TRÊS	2	2	0	1	5	0	0	0	0	2	2
35	CADA	0	0	1	3	0	0	0	8	1	0	0
36	RECEBERAM	0	1	0	0	0	0	2	3	6	1	0
37	DUAS	1	0	2	0	9	0	0	0	0	0	0
38	NÚMERO	0	0	2	10	0	0	0	0	0	0	0
39	SEGUNDO	0	0	0	0	0	0	2	2	3	4	1
40	ESCOLA	0	1	0	0	0	0	0	2	2	5	1
41	PAIS	0	0	0	8	0	0	0	1	0	1	1
42	SIGNIFICATIVAMENTE	1	4	3	0	0	0	0	1	0	1	1
43	TIVERAM	2	0	0	0	0	0	0	4	1	3	1
44	ANO	2	1	0	0	0	0	0	0	0	6	1
45	APRESENTAVAM	0	1	0	0	0	0	5	1	1	0	2
46	AVALIADAS	0	0	0	1	0	0	7	1	0	1	0
47	CARACTERÍSTICAS	1	2	3	2	0	0	0	1	0	1	0
48	FEMININO	1	0	0	0	0	0	0	0	7	2	0
49	FIGURA	3	4	1	0	0	0	0	0	0	2	0
50	MAIORES	0	2	0	1	0	0	6	0	1	0	0

Quadro 17 – Quadro de resultado do cálculo MI

	Colocado	MI	T
0001	ACOMPANHADAS	5.146	2.572
0002	BRANCAS	5.094	2.913
0003	CRIANÇAS	4.984	23.503

0004	OSTOMIZADAS	4.923	2.370
0005	INTERNADAS	4.923	2.164
0006	PORTADORAS	4.701	2.356
0007	QV	4.465	2.701
0008	EXCLUÍDAS	4.410	2.523
0009	AMAMENTADAS	4.371	3.688
0010	IMIP	4.339	2.329
0011	SO	4.339	2.329
0012	SOROLOGIA	4.186	2.317
0013	GMTLMAI	4.186	2.317
0014	CUJAS	4.117	2.666
0015	PARTICIPARAM	4.075	2.104
0016	OBESAS	4.075	2.104
0017	IGUAIS	4.049	2.302
0018	INCLUÍDAS	3.924	2.288
0019	TRATADAS	3.924	2.090
0020	AVALIADAS	3.853	2.944
0021	ADOLESCENTES	3.735	5.233
0022	CIRROSE	3.661	2.059
0023	PROTOCOLO	3.644	2.436
0024	MENORES	3.626	4.503
0025	ELEVADO	3.602	2.249
0026	NEGATIVA	3.439	2.030
0027	NAS	3.355	6.814
0028	SOBREPESO	3.315	4.023
0029	NORMAIS	3.075	2.787
0030	AGUDA	3.049	2.638

4.2.4. Em Discussão

Quadro 18 – Quadro de colocados de crianças em Discussão

N	WORD	L5	L4	L3	L2	L1	*	R1	R2	R3	R4	R5
1	CRIANÇAS	1	3	1	0	0	778	1	0	1	3	1
2	ENTRE	3	1	2	33	13	0	2	4	1	5	3
3	ADOLESCENTES	1	0	0	1	0	0	0	63	0	1	0
4	NÃO	8	6	9	0	0	0	11	13	6	4	3
5	ANOS	0	1	2	0	0	0	0	0	3	28	15
6	MAIS	4	6	11	1	0	0	13	4	3	4	3
7	ESTUDO	13	3	7	10	0	0	0	1	6	3	2
8	NAS	2	0	0	0	37	0	0	0	1	0	2
9	FOI	6	5	9	0	0	0	2	1	3	5	8
10	IDADE	3	3	3	3	0	0	0	6	6	5	4
11	MAIOR	4	2	9	5	0	0	0	4	2	2	4
12	FORAM	1	2	8	0	0	0	5	5	4	3	3
13	GRUPO	2	0	2	8	0	0	0	11	1	2	2
14	COMO	5	6	3	0	1	0	1	2	4	2	2
15	AL	7	7	5	0	0	0	0	0	0	2	4
16	DBP	2	0	0	0	0	0	0	14	4	0	0
17	DIETA	0	0	0	1	0	0	0	14	5	0	0
18	MENORES	0	0	0	0	0	0	17	1	1	1	0
19	AMAMENTADAS	0	0	0	1	0	0	10	5	2	1	0
20	ET	7	5	0	0	0	0	0	0	2	4	1

21	MÃES	1	1	0	4	1	0	0	8	1	2	0
22	SEM	1	1	1	0	0	0	10	2	0	0	2
23	MESES	0	1	1	3	1	0	0	0	0	6	4
24	ASMA	1	1	2	4	0	0	0	3	2	0	2
25	ATÉ	1	0	2	0	0	0	2	3	4	2	1
26	LEITE	1	0	0	1	0	0	0	0	2	3	8
27	PESO	1	0	2	3	0	0	0	1	4	0	4
28	QUANDO	0	2	6	0	0	0	1	1	1	1	3
29	DESSAS	0	1	1	1	10	0	0	0	0	0	1
30	DOIS	1	1	3	0	0	0	0	3	3	2	1
31	ESTUDADAS	0	0	0	1	0	0	11	0	0	0	2
32	MAIORES	0	0	1	1	0	0	8	2	0	0	2
33	MUITO	1	3	0	0	0	0	0	2	7	1	0
34	PRINCIPALMENTE	1	1	0	7	0	0	0	1	2	1	1
35	DIARRÉIA	2	0	0	2	0	0	0	4	2	3	0
36	NOSSO	1	2	4	0	0	0	0	2	1	1	2
37	SÃO	2	1	2	1	0	0	2	1	2	2	0
38	SOBRE	3	2	2	1	0	0	0	0	0	2	3
39	SUBMETIDAS	0	0	0	0	0	0	8	2	0	1	2
40	TAMBÉM	1	1	0	3	0	0	2	2	0	1	3
41	USO	3	3	0	1	0	0	0	3	2	0	1
42	ANO	1	3	0	1	0	0	0	1	0	6	0
43	APRESENTARAM	0	0	0	0	0	0	2	4	2	3	1
44	DIAGNÓSTICO	1	4	3	2	0	0	0	0	0	2	0
45	MENOR	1	2	0	4	0	0	0	3	0	1	1
46	PRESENTE	1	2	2	1	0	0	0	0	2	0	4
47	VIDA	3	4	1	4	0	0	0	0	0	0	0
48	EXCLUSÃO	1	0	0	0	0	0	0	1	0	7	2
49	FATO	4	0	0	3	0	0	0	0	2	1	1
50	GRUPOS	0	1	0	6	0	0	0	1	1	2	0

Quadro 19 – Quadro de resultado do cálculo MI

	Colocado	MI	T
0001	NÃO	10.970	7.743
0002	ADMITIDAS	5.392	2.184
0003	CRIANÇAS	5.091	27.266
0004	OSTOMIZADAS	5.071	2.171
0005	PREMATURAS	5.071	2.171
0006	NASCIDAS	5.071	2.171
0007	VELHAS	5.071	2.171
0008	SADIAS	5.071	2.171
0009	INFECTADAS	4.919	2.901
0010	INTELECTUAL	4.808	2.157
0011	APROPRIADOS	4.808	2.157
0012	VALENTE	4.808	2.157
0013	TVP	4.808	3.050
0014	QV	4.749	2.724
0015	AMAMENTADAS	4.734	4.196
0016	TRATADAS	4.657	2.354
0017	ALIMENTADAS	4.586	2.144
0018	INTERNADAS	4.486	3.021
0019	ESTUDADAS	4.178	3.536

0020	EXCLUSIVAMENTE	4.071	2.305
0021	SUBMETIDAS	4.071	3.392
0022	PORTADORAS	4.071	2.660
0023	INCLUÍDAS	4.071	2.104
0024	PEQUENAS	4.071	2.305
0025	ISENTA	4.071	2.660
0026	ADOLESCENTES	3.986	7.612
0027	CUJAS	3.971	2.478
0028	PROVENIENTES	3.934	2.090
0029	RECEBIAM	3.934	2.090
0030	EXCLUSÃO	3.830	3.085
0031	COLESTASE	3.808	2.077
0032	AQUELAS	3.749	2.928
0033	DBP	3.693	4.126
0034	TROMBOSE	3.655	2.256
0035	JOVENS	3.631	2.434
0036	CRECHES	3.586	2.050
0037	HIV	3.540	2.743
0038	DESSAS	3.486	3.409
0039	COMPARADAS	3.409	2.221
0040	CICATRIZES	3.409	2.221
0041	RENAIS	3.306	2.010
0042	FAMÍLIAS	3.264	2.196
0043	CRÔNICA	3.264	2.535
0044	MENORES	3.222	3.993
0045	DIARRÉIA	3.186	3.210
0046	PLASMÁTICOS	3.145	1.985
0047	PROPORÇÃO	3.111	2.653
0048	DIETA	3.090	3.947
0049	ABAIXO	3.071	2.332
0050	RGEP	3.071	2.160
0051	PAIS	3.000	2.768

4.3. Resultados para a análise da ocorrência de *crianças* em todas as seções

O objetivo do experimento de avaliar os colocados e o cálculo de MI para a mesma palavra nas quatro sessões era o de perceber, estatisticamente, quais as palavras que se associavam a *crianças* e se elas seriam diferentes para cada sessão.

Após análise dos quadros, foi possível elaborar algumas conclusões:

- as palavras que se associam a *crianças* em cada uma das seções são, em sua maioria, exclusivas de cada seção. Isso equivale a dizer que as palavras se associam de formas distintas conforme a seção. Assim, por exemplo, palavras que se mostraram associadas a *crianças* só na *Introdução* foram: *adultos, brasileiras, saudáveis*, entre outras. Já na seção *Métodos* foram: *constituída, controle, coorte, selecionada*. Na seção *Resultados*, encontramos: *elevado, negativas, normais, participaram*. Por fim, na seção *Discussão* obtivemos: *admitidas, comparadas, alimentadas*;

- a seção *Introdução* e a seção *Discussão* são as que mais apresentam coincidência nas palavras que se associam a *crianças*. Acreditamos que isso se deva à relação que ambas têm, pois os tópicos mencionados na *Introdução* geralmente são retomados na *Discussão*, conforme orientam os Requisitos Uniformes: “Relacione as conclusões com os objetivos do estudo”;
- a seção *Discussão* é aquela em que as palavras associadas a *crianças* são menos “características” da seção. Por outro lado, a seção *Métodos* demonstrou ter palavras bem prototípicas, como *coorte*, *controle*, *constituída*. Em *Resultados* também encontramos palavras bem características, como *elevado*, *negativas*, *normais*, palavras que fazem uma apreciação/caracterização dos resultados.

5. Conclusões

Após as experiências realizadas nesses quatro estudos-piloto, concluímos que poderíamos responder apenas parcialmente nossa primeira questão de pesquisa, que questiona a possibilidade de identificar essas associações relevantes de nível sintagmático presentes nos artigos médicos de Pediatria. Mencionamos *parcialmente* porque os resultados que obtivemos demonstrariam, a princípio, nossos pressupostos e hipótese. Porém, havia muita subjetividade, e a detecção das associações necessitava ainda de um esforço manual muito grande, e nossa intenção era identificar as associações em larga escala, visando a alimentação de um banco de dados maior. Esse mesmo estudo que fizemos com a palavra *crianças*, fizemos com outras quatro palavras, que detectamos como exclusivas de cada seção⁴⁷. Esses resultados não se encontram nesta tese, pois acreditamos que eram uma extensão do terceiro estudo-piloto e, apesar de termos identificado as associações, como dissemos, o método necessitava uma análise qualitativa de cada linha de contexto para estabelecer uma análise sistemática dos resultados.

Depois de realizar esses estudos iniciais, deparamo-nos com o software *ConcGram*, que fazia exatamente o que havíamos imaginado como uma necessidade: detectava decontinuidades. Assim, utilizamos esse recurso para fazer um levantamento das associações

⁴⁷ A detecção das palavras exclusivas foi feita em uma comparação de listas de *keyness*. Como essas palavras foram usadas na Etapa 2 de nossa investigação em larga escala, tal método de identificação de palavras-nodo será descrito na seção apropriada no capítulo a seguir.

tanto para fins de comprovar a hipótese, o que descrevemos na Parte I, como para alimentar a proposta da ferramenta que será descrita na Parte II desta tese.

CAPÍTULO V

INVESTIGAÇÃO DA HIPÓTESE: MÉTODOS E RESULTADOS

A hipótese que nos propomos a investigar é a de que existem, nos textos especializados, associações no nível sintagmático que:

- a) são combinações mais ou menos fixas e frequentes de palavras;
- b) não são necessariamente contínuas;
- c) estão associadas, no nível sintagmático, por uma atração semântica (palavras que andam juntas no mesmo contexto - no limite da frase), e/ou pragmática (palavras que andam juntas para desempenhar funções no texto - também no limite da frase), e/ou gramatical;
- d) não necessariamente contêm um termo em sua composição, aqui termo entendido como uma expressão lexical associada a um conceito.

Para evidenciar a hipótese e o desdobramento de que essas associações podem ser também Unidades de Tradução Especializadas, a análise dos dados será apresentada em diferentes etapas.

Na Etapa 1, para comprovar a flexibilidade e as possibilidades de formação dessas associações (itens “a” e “b”), apresentamos os *congrams* das palavras-nodo e seus contextos nos subcorpú de ocorrência de cada uma. Optamos por uma abordagem que seria mais *direcionada pelo córpus*⁴⁸, ou seja, as palavras-nodo que selecionamos são os substantivos mais frequentes de cada seção, elas não foram previamente selecionadas por outro critério. Depois, processamos os *congrams* para cada uma dessas palavras-nodo e analisamos os

⁴⁸ Esse termo é conhecido em Linguística de córpus também pelo termo em inglês *corpus-driven* (Tognini-Bonelli, 2001).

resultados. Cada palavra-nodo resulta em uma série de *concgrams*. Procuramos analisar os 8 primeiros *concgrams* para cada palavra-nodo, desde que eles realmente representassem uma associação e não fossem meras coocorrências de palavras.

Na Etapa 2, que trata do item “c”, fazemos uma análise qualitativa, usando palavras-nodo pré-selecionadas que julgamos com potencial para demonstrar com mais evidência a característica em foco. Poder-se-ia dizer que essa segunda etapa é mais *baseada em córpus*, pois o utilizamos como elemento de evidência. As palavras selecionadas para análise, nessa etapa, foram palavras que, além de serem palavras-chave detectadas em comparações de uma lista de *keyness*, são também exclusivas de cada seção dos artigos.

O item “d” envolve avaliar o caráter especializado dessas unidades, mesmo que elas não contenham em sua composição uma palavra-termo. Para isso, na Etapa 3, fazemos comparações entre os *concgrams* de determinadas palavras-nodo em um córpus especializado (do Jornal de Pediatria), em um córpus de textos jornalísticos e em um outro córpus especializado, de contraste. Isso para demonstrar em que medida o significado especializado também se constrói por meio de uma rede de relações, mesmo sem a presença de uma palavra-termo.

Na Etapa 4, tratamos de demonstrar como essas unidades podem ser consideradas unidades de tradução. Para isso, fazemos um cotejamento entre algumas unidades e como foram traduzidas.

A nossa hipótese e investigação têm o intuito de demonstrar mais um critério, ou particularidade, dos textos especializados, usando para isso ferramentas da Linguística de Córpus que não estavam disponíveis há 10 anos, como o software *ConcGram*.

Iniciamos esse capítulo de investigação da hipótese com uma descrição do software que utilizamos nas Etapas de 1 a 3, considerando também que o próprio se configura como uma resposta à nossa questão de pesquisa sobre a possibilidade de detectar as associações em uma sequência não contínua. Em seguida, descrevemos os córpus que utilizamos, embora nem todos sejam utilizados em todas as etapas. O restante do capítulo é organizado conforme as etapas de investigação, uma vez que cada uma delas tem particularidades metodológicas e resultados parciais. Dentro de cada etapa, descrevemos o método seguido para o objetivo em questão, as análises conduzidas e nossas conclusões a respeito. Ao fim do capítulo, é feito um fechamento, considerando a confirmação da hipótese com base nas análises apresentadas.

1. Descrição do aplicativo *ConcGram*

Para conseguir responder à nossa questão de pesquisa quanto à possibilidade de detectar as unidades delineadas acima, comprovar sua presença nos textos do *córpus* e comprovar nossa hipótese, foram feitos uma série de experimentos utilizando as ferramentas e métodos disponibilizados pela Linguística de *Córpus* (conforme já descritos no capítulo anterior). Todos os experimentos, embora tenham sido válidos e tenham nos mostrado dados interessantes, alguns secundários a essa pesquisa, não se mostravam muito práticos, e a entrada de pesquisas manuais era necessária muito cedo no processo, o que acabaria tomando muito tempo para sistematizar os resultados, além do que estes seriam talvez mais subjetivos.

A ferramenta *ConcGram* apresentou-se como a solução para a nossa questão de pesquisa. Antes, tínhamos uma pergunta aberta: como detectar esse tipo de estrutura, de associação? Até onde sabíamos, nenhuma ferramenta disponível poderia detectar associações que não fossem contíguas e sem delimitação do número de palavras entre os elementos. Por isso, fizemos abordagens estatísticas usando apenas listas de palavras e a ferramenta *Concord* do *WordSmith Tools*.

O *ConcGram*, por sua vez, possibilita fazer buscas por palavras que ocorrem em um mesmo contexto (um limite de 50 caracteres, aproximadamente 12 palavras). Pode-se selecionar quantas palavras coocorrentes se deseja incluir na busca (de 2 a 5). É possível fazer buscas a partir de palavras selecionadas e também buscar todas as associações, ou seja, todas as palavras coocorrentes (ou *concgrams*), o que permite uma busca totalmente *corpus-driven*. Essas coocorrências são denominadas *concgrams* e incluem tanto variação de constituição (exemplo extraído do manual: *increase in expenditure, increase in the share of public expenditure*) quanto de posição (*expenditure would inevitably increase*) (Greaves, 2009). Com esse sistema, obtêm-se listas de *concgrams* compostos de 2, 3, 4 e 5 palavras.

Todo o processo de busca de *concgrams* acontece a partir da criação da lista de *ConcGrams* de 2 palavras. A seguir, mostramos algumas figuras com telas do software, usando *crianças* como palavra de busca no *córpus* composto das *Introduções* dos artigos. A criação das listas é progressiva, isto é, só é possível fazer a lista subsequente (*concgrams* de 3 palavras) após fazer a lista dos *concgrams* de 2 palavras e assim por diante até 5.

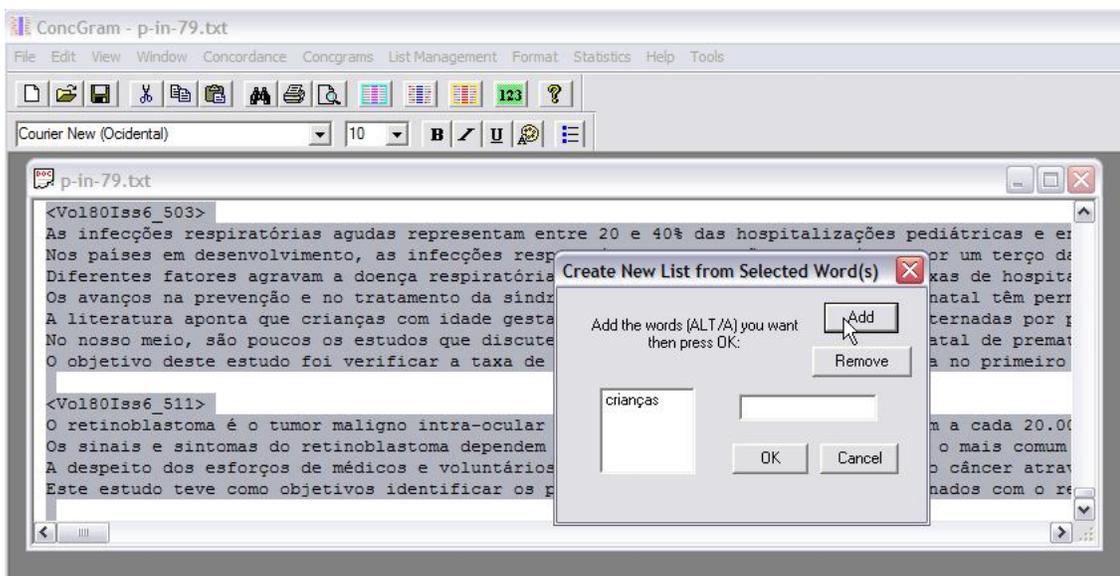


Figura 2 – Texto a ser investigado (p-in-79.txt) e seleção da palavra de busca

Ao clicar no botão OK, o software inicia as buscas dos concgrams para a palavra selecionada e apresenta os resultados em uma listagem.

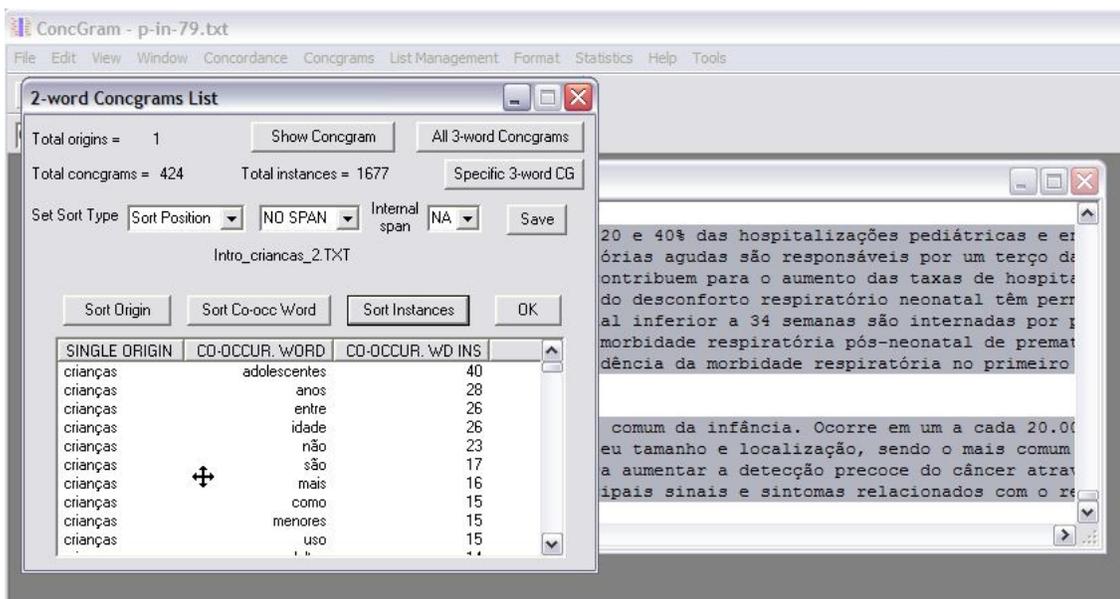


Figura 3 – Lista dos concgrams de duas palavras

A seguir mostramos exemplos das janelas com os concgrams de três, quatro e cinco palavras para *crianças* na sessão *Introdução*:

3-Word Congrams List

Total origins = 209 Show Congram All 4-word Congrams

Total congams = 550 Total instances: 1269 Specific 4-word CG

Set Sort Type Sort Position NO SPAN Internal span NA Save

Intro_criancas_3.TXT

Sort Double Origin Sort Co-occ Word Sort Instances OK

DOUBLE ORIGIN	CO-OCC WORD	CO-OCC WORD INST
crianças / abaixo	idade	5
crianças / adolesce...	sobrepeso	4
crianças / adolesce...	amostra	3
crianças / adolesce...	assim	3
crianças / adolesce...	Brasil	3
crianças / adolesce...	obesidade	3
crianças / adolesce...	peroxidação	3
crianças / adolesce...	TMD	3
crianças / adolesce...	arterial	2

Figura 4 – Lista dos *congams* de três palavras

4-Word Congrams List

Unique origins = 167 Show Congram All 5-word Congrams

Unique congams = 239 Total instances: 508 Specific 5-word CG

Set Sort Type Sort Position NO SPAN Set span NA Save

Intro_criancas_4.TXT

Sort Triple Origin Sort Co-occ Word Sort Instances OK

TRIPLE ORIGIN	CO-OCC ...	CO-OC WD INS
crianças / adolescentes / arterial	pressão	2
crianças / adolescentes / aterosclerose	fatores	2
crianças / adolescentes / aterosclerose	risco	2
crianças / adolescentes / causa	mortalidade	2
crianças / adolescentes / DAC	familiar	2
crianças / adolescentes / DAC	história	2
crianças / adolescentes / DM0	DC	2
crianças / adolescentes / estatura	peso	2
crianças / adolescentes / familiar	história	2
crianças / adolescentes / história	risco	2

Figura 5 – Lista dos *congams* de quatro palavras

5-Word Congrams

Total origins = 61 Total congams = 81 Total instances = 168 Save

Set Sort Type Sort Position NO SPAN Set span NA Show Congram

Intro_criancas_5.TXT

Sort Quadruple Origin Sort Co-occ Word Sort Instances OK

QUADRUPLE ORIGIN	CO-OCC WORD	CO-OC WD INS
crianças / adolescentes / aterosclero...	risco	2
crianças / adolescentes / DAC / famil...	história	2
crianças / adulta / arterial / hipertensão	idade	2
crianças / adulta / arterial / hipertensão	pressão	2
crianças / adulta / arterial / idade	pressão	2
crianças / adulta / hipertensão / idade	pressão	2
crianças / al / Anene / et	estudo	2
crianças / alto / c / entre	lkpeng	2
crianças / alto / c / entre	obtido	2
crianças / alto / c / entre	Xingu	2

Figura 6 – Lista dos *congams* de cinco palavras

Ao clicar no botão *Save*, o sistema salva uma lista no formato .txt com todas as coocorrências das palavras e a frequência com que aquela sequência se repete.

Para ver as linhas de concordância de cada coocorrência, é necessário selecionar o item desejado e clicar em *Show Concgram*. Uma janela com os contextos para aquela coocorrência é exibida:

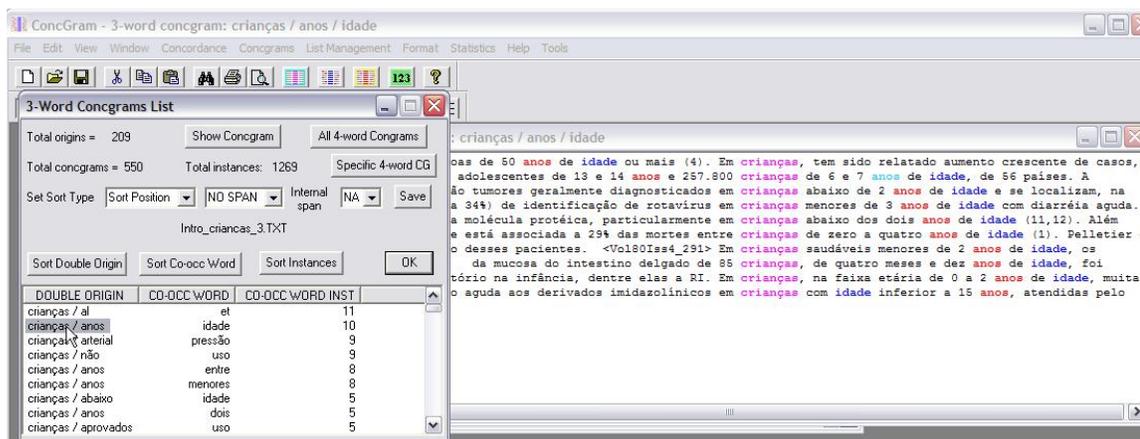


Figura 7 – Linhas com os contextos do *concgram* selecionado

Como já esclarecemos no Capítulo I, é importante salientar que os resultados do *ConcGram* compreendem todas as coocorrências de palavras, mas não determinam se elas têm uma associação em termos de significado ou se são apenas coocorrências casuais de palavras. O termo coocorrência diz respeito a qualquer palavra que ocorre dentro de um espaço de vizinhança em relação à outra palavra; uma palavra associada é aquela cuja coocorrência não é acidental, mas forma um colocado com a outra palavra (Greaves, 2009). Levando essa característica em consideração, após fazer as listas dos *ConcGrams* de duas, três, quatro e cinco coocorrências, é preciso fazer uma avaliação manual dos resultados, acessando as linhas de concordância de cada coocorrência.

Levando em conta as possibilidades apresentadas pelo *ConcGram*, nossa investigação poderia seguir duas frentes: fazer buscas das associações usando palavras-nodo previamente selecionadas (como mostram os exemplos acima) ou usar o recurso de buscar todas as palavras coocorrentes existentes e fazer um recorte das mais frequentes. A princípio, pensamos em trabalhar com os *concgrams* mais frequentes sem selecionar uma palavra-nodo, mas, infelizmente, o software não opera adequadamente essa função quando o idioma apresenta caracteres acentuados, como é o caso da Língua Portuguesa.

2. Descrição dos corpus

2.1. Jornal de Pediatria

O corpus do Jornal de Pediatria que utilizamos é, na verdade, um recorte do corpus inicialmente compilado no trabalho de mestrado de Robert Coulthard (2005). Em nosso recorte, excluimos as Cartas ao editor e os Artigos de revisão. Sendo assim, o corpus ficou composto de 120 artigos originais, em português, com um total de 348.862 palavras. Os textos foram publicados entre os anos de 2003 e 2004.

Quadro 20 – Número de palavras do corpus em Português

	Introdução	Métodos	Resultados	Discussão	Total
Tokens	51.971	83.367	83.076	30.448	
Types	6.995	8.648	5.859	10.473	
					348.862

O Jornal de Pediatria é uma publicação bimensal da Sociedade Brasileira de Pediatria, em circulação desde 1934. O JPED é a versão eletrônica do Jornal de Pediatria, considerada a “maior e mais completa revista de pediatria da América Latina, com mais de 15.000 leitores”⁴⁹.

Os artigos publicados no Jornal de Pediatria são indexados pelos seguintes órgãos: Index Medicus/MEDLINE; SciELO; EMBASE/Excerpta Medica; Medical Research Index; University Microfilms International; LILACS - Literatura Latino-Americana e do Caribe em Ciências da Saúde; Sociedad Iberoamericana de Información Científica (SIIC) Data Bases; Science Citation Index Expanded; Journal Citation Reports/Science Edition.

2.1.1. Jornal de Pediatria traduzido

Todos os artigos que compõem o corpus em português do Jornal de Pediatria foram também traduzidos por tradutores da empresa responsável pela editoração da revista e publicados no site do Jornal de Pediatria. Eles são, em sua grande maioria, submetidos em português e traduzidos por tradutores profissionais para configurarem, no site, uma versão bilíngue. O corpus é composto conforme o Quadro a seguir. O JPED disponibiliza a versão completa da revista em inglês desde 2000.

⁴⁹ Informação extraída da página na Web da revista: www.jpmed.com.br

Quadro 21 – Número de palavras do *cópus* em Inglês

	Introdução	Métodos	Resultados	Discussão	Total
Palavras	49.650	76.976	80.621	122.819	330.066

2.2. Lácio-Web

O Projeto Lácio-Web é uma parceria entre o NILC (Núcleo Interinstitucional de Linguística Computacional), localizado no ICMC-USP, o IME (Instituto de Matemática e Estatística) e a FFLCH (Faculdade de Filosofia, Letras e Ciências Humanas). Segundo informações do site onde é disponibilizado o *cópus*, o objetivo do projeto inclui disponibilizar livremente, na Web, *cópus* do português brasileiro escrito contemporâneo.

O Lácio-Web é composto de diferentes *cópus*, mas utilizamos para nossa pesquisa parte do *cópus* Lácio-Ref, que é composto de textos escritos em português brasileiro culto. Dentro desse *cópus*, os textos estão distribuídos em tipologias de domínio, gênero textual, tipo de texto, e meio de distribuição. Sendo assim, a seção do Lácio que utilizamos, denominada Generalidades, inclui, em termos de domínio, “textos com temas variados e, de modo geral, inseridos num campo tematizado pelo senso comum (ex.: entretenimento). Inclui, além disso, os textos que abordam, de forma não-analítica, temas considerados pela ciência (ex.: ciência e tecnologia, saúde, esporte, etc.)”⁵⁰. O *cópus* Lácio-Generalidades é composto principalmente por textos de jornal e revistas, e o gênero predominante é o informativo (reportagem, notícia). As áreas que esse *cópus* abrange são: Ambiente doméstico, Cotidiano/Comunidade, Culinária, Cultura, Empreendimento, Entretenimento, Lazer e Recreação, Esporte, Informática e Internet, Meio ambiente, Moda, Policial, Sociedade, Viagem, Celebidades/Solenidades, Educação, Dinheiro/Finanças, Saúde, Ciência e tecnologia, Justiça, entre outros. Esse *cópus* é composto por 1.683.706 palavras.

2.3. *Cópus* de contraste: Química Nova

Cópus composto por artigos científicos publicados no ano de 2004, de janeiro a agosto, da Revista Química Nova. A revista Química Nova é um periódico nacional, produzido pela sociedade Brasileira de Química e direcionado para profissionais da área da Química. O *cópus* que utilizamos foi compilado no âmbito do projeto TextQuim

⁵⁰ O *cópus* pode ser obtido na página do NILC, <http://www.nilc.icmc.usp.br/lacioweb/>.

(www.ufrgs.br/textquim) e foi aplicado por Susana de Azeredo na ocasião de desenvolvimento de pesquisa de mestrado (Azeredo, 2007). É composto por um total de 406.289 palavras. A revista inclui artigos que se encaixam nos seguintes temas: Química e sociedade, Educação em Química e multimídia, Espaço aberto, Conceitos científicos em destaque, História da Química, Atualidades em Química, Relatos de sala de aula, Pesquisa em ensino, Experimentação no ensino de Química, Elemento Químico, contando também com editoriais e resenhas. Quanto aos tipos de artigos, contém artigos⁵¹, artigos de divulgação, artigos de nota técnica, artigos de assuntos gerais, artigos de educação e artigos de revisão.

3. Etapa 1: *concgrams* e a flexibilidade das unidades de tradução

Nesta etapa, investigamos os itens *a* e *b* mencionados no início deste Capítulo, ou seja, estamos concentrados em demonstrar como associações sintagmáticas podem ter uma estrutura flexível, "não engessada", quando estão associadas para transmitirem uma mesma ideia ou para cumprirem um mesmo papel. Na Metafunção Ideacional, a "realidade" é representada por orações com três tipos de componentes: os processos, os participantes e as circunstâncias.

Nos termos da Linguística Funcional, as associações estariam transmitindo o mesmo sentido ideacional, mas diferindo no sentido textual. Na Metafunção Ideacional, a linguagem tem a função de representação da situação, ou seja, ela ajuda a codificar a vivência e a experiência do mundo.

Além das variações dos itens que acompanham as associações, também não deixamos de observar outras regularidades que se apresentaram, quando as consideramos relevantes para o trabalho. É importante ressaltar que essas outras observações (colocados, prosódia semântica, associações semânticas) devem ser aqui consideradas como itens extras à análise, por isso não foram sistematizadas.

Nessa Etapa 1, descrevemos, então, como foi feita a seleção de palavras-nodo, o processamento dos *concgrams* e apresentamos a análise dos resultados.

⁵¹ A descrição do *cópus* não define, mas entendemos que esses *artigos* corresponderiam ao que, em Medicina, se costuma denominar artigos originais.

3.1. Metodologia

3.1.1. Seleção das palavras-nodo

Embora nos estudos-piloto tivéssemos feito uma seleção de palavras-nodo com base na sua “chavidade” (*keyness*) e exclusividade em artigos médicos e seções, não tínhamos certeza se esse método apresentaria as melhores palavras-chave para nossa análise. Acabamos considerando que daquela forma poderíamos estar excluindo palavras e agrupamentos significativos cujas palavras-nodo poderiam ocorrer em duas ou mais sessões. Optamos então, na Etapa 1, por buscar os *congrams* dos primeiros substantivos mais frequentes de cada seção dos artigos médicos, extraídos de uma lista simples de palavras com suas frequências, feita com o *WordSmith Tools*.

Após fazer uma lista de frequência das palavras pertencentes a cada seção (o Anexo I apresenta as primeiras ocorrências das listas de frequências de cada uma delas), selecionamos os substantivos mais frequentes como palavras-nodo. O quadro a seguir mostra os 5 primeiros substantivos mais frequentes de cada seção, mostrando a frequência absoluta (FA) e a frequência relativa (FR) no *corpus*. Por FR entendemos a frequência normalizada por 10 mil palavras.

Quadro 22 – Cinco primeiros substantivos mais frequentes de cada seção

Introdução	FA	FR	Métodos	FA	FR	Resultados	FA	FR	Discussão	FA	FR
CRIANÇAS	294	56,57	ESTUDO	429	51,46	TABELA	746	89,80	CRIANÇAS	778	255,52
PACIENTES	151	29,05	CRIANÇAS	357	42,82	PACIENTES	729	87,75	ESTUDO	729	87,75
ESTUDOS	143	27,52	IDADE	291	34,91	CRIANÇAS	565	68,01	PACIENTES	686	82,57
ESTUDO	140	26,94	PACIENTES	291	34,91	GRUPO	461	55,49	ESTUDOS	295	35,51
SAÚDE	127	24,22	TESTE	229	27,47	IDADE	370	44,54	IDADE	287	34,55

Como Berber-Sardinha, em sua investigação (1999), não consideramos as palavras gramaticais como pontos de partida, pois pensamos que itens lexicais de conteúdo (ou “cheios”) poderiam ser mais produtivos. No caso de nossa investigação, acreditamos que eles determinam a organização das frases, eles seriam os itens fixos das unidades que buscamos. Nos referimos a *fixos* no sentido de que são os elementos que sempre estão presentes, mesmo que assumam diferentes posições dentro da frase e tenham em seu entorno diferentes palavras, gramaticais ou não. Gledhill (2000), em sua pesquisa sobre a fraseologia na área de Farmácia fez as buscas usando as palavras gramaticais mais frequentes, que ele chama de *salientes*. Nós também optamos por usar palavras salientes, mas optamos pelos substantivos. Essa decisão também é necessária nos estudos voltados à prosódia semântica, pois se diz que somente as palavras de conteúdo podem ser interpretadas como positivas, negativas, ou neutras (Berber-

Sardinha, 1999), embora haja algumas palavras lexicais que fazem parte de sintagmas com uma relação negativa, por exemplo *in broad daylight*.

3.1.2. Processamento do texto para obter *concgrams*

Após decidirmos quais seriam as palavras-nodo a serem utilizadas, o próximo passo foi fazer as buscas das palavras que mais se associam a elas, utilizando o software *ConcGram*. Como dissemos, esse software é capaz de demonstrar até 4 palavras associadas a uma palavra de busca (o que conformaria um *concgram* de cinco palavras). Consideramos que não conseguiríamos fazer uma análise consistente e bem direcionada se analisássemos todos os *concgrams* (de duas a cinco palavras) de todas as palavras-nodo selecionadas para este estudo, pois teríamos muitos dados e não é o resultado quantitativo das análises que demonstrará nossa tese, mas sim as avaliações de fundo mais qualitativo e comparativo. Assim, optamos por avaliar os *concgrams* compostos de 3 palavras, pois não restringe tanto a investigação (no caso dos *concgrams* de 5 palavras, que são menos frequentes) e também não a torna inviável (no caso dos *concgrams* de 2 palavras, que são bem mais frequentes). Nem todos os *concgrams* de 3 palavras foram analisados, estabelecemos como ponto de corte analisar até 8 *concgrams* para cada palavra-nodo. Lembramos também que algumas associações não foram analisadas pois constituíam-se apenas como coocorrências ocasionais de palavras.

Ressaltamos ainda que usamos uma lista de palavras a serem excluídas da busca, com isso evitando que o sistema apresentasse como *concgram* associações com palavras que são muito frequentes na Língua Portuguesa, como artigos e preposições. A lista de palavras excluídas é apresentada no Anexo II.

As análises são feitas através de uma observação dos contextos em que os *concgrams* ocorrem. Decidimos, na maioria dos casos, não extrapolar o limite do contexto (ou linha de concordância) apresentado pelo software. Assim, concentramos a observação na própria associação. Após cada grupo de contextos, fazemos nossas apreciações a respeito das associações. Nosso foco principal é demonstrar os pontos em que elas são flexíveis no que diz respeito à ordem sintática e às palavras em seu entorno, que se constituem de colocações, coligações, preferências semânticas (que são conjuntos semânticos de colocados) e até mesmo prosódias semânticas.

3.2. Resultados da investigação da Etapa 1

3.2.1. Crianças na seção Introdução

A palavra da qual pesquisamos os *congrams* na seção *Introdução* foi *crianças*. A lista dos primeiros *congrams* de 3 palavras para essa palavra-nodo pode ser vista no quadro a seguir, que também mostra a frequência de ocorrência das associações e se elas foram consideradas para análise (A) ou não (NA). A lista completa encontra-se no Anexo III.

Quadro 23 – Congrams de 3 palavras para a palavra-nodo crianças

Palavra-nodo			Frequência	Analisados
crianças	al	Et	11	NA
crianças	anos	idade	10	A
crianças	arterial	pressão	9	A
crianças	não	uso	9	A
crianças	anos	entre	8	A
crianças	anos	menores	8	A
crianças	abaixo	idade	5	A
crianças	anos	dois	5	A
crianças	aprovados	uso	5	NA
crianças	contra	uso	5	NA
crianças	estudo	Et	5	NA
crianças	idade	meses	5	A

3.2.1.1. crianças/anos/idade

1 e em pessoas de 50 anos de idade ou mais (4). Em crianças, tem sido relatado aumento crescente de casos,
 2 463.801 adolescentes de 13 e 14 anos e 257.800 crianças de 6 e 7 anos de idade, de 56 países. A
 3 são tumores geralmente diagnosticados em crianças abaixo de 2 anos de idade e se localizam, na
 4 71% (média 34%) de identificação de rotavírus em crianças menores de 3 anos de idade com diarreia aguda.
 5 com a molécula protéica, particularmente em crianças abaixo dos dois anos de idade (11,12). Além
 6 e social e está associada a 29% das mortes entre crianças de zero a quatro anos de idade (1). Pelletier e
 7 tratamento desses pacientes. <Vol180Iss4_291> Em crianças saudáveis menores de 2 anos de idade, os
 8 da mucosa do intestino delgado de 85 crianças, de quatro meses e dez anos de idade, foi
 9 digestório na infância, dentre elas a RI. Em crianças, na faixa etária de 0 a 2 anos de idade, muitas
 10 exposição aguda aos derivados imidazolinicos em crianças com idade inferior a 15 anos, atendidas pelo

Figura 8 – Linhas de concordância do congram crianças/anos/idade

A parte variável que encontramos nesses exemplos são os qualificadores⁵² para *crianças*, que delimitam a faixa etária. Há uma predominância de uma construção seguindo a ordem *crianças + classificador + anos de idade*, porém há também a possibilidade da ordem ser *crianças + idade + anos*. Observamos ainda, na linha 7, a inclusão de *saudáveis* entre os *congrams*. Nos exemplos da linha 2 a 10, percebemos que a descrição da idade das crianças

⁵² Qualificadores é um termo que tomamos emprestado da Linguística Funcional.

é feita usando diferentes palavras e expressões, mas todas elas tem a função de definir a idade dos pacientes. Assim, pode-se dizer que esse *concgram* tem uma associação semântica com itens numéricos, representativos da idade dos sujeitos em questão.

3.2.1.2. crianças/arterial/pressão

```

1 correlação entre pressão arterial elevada em crianças e adolescentes e hipertensão na idade adulta
2 mais elevados de pressão arterial, sendo que crianças e adolescentes da cor negra apresentam níveis
3 pediátrica deverá medir a pressão arterial das crianças maiores de 3 anos no mínimo uma vez ao ano e
4 sobre o controle da pressão arterial em crianças (4), que apresentava uma proposta de padronização
5 com a avaliação da pressão arterial em crianças nas últimas décadas. Em 1977, nos Estados Unidos,
6 curvas de distribuição da pressão arterial em crianças normais, organizadas em gráficos de percentis de
7 com níveis elevados da pressão arterial em crianças de dois estabelecimentos de ensino (público e
8 sobre o controle da pressão arterial em crianças (5). Em 1993, Rosner et al. publicaram uma
9 e hipertensão na idade adulta (3,4). Crianças com pressão arterial acima do percentil 90 têm um

```

Figura 9 – Linhas de concordância do *concgram* crianças/arterial/pressão

As variações observadas nessas linhas são o uso da preposição *em* (linhas 1 e de 4 a 8) e *de + as* (linha 3). Poderíamos também dizer diferente, que estas seriam coligações do *concgram* em questão.

Na linha 1, há inclusão de um classificador para *pressão arterial: elevada*. As linhas 2 e 9 foram excluídas da análise, pois não tem o mesmo sentido ideacional das outras associações.

As regularidades observadas em torno do *concgram* crianças/arterial/pressão são a presença de verbos (*medir, controlar, avaliar, distribuir*) ou substantivos que derivam desses verbos. Também parece haver uma associação semântica com um conceito de pressão elevada, como percebemos nos exemplos 1 e 7.

3.2.1.3. crianças/não/uso

```

1 não são aprovados ou padronizados para uso em crianças (1). No nosso meio, suspeita-se que a taxa de uso
2 para uso como um todo, não aprovados para uso em crianças, contra-indicados para uso em crianças,
3 taxa de uso de medicamentos não apropriados para crianças, em pacientes pediátricos, também seja elevada,
4 do uso de medicamentos não apropriados para crianças em prescrições de uma unidade de tratamento
5 uso de enzimas pancreáticas (8-10). No entanto, crianças com FC geralmente não atendem a esses requisitos
6 medicações e do risco potencial de seu uso em crianças, não há consenso quanto a posologia ou duração do

```

Figura 10 – Linhas de concordância do *concgram* crianças/não/uso

Podemos ver que cada item do *concgram* pode ocupar posições diferentes nas frases e, mesmo assim, estão cumprindo a mesma função (linhas 1, 2, 3 e 4), qual seja, apontar que o uso do medicamento não é indicado para crianças. Há uma variação de posição dos elementos

dos *congrams* em: *não aprovados para uso* (linhas 1 e 2) e *uso de medicamentos não apropriados*.

São colocados desses *congrams* palavras como: *aprovados*, *apropriados*, *padronizados*, *medicamentos* e *medicações*. Percebemos que esses colocados pertencem a dois campos semânticos semelhantes.

3.2.1.4. crianças/anos/entre

1 de 1965 e 1980, constata-se que a obesidade nas crianças de 6 a 11 anos aumentou em 67% entre os meninos e
 2 São Paulo, prevalência de 2,5% de obesidade em crianças menores de 10 anos, entre as classes econômicas
 3 entre 1 e 2 milhões de óbitos ocorrem em crianças menores que cinco anos, em decorrência de
 4 Nos Estados Unidos, entre 1963 e 1980, em crianças de 6 a 11 anos, a prevalência de asma referida
 5 Unidos da América, onde a prevalência entre crianças menores de cinco anos era de 10,3% em 1997 (3),
 6 e social e está associada a 29% das mortes entre crianças de zero a quatro anos de idade (1). Pelletier e
 7 a tendência de aumento de peso entre crianças num período de 20 anos, já valorizando a
 8 conjugadas contra Hib foram introduzidas para crianças entre 18 meses e cinco anos. A partir de 1990, o

Figura 11 – Linhas de concordância do congram crianças/anos/entre

Nesse caso, encontramos exemplos em que os *congrams* estão associados. No entanto, *entre*, por exemplo, não está semanticamente associado com *crianças* e *anos*, como nos parece o caso das linhas 1, 2, 3, 4. As linhas 5, 6 e 7 expressam um mesmo sentido ideacional, e a frase 8 expressa outro sentido, mais semelhante ao do primeiro *congram*.

Neste caso, nas linhas de 5 a 7, as palavras que se posicionam entre *entre crianças* e *anos* são também pertencentes ao um mesmo campo semântico de delimitação de faixa etária.

3.2.1.5. crianças/anos/menores

1 71% (média 34%) de identificação de rotavírus em crianças menores de 3 anos de idade com diarreia aguda.
 2 São Paulo, prevalência de 2,5% de obesidade em crianças menores de 10 anos, entre as classes econômicas
 3 no ano de 1996 revelou grande adesão das mães de crianças menores de dois anos à prática da amamentação.
 4 a imunogenicidade da vacina pneumocócica em crianças menores de dois anos, por permitir uma resposta
 5 num estudo realizado nas Filipinas, com 9.942 crianças menores de dois anos, avaliaram o efeito da
 6 Unidos da América, onde a prevalência entre crianças menores de cinco anos era de 10,3% em 1997 (3),
 7 entre 1 e 2 milhões de óbitos ocorrem em crianças menores que cinco anos, em decorrência de
 8 tratamento desses pacientes. <Vol180Iss4_291> Em crianças saudáveis menores de 2 anos de idade, os

Figura 12 – Linhas de concordância do congram crianças/anos/menores

Nesse exemplo, as variáveis são bem evidentes: *anos* pode estar acompanhado ou não do complemento *de idade* e um qualificador foi incluído no contexto da linha 8. Outra variação é o uso de *que*. Podemos também dizer que um colocado para essas associações seria a palavra *prevalência*.

3.2.1.6. crianças/abaixo/idade

1 por um expressivo número de hospitalizações em **crianças abaixo** de 1 ano de **idade** (2). O vírus
 2 são tumores geralmente diagnosticados em **crianças abaixo** de 2 anos de **idade** e se localizam, na
 3 comuns de infecção das vias aéreas inferiores em **crianças abaixo** de um ano de **idade** e determina um
 4 com a molécula protéica, particularmente em **crianças abaixo** dos dois anos de **idade** (11,12). Além
 5 de **idade**, relataram a presença da RI em 50% das **crianças abaixo** dos 3 meses, em 67% na faixa etária de 4 a

Figura 13 – Linhas de concordância do *concgram* crianças/abaixo/idade

O exemplo na linha 5 deve ser excluído, pois *idade* não está vinculado as duas outras palavras. Não há variação na ordem do *concgram*, mas sim nas palavras associadas a ele. Nos casos de 1 a 4, as variáveis foram os números e o uso da aglutinação de preposição + artigo. Uma observação sobre o tema da convencionalidade é que, no caso do *concgram* anterior a esse, a combinação *crianças/menores* não era seguida da expressão *anos de idade* enquanto para *crianças/abaixo* todas as coocorrências de anos foram seguidas por *de idade*.

3.2.1.7. crianças/anos/dois

1 a imunogenicidade da vacina pneumocócica em **crianças** menores de **dois anos**, por permitir uma resposta
 2 num estudo realizado nas Filipinas, com 9.942 **crianças** menores de **dois anos**, avaliaram o efeito da
 3 no ano de 1996 revelou grande adesão das mães de **crianças** menores de **dois anos** à prática da amamentação.
 4 com a molécula protéica, particularmente em **crianças** abaixo dos **dois anos** de idade (11,12). Além
 5 em pessoas de todas as idades ao redor do mundo. **Crianças** com menos de **dois anos**, indivíduos com mais de 65

Figura 14 – Linhas de concordância do *concgram* crianças/anos/dois

Nesse exemplo, também há uma pequena variação nos qualificadores para *crianças*, todos eles indicando um limite de idade dos sujeitos estudados ou referidos no estudo. Esse é, portanto, um *concgram* semelhante aos que analisamos anteriormente, com a diferença de que, nesse caso, em nosso cópulo, talvez a idade de até dois anos seja mais frequentemente referida como uma idade – ou um parâmetro – importante no desenvolvimento da criança ou uma data limítrofe ou que sirva como parâmetro para determinados estudos.

3.2.1.8. crianças/idade/meses

1 de **idade**, relataram a presença da RI em 50% das **crianças** abaixo dos 3 **meses**, em 67% na faixa etária de 4 a
 2 de uma doença das vias aéreas inferiores em **crianças** com **idade** inferior a 12 **meses**, precedida por um
 3 aos quatro **meses** de **idade**, comparadas com **crianças** amamentadas exclusivamente ao seio até o 6º mês
 4 da mucosa do intestino delgado de 85 **crianças**, de quatro **meses** e dez anos de **idade**, foi
 5 analisando dados de 948 entrevistas com pais de **crianças** menores de 13 **meses** de **idade**, relataram a

Figura 15 – Linhas de concordância do *concgram* crianças/idade/meses

Novamente, esse *concgram* está associado para apresentar dados da faixa etária das crianças referidas no estudo (exceto itens 1 e 3). Nas linhas 2, 4 e 5, para descrever o limite de idade, temos três possibilidades, sendo que, nas linhas 4 e 5, temos a ordem *crianças meses idade* e, na linha 2, *crianças idade meses*.

3.2.2. *Estudo na seção Métodos*

A palavra da qual pesquisamos os *concgrams* na seção *Métodos* foi *estudo*. A lista dos primeiros *concgrams* de 3 palavras para essa palavra-nodo pode ser vista no quadro a seguir, que também mostra a frequência de ocorrência das associações e se elas foram consideradas para análise (A) ou não (NA). A lista completa encontra-se no Anexo III.

Quadro 24 – Concgrams de 3 palavras para a palavra-nodo estudo

Palavra-nodo			Frequência	Analizados
estudo	foi	realizado	38	A
estudo	foi	pelo	35	A
estudo	aprovado	foi	32	A
estudo	foi	Vollss	32	NA
estudo	aprovado	pelo	28	A
estudo	realizado	Vollss	28	NA
estudo	Comitê	pelo	27	NA
estudo	Comitê	foi	26	NA
estudo	aprovado	Comitê	23	NA
estudo	transversal	Vollss	19	NA
estudo	crianças	Foram	16	A
estudo	foi	Pesquisa	16	NA
estudo	foram	pacientes	16	A
estudo	Este	Vollss	15	NA
estudo	aprovado	Pesquisa	14	NA
estudo	Este	foi	14	A
estudo	Comitê	Pesquisa	13	NA
estudo	foi	transversal	13	A
estudo	se	realizou	13	A

3.2.2.1. **estudo/foi/realizado**

1 médico. <Vol79Iss2_159> Foi realizado um estudo retrospectivo de todos os casos de enterocolite
2 a 0,05. <Vol79Iss2_123> Foi realizado um estudo transversal, de caráter observacional, envolvendo
3 materna e gemelaridade. Foi realizado um estudo descritivo, através da análise deste banco de
4 ou de recombinação MLL-AF4. Foi realizado um estudo de coorte mista com informação oriunda de
5 em 5% (12). <Vol79Iss3_245> Foi realizado um estudo transversal controlado, cujo fator em estudo foi
6 Fluminense. <Vol79Iss2_165> Foi realizado um estudo transversal durante um período de dois anos
7 Alegre (HCPA). <Vol80Iss5_355> Foi realizado um estudo transversal nas UTIPs e UTINs do Hospital São
8 estudadas. <Vol80Iss4_305> Foi realizado um estudo epidemiológico tipo transversal por meio da
9 ou $p < 0,05$. <Vol79Iss4_297> Foi realizado um estudo caso-controle, incluindo todos os recém-nascidos
10 dois grupos). <Vol80Iss4_315> Foi realizado um estudo de coorte durante o pico sazonal do VRS em Porto
11 (SPSS Inc.). <Vol80Iss1_23> Foi realizado um estudo caso-controle pareado com crianças e adolescentes
12 de dados. <Vol80Iss2_147> Foi realizado um estudo transversal com 91 pacientes, de ambos os sexos,
13 o mesmo ventilador mecânico. Foi realizado um estudo de experiências controladas através de simulações
14 de campo (12). <Vol79Iss5_397> Foi realizado um estudo transversal, observacional, baseado na avaliação
15 foi desenvolvida pelos mesmos pediatras em um estudo piloto, realizado seis meses antes. Cálculo da
16 foi definido com base nos resultados de um estudo prévio realizado em nosso laboratório (13), onde
17 foi de 5%. <Vol80Iss2_113> Trata-se de um estudo prospectivo, por ensaio clínico, realizado no
18 foi de 5%. <Vol80Iss6_511> Este é um estudo do tipo coorte retrospectivo, realizado no
19 do tamanho da amostra foi feito a partir de um estudo piloto realizado com 20 pacientes. Para
20 <Vol79Iss1_13> O estudo foi realizado na maternidade do Hospital Regional
21 de confiança de 95%. <Vol80Iss2_123> Este estudo foi realizado na Maternidade do Hospital
22 o software SAS, versão 8.2. <Vol80Iss1_17> O estudo foi realizado no ano de 2000, envolvendo
23 Prof. Hosannah de Oliveira (CPPHO - UFBA). O estudo foi realizado em duas salas: em uma delas ficava
24 de excesso de peso pelos responsáveis diretos. O estudo foi realizado em escolas da rede de ensino
25 artesanalmente através do sistema selo d'água. O estudo foi realizado no laboratório de manutenção da
26 de Fisher e Kruskal-Wallis. <Vol80Iss3_229> O estudo foi realizado na cidade de Pontal, no estado de
27 livre e esclarecido). <Vol80Iss4_291> O estudo foi realizado no Hospital de Pediatria Professora
28 e o IC 95% do RGEF. <Vol80Iss4_296> Este estudo foi realizado na Unidade Neonatal e de Neurologia
29 Paulista de Medicina. <Vol80Iss3_189> Este estudo foi realizado no laboratório experimental da
30 tabagistas ou estilistas durante a gravidez. O estudo foi realizado em duas etapas: 1ª etapa:
31 e testosterona foram inferiores a 5%. Este estudo foi realizado de acordo com a revisão atual da
32 período de realização e população estudada O estudo foi realizado no serviço de pneumologia e alergia
33 imunoglobulinas no colostro. <Vol80Iss2_129> O estudo foi realizado em Brasília (DF), entre os anos de
34 reformuladas na linguagem sugerida por eles. O estudo piloto foi realizado com 20 adolescentes. O
35 e construção das figuras. <Vol80Iss5_407> Um estudo prospectivo foi realizado a fim de avaliar
36 encontrados representassem a população em estudo, o cálculo da amostra foi realizado com base num
37 abordagem das famílias e na coleta de dados. Um estudo piloto, com 10 pacientes, foi realizado em
38 Federal de Minas Gerais. <Vol79Iss3_239> Este estudo clínico prospectivo e randomizado foi realizado

Figura 16 – Linhas de concordância do *concgram* estudo/foi/realizado

Nas linhas de concordância acima, se observa que o *concgram* se associa à esquerda, predominantemente, a um artigo definido *o* ou indefinido *um*, e, em alguns casos, também ao demonstrativo *este*. Podemos perceber que, quando o artigo indefinido é usado, são colocadas palavras que qualificam os estudos: *transversal*, *descritivo*, *epidemiológico*. Há uma variação desta estrutura na linha 14, em que o artigo indefinido é excluído. As linhas 35, 37 e 38 são variações textuais para uma mesma função desempenhada pelo *concgram*, a de descrever o estudo. Elas apresentam o verbo após *estudo*.

Nas linhas de 15 a 16 e 36, a coocorrência dos itens é ocasional, uma vez que *foi* não se refere à *estudo*.

Quando é usado o artigo definido, percebemos que os colocados apresentam um conteúdo semântico de local ou data de realização do *estudo*. Podemos, então, concluir que a ordem dos itens que compõem o *concgram* varia e associa-se a diferentes colocados ou coligados conforme o sentido ideacional da frase. Na linha 34, observamos a inclusão de um classificador para estudo (*piloto*)

3.2.2.2. estudo/foi/pelo

1 mínima aceito foi p < 0,05. O protocolo do estudo foi aprovado pelo Comitê de Ética da nossa
 2 a 22 anos e tratamento concluído. Este foi um estudo retrospectivo, aprovado pelo Comitê de Ética do
 3 Ltda.), com o apoio da Intermed. O protocolo do estudo foi aprovado pelo Comitê de Ética em Pesquisa do
 4 (Child-Pugh B) ou grave (Child-Pugh C). O estudo foi aprovado pelo Comitê de Ética do hospital,
 5 1) foram considerados positivos (10,12). O estudo foi aprovado pelo Comitê de Ética, e os pacientes
 6 as variáveis para a análise multivariada (17). O estudo foi aprovado pelo Comitê de Ética em Pesquisa do
 7 a 37 semanas, segundo o método Capurro (15). O estudo foi aprovado pelo Comitê de Ética da Santa Casa
 8 responsáveis pelos pacientes e controles, e o estudo foi aprovado pelo Comitê de Ética da Universidade
 9 (odds ratio) <Vol79Iss6_513> O presente estudo foi aprovado pelo Comitê de Ética em Pesquisa da
 10 em 103 RNMBP, considerando $\alpha = 5\%$ e $\beta = 20\%$. O estudo foi aprovado pelo Comitê de Ética em Pesquisa da
 11 segundo a classificação de Tanner (10). O estudo foi aprovado pelo Comitê de Ética em Pesquisa da
 12 e seguia-se a conduta de rotina do serviço. O estudo foi aprovado pelo Comitê de Ética em Pesquisa do
 13 nível de rejeição da hipótese de nulidade. Este estudo foi aprovado pelo Comitê de Ética em Pesquisa da
 14 p (erro alfa) menor ou igual a 0,05. O presente estudo foi aprovado pelo Comitê de Ética Médica da
 15 das amostras de LHC e de LHP. A elaboração do estudo foi aprovada pelo Comitê de Ética em Pesquisa do
 16 0,35 com alfa 0,05 e beta 0,20. O protocolo do estudo foi aprovado pelo Comitê de Ética em Pesquisa do
 17 estudantes de classe média e média alta. O estudo foi aprovado pelo comitê de ética da UCB e
 18 Williams entre hepatopatas e não-hepatopatas. O estudo foi aprovado pelo Comitê de Ética em Pesquisa
 19 O responsável legal por cada criança incluída no estudo foi informado pelo pesquisador quanto aos
 20 da Universidade de São Paulo. O protocolo do estudo foi previamente aprovado pelo Comitê de Ética de
 21 HIV poderia tornar não ético o estudo (13,17). O estudo foi submetido e aprovado pelo Comitê de Ética em
 22 em região de alto IQVU (0,52, classe II). O estudo foi iniciado após a aprovação do trabalho pelo
 23 berçários das maternidades do mesmo município. O estudo foi iniciado após a aprovação do protocolo pelo
 24 dosagens laboratoriais. Na primeira fase do estudo, foi preenchido pelo pesquisador um questionário
 25 pacientes abordadas se recusou a participar do estudo, que foi aprovado pelo Comitê de Ética e Pesquisa
 26 semana de vida e óbito durante o período de estudo. Esta investigação foi aprovada pelo Comitê de
 27 há pelo menos 3 semanas, foram admitidas no estudo. A amostra foi do tipo não-probabilística por
 28 de infecção pelo HIV poderia tornar não ético o estudo (13,17). O estudo foi submetido e aprovado pelo
 29 descritas pelo cálculo da proporção (10). O estudo foi aprovado pela Comissão de Ética Médica da
 30 conforme prescrito pelo médico responsável. O estudo foi aprovado pelo comitê de ética médica de nosso
 31 livre e esclarecido pelo diretor da escola. O estudo foi aprovado pelo Comitê de Ética em Pesquisa do
 32 para 16 kg, aferida pelo Inmetro no início do estudo. O comprimento foi aferido com régua

Figura 17 – Linhas de concordância do concgram estudo/foi/pelo

Nessas linhas, notamos uma sequência de ocorrências com uma sequência fixa (de 4 a 18). Queremos chamar a atenção que, mesmo nesse grupo, encontramos variações de denominação do Comitê de Ética: *Comitê de Ética em Pesquisa da Universidade X*, *Comitê de Ética Médica* etc.

Encontramos uma série de variações nos itens em torno desse *concgram*. Primeiramente, identificamos três “organizações” diferentes para a posição dos itens que compõem o *concgram*:

- 1) estudo foi [verbo] pelo (linhas 1, 3-18, 20, 21, 30, 31)
- 2) foi um estudo [verbo] pelo (linha 2)
- 3) estudo, que foi [verbo] (linha 25)
- 4) estudo foi [verbo] após (linhas 22 e 23)

A estrutura mais frequente é a primeira e como variações encontramos especificações de participantes: *protocolo do estudo*, *elaboração do estudo*, *o estudo*, *o presente estudo*. Também há variações no processo: *foi previamente aprovado*, *foi submetido e aprovado*, *foi iniciado*. Nota-se como a associação com o verbo *aprovar* é frequente.

Observamos que são colocados para esse *concgram* palavras como: *aprovado*, *aprovação*, *comitê de ética em pesquisa*, *protocolo*.

3.2.2.3. estudo/aprovado/foi

1 mínima aceito foi p < 0,05. O protocolo do estudo foi aprovado pelo Comitê de Ética da nossa
 2 a 22 anos e tratamento concluído. Este foi um estudo retrospectivo, aprovado pelo Comitê de Ética do
 3 segundo a classificação de Tanner (10). O estudo foi aprovado pelo Comitê de Ética em Pesquisa da
 4 a 37 semanas, segundo o método Capurro (15). O estudo foi aprovado pelo Comitê de Ética da Santa Casa
 5 em 103 RNMBP, considerando $\alpha = 5\%$ e $\beta = 20\%$. O estudo foi aprovado pelo Comitê de Ética em Pesquisa da
 6 responsáveis pelos pacientes e controles, e o estudo foi aprovado pelo Comitê de Ética da Universidade
 7 ao final do primeiro mês de vida (n= 219). O estudo foi aprovado pela Comissão Científica e pela
 8 10.0, e Sigma Plot v. 2.1 (19). O presente estudo foi aprovado pela Comissão de Ética em Pesquisa
 9 admissão, tendo um bom poder preditivo (13). O estudo foi aprovado pela Comissão de Ética e Pesquisa do
 10 um nível de significância de 0,05. Ética Este estudo foi aprovado pela comissão de ética em pesquisa
 11 livre e esclarecido pelo diretor da escola. O estudo foi aprovado pelo Comitê de Ética em Pesquisa do
 12 0,35 com alfa 0,05 e beta 0,20. O protocolo do estudo foi aprovado pelo Comitê de Ética em Pesquisa do
 13 (odds ratio) <Vol179Iss6_513> O presente estudo foi aprovado pelo Comitê de Ética em Pesquisa da
 14 Williams entre hepatopatas e não-hepatopatas. O estudo foi aprovado pelo Comitê de Ética em Pesquisa
 15 estudantes de classe média e média alta. O estudo foi aprovado pelo comitê de ética da UCB e
 16 nível de rejeição da hipótese de nulidade. Este estudo foi aprovado pelo Comitê de Ética em Pesquisa da
 17 (Child-Pugh B) ou grave (Child-Pugh C). O estudo foi aprovado pelo Comitê de Ética do hospital,
 18 e seguia-se a conduta de rotina do serviço. O estudo foi aprovado pelo Comitê de Ética em Pesquisa do
 19 p (erro alfa) menor ou igual a 0,05. O presente estudo foi aprovado pelo Comitê de Ética Médica da
 20 1) foram considerados positivos (13,14). O estudo foi aprovado pelos Comitês de Ética, e obteve-se
 21 1) foram considerados positivos (10,12). O estudo foi aprovado pelo Comitê de Ética, e os pacientes
 22 as variáveis para a análise multivariada (17). O estudo foi aprovado pelo Comitê de Ética em Pesquisa do
 23 Ltda.), com o apoio da Intermed. O protocolo do estudo foi aprovado pelo Comitê de Ética em Pesquisa do
 24 descritas pelo cálculo da proporção (10). O estudo foi aprovado pela Comissão de Ética Médica da
 25 conforme prescrito pelo médico responsável. O estudo foi aprovado pelo comitê de ética médica de nosso
 26 da Universidade de São Paulo. O protocolo do estudo foi previamente aprovado pelo Comitê de Ética de
 27 HIV poderia tornar não ético o estudo (13,17). O estudo foi submetido e aprovado pelo Comitê de Ética em
 28 Santiago de Jinotepe na Nicarágua. O presente estudo foi submetido e aprovado pela Comissão de
 29 pacientes abordadas se recusou a participar do estudo, que foi aprovado pelo Comitê de Ética e Pesquisa
 30 de infecção pelo HIV poderia tornar não ético o estudo (13,17). O estudo foi submetido e aprovado pelo
 31 foram estudados 60 filmes. Aspectos éticos O estudo inicial sobre pneumonia foi avaliado e aprovado

Figura 18 – Linhas de concordância do concgram estudo/aprovado/foi

Esses exemplos são praticamente os mesmos do item anterior, no entanto os incluímos porque demonstram como as palavras-nodo são apenas um recurso metodológico, não sendo portadoras de um valor privilegiado nas sentenças. Nessas linhas, podemos ver palavras que foram incluídas entre os *concgrams*, como o qualificador *retrospectivo* (linha 2) e *inicial sobre pneumonia* (linha 31), para estudo, e *previamente* para *aprovado*. Também os verbos *submeter* e *avaliar*, nas linhas 27, 28 e 31, foram incluídos.

A linha 31 mostra uma variação para a estrutura mais frequente de se apresentar esse tipo de informação, desempenhando a mesma função, mas trazendo mais dados sobre o estudo.

Os próximos *concgrams* são *estudo/aprovado/pelo*, *estudo/comitê/pelo*, *estudo/comitê/foi*, *estudo/aprovado/comitê*, *estudo/foi/pesquisa*, *estudo/aprovado/pesquisa*, *estudo/comitê/pesquisa*. Como os itens de destaque são praticamente os mesmos, não os incluímos em uma análise mais detalhada.⁵³

3.2.2.4. estudo/crianças/foram

⁵³ Os *concgrams* estudo/foi/Volls, estudo/realizado/Volls, estudo/transversl/Volls, estudo/este/Volls e outros com a palavra *Volls* foram excluídos, uma vez que *Volls* é uma palavra do córpus que indica a descrição de volume em que o artigo foi publicado. É um metadado do córpus e, portanto, não deve entrar na análise.

1 Após a seleção, 192 crianças foram incluídas no estudo. Ambas as creches atendem uma população de baixa
 2 de crianças para o grupo controle, o poder do estudo não ficou comprometido, porque foram recrutadas
 3 pelas crianças e adolescentes incluídos no estudo. Os pacientes com TVP foram avaliados sob ponto
 4 avaliação clínica das crianças submetidas ao estudo. As amostras foram processadas no setor de
 5 Foram avaliadas, prospectivamente, através de um estudo observacional, dez crianças consecutivamente
 6 os pacientes foram incluídos prospectivamente no estudo. As crianças foram submetidas a um protocolo
 7 para cada paciente. Foram incluídas neste estudo todas as crianças de 0 a 12 meses de idade, com
 8 portadores de DAC prematura. Foram incluídas no estudo as crianças e adolescentes com idades entre 2
 9 Convivência Infantil (CCI). Foram excluídas do estudo 23 crianças que apresentaram doença grave nas
 10 conclusão dos exames). Assim, foram incluídas no estudo 45 crianças com SD. Realizou-se um corte
 11 não era o genitor (a), foram incluídas no estudo apenas as crianças cujos entrevistados eram
 12 com duração mínima de 1 mês. Foram incluídas no estudo apenas as crianças com exame clínico normal e
 13 faz parte da equipe de amamentação do GAMA. No estudo foram incluídas crianças nascidas no Hospital
 14 de 6 a 10 anos (15,16) Para a segunda fase do estudo, foram selecionadas todas as crianças com
 15 (p < 0,05). <Vol180Iss1_29> Este é um estudo transversal no qual foram avaliadas crianças de

Figura 19 – Linhas de concordância do concgram estudo/aprovado/foi

Nessas linhas, é possível identificar que a maioria dos casos trata de introduzir a apresentação dos critérios de inclusão e exclusão dos sujeitos do estudo. Para isso, diferentes estruturas textuais foram empregadas:

- 1) x crianças foram [verbo] no estudo (1)
- 2) foram [verbo] no estudo x crianças (7, 8, 9, 10, 11, 12)
- 3) no estudo foram [verbo] crianças (13)
- 4) foram [verbo], através de um estudo, x crianças (5)
- 5) para o estudo, foram [verbo] x crianças (14)
- 6) este é um estudo no qual foram [verbo] crianças (15)

A estrutura mais frequente é a segunda que, em todas as ocorrências, usa o mesmo verbo *incluir*, com exceção da linha 9, que emprega *excluir*. Podemos também perceber que as três primeiras estruturas diferem apenas na posição dos elementos na frase, e, na linha 13, o qualificador de crianças não é um numeral, mas *nascidas no Hospital...*

Como colocados podemos destacar o verbo *incluir*, mas outros como *submeter*, *excluir*, *selecionar*, *avaliar*.

3.2.2.5. estudo/foram/pacientes

1 1) foram considerados positivos (10,12). O estudo foi aprovado pelo Comitê de Ética, e os pacientes
 2 em dezembro de 2000, foram incluídas neste estudo 1.261 pacientes que haviam realizado exame
 3 de inclusão e exclusão. Foram incluídos no estudo 63 pacientes de quatro a catorze anos, no primeiro
 4 de p < 0,05. <Vol179Iss2_149> Foram incluídos no estudo 63 pacientes que tiveram o diagnóstico de LLA com
 5 para estudo histológico. Foram incluídos no estudo os pacientes cujos fragmentos de mucosa de
 6 (p < 0,05). <Vol179Iss6_543> Foram incluídos no estudo todos os pacientes que, admitidos na Unidade de
 7 a 14 anos) o consentimento para participação no estudo. <Vol180Iss5_387> Foram estudados nove pacientes
 8 os pacientes foram incluídos prospectivamente no estudo. As crianças foram submetidas a um protocolo
 9 pais e pacientes foram informados da natureza do estudo e consentiram na sua participação. O escore
 10 no mínimo, 23 pacientes (14). Foram excluídos do estudo os pacientes: - portadores de: malformação
 11 As demais pacientes foram excluídas do estudo, porque já haviam realizado exame sorológico fora
 12 anos anteriores. Os pacientes foram excluídos do estudo caso se encontrassem em algumas das situações
 13 (SUS). Entre os 1.050 pacientes elegíveis para o estudo, foram selecionados, de forma aleatória simples,
 14 pelas crianças e adolescentes incluídos no estudo. Os pacientes com TVP foram avaliados sob ponto
 15 uma recusa, por parte dos pais, em participar do estudo, e dados de cinco pacientes não foram obtidos. A

Figura 20 – Linhas de concordância do concgram estudo/foram/pacientes

Essas linhas não diferem muito das do exemplo anterior, porém aqui se utiliza *pacientes* ao invés de *crianças*. Como nas linhas do *concgram* anterior, os itens desse *concgram* também se associam em frases que apresentam critérios de inclusão e exclusão no estudo (excluindo-se as linhas 1, 7, 14, 15).

Algumas variações de posição dos itens pode ser observada comparando-se as linhas 2 e 10, por exemplo. Na linha 2 (assim como em 3, 4, 5 e 6), temos *Foram incluídos no estudo pacientes* e, na linha 10, *Os pacientes foram excluídos do estudo*. Além desse exemplo, temos a linha 13, *Entre os X pacientes elegíveis para o estudo, foram selecionados...*, que mostra uma formação frasal ainda mais diferente das duas anteriormente citadas. Os colocados para os *concgrams* concentram-se em dois: *excluídos* e *incluídos*.

3.2.2.6. estudo/este/foi

```

1      (versão 8.0). <Vol79Iss1_49> Este foi um estudo caso-controle prospectivo, em 100 crianças com
2      a 22 anos e tratamento concluído. Este foi um estudo retrospectivo, aprovado pelo Comitê de Ética do
3      (pequenos para a idade gestacional). Este foi um estudo longitudinal prospectivo que incluiu todos os
4      específico para este estudo. A população do estudo foi composta por todos os pacientes com
5      um nível de significância de 0,05. Ética Este estudo foi aprovado pela comissão de ética em pesquisa
6      nível de rejeição da hipótese de nulidade. Este estudo foi aprovado pelo Comitê de Ética em Pesquisa da
7      de confiança de 95%. <Vol80Iss2_123> Este estudo foi realizado na Maternidade do Hospital
8      Paulista de Medicina. <Vol80Iss3_189> Este estudo foi realizado no laboratório experimental da
9      de 5% (alfa = 0,05). <Vol80Iss3_197> Este estudo foi conduzido no Serviço de Gastroenterologia
10     e o IC 95% do RGEF. <Vol80Iss4_296> Este estudo foi realizado na Unidade Neonatal e de Neurologia
11     e testosterona foram inferiores a 5%. Este estudo foi realizado de acordo com a revisão atual da
12     um banco de dados específico para este estudo. A população do estudo foi composta por todos os
13     Federal de Minas Gerais. <Vol79Iss3_239> Este estudo clínico prospectivo e randomizado foi realizado
14     foi de 5%. <Vol80Iss6_511> Este é um estudo do tipo coorte retrospectivo, realizado no

```

Figura 21 – Linhas de concordância do concgram estudo/este/foi

O *concgram estudo/este/foi* aparece em frases com três significados ideacionais diferentes. O primeiro (linhas 1, 2 e 3) apresenta o tipo de estudo e, por isso, qualificadores como *caso-controle prospectivo*, *retrospectivo*, *longitudinal* são colocados dessa associação. O segundo sentido (linhas 5, 6, 11) é o de informar que o estudo foi realizado de acordo com algum parâmetro ou tipo de controle, já analisado nos exemplos anteriores, com colocados variáveis como *comissão* e *comitê*, *aprovado*. O terceiro informa o local de realização do estudo (linhas 7, 8, 9, 10, 13). As linhas 7, 8, 9 e 10 seguem uma mesma estrutura, enquanto a frase 13 classifica o tipo de estudo e descreve onde ele foi realizado: *Este estudo clínico prospectivo e randomizado foi realizado na unidade de terapia intensiva (UTI) neonatal*. Na linha 13, parece haver também uma mescla do primeiro com o terceiro sentidos.

3.2.2.7. estudo/foi/transversal

```

1 Alegre (HCPA). <Vol180Iss5_355> Foi realizado um estudo transversal nas UTIPs e UTINs do Hospital São
2 a 0,05. <Vol179Iss2_123> Foi realizado um estudo transversal, de caráter observacional, envolvendo
3 Fluminense. <Vol179Iss2_165> Foi realizado um estudo transversal durante um período de dois anos
4 em 5% (12). <Vol179Iss3_245> Foi realizado um estudo transversal controlado, cujo fator em estudo foi
5 de dados. <Vol180Iss2_147> Foi realizado um estudo transversal com 91 pacientes, de ambos os sexos,
6 estudadas. <Vol180Iss4_305> Foi realizado um estudo epidemiológico tipo transversal por meio da
7 de campo (12). <Vol179Iss5_397> Foi realizado um estudo transversal, observacional, baseado na avaliação
8 anabolizantes ou anticoncepcionais (11). Foi um estudo transversal em que se avaliaram as variáveis
9 em ordem decrescente. O delineamento do estudo foi de corte transversal, retrospectivo. Os
10 de janeiro a agosto de 2002. O desenho do estudo foi um corte transversal, para a estimativa de
11 o termo de consentimento informado. O desenho de estudo adotado para o trabalho foi do tipo transversal,
12 um estudo transversal controlado, cujo fator em estudo foi colestase crônica, e o desfecho é o nível
13 do Recife-PE. Inicialmente realizou-se um estudo do tipo corte transversal, no qual foi possível

```

Figura 22 – Linhas de concordância do concgram estudo/foi/transversal

Dessas linhas, excluímos da análise apenas a linha 12, em que a associação é ocasional. Nesse exemplo, podemos observar, novamente, como o mesmo *concgram* pode realizar a mesma função em combinações diferentes de texto. Em termos de flexibilidade na formação, podemos observar as diferenças nas maneiras de apresentar o estudo: *foi realizado um estudo transversal* (linhas 1-6), a ausência do artigo indefinido (linha 7), a ausência do verbo *realizado* (linha 8), a introdução de um sujeito (linhas 8, 9 e 10), e a forma impessoal *realizou-se*. Chama a atenção que, nos casos de 1 a 6, não foi utilizada a especificação *corte transversal*, embora seria aceitável uma frase como *Foi realizado um estudo de corte transversal*, e quando se utilizou *desenho* ou *delineamento* há o uso de *corte*. Como colocados para esses *concgrams* citaríamos *realizar*, *corte*, *desenho*, podendo haver outros como *controlado*, *de caráter observacional*, *retrospectivo*, ou seja, palavras que especificam o tipo de estudo.

3.2.2.8. estudo/se/realizou

```

1 etária <Vol180Iss5_391> Realizou-se no Recife um estudo do tipo corte transversal de base populacional,
2 do Recife-PE. Inicialmente realizou-se um estudo do tipo corte transversal, no qual foi possível
3 por faixa etária; posteriormente, realizou-se um estudo comparativo entre o grupo de crianças com
4 de associação. <Vol179Iss1_29> Realizou-se um estudo prospectivo descritivo a partir do ano de 1999,
5 probabilidade. <Vol179Iss6_550> Realizou-se um estudo prospectivo de uma coorte de 153 RN, tratados no
6 de Alagoas. <Vol180Iss1_41> Realizou-se um estudo transversal, observacional, contemporâneo, com
7 e as expectativas de futuro. Realizou-se um estudo-piloto com 10 pacientes (que não participaram da
8 de maio a outubro de 2001, realizou-se um estudo de coorte que incluiu 111 pacientes entre 1 e 12
9 responsáveis. <Vol180Iss5_371> Realizou-se um estudo descritivo, retrospectivo e de corte transversal
10 de Porto Alegre. <Vol179Iss1_75> Realizou-se estudo longitudinal, comparativo, de 127 recém-nascidos
11 segundo cada aspecto avaliado <Vol179Iss1_21> O estudo realizou-se no ambulatório geral do Hospital
12 conclusão dos exames). Assim, foram incluídas no estudo 45 crianças com SD. Realizou-se um corte
13 classes (21). Quando as crianças avaliadas no estudo perinatal completaram três meses, realizou-se uma

```

Figura 23 – Linhas de concordância do concgram estudo/se/realizou

Excluindo-se as linhas 11, 12 e 13, em termos de variações na associação dos itens do *concgram*, as linhas 1 e 10 diferem das demais. Na linha 1, há a inclusão de um dado sobre o local de realização do estudo, na linha 10, houve a exclusão do artigo indefinido. No que diz

respeito aos colocados, podemos afirmar que aqueles posteriores a *estudo* são todos tipos de estudo: *transversal, comparativo, prospectivo, longitudinal* etc.

3.2.3. Pacientes⁵⁴ na seção Resultados

A palavra da qual pesquisamos os *concgrams* na seção *Resultados* foi *pacientes*. A lista dos primeiros *concgrams* de 3 palavras para essa palavra-nodo pode ser vista na tabela a seguir, que também mostra a frequência de ocorrência das associações e se elas foram consideradas para análise (A) ou não (NA). A lista completa encontra-se no Anexo III.

Quadro 25 – Concgrams de 3 palavras para a palavra-nodo pacientes

Palavra-nodo			Frequência	Analizados
pacientes	masculino	sexo	20	A
pacientes	foi	grupo	15	NA
pacientes	se	Observou	15	A
pacientes	distribuição	Tabela	14	NA
pacientes	dois	foi	12	NA
pacientes	Entre	Tabela	12	NA
pacientes	foi	três	12	NA
pacientes	anos	idade	11	A
pacientes	B	grupo	11	NA
pacientes	encaminhados	foram	11	A
pacientes	Entre	foi	11	NA
pacientes	etária	faixa	11	A
pacientes	foram	incluídos	11	A
pacientes	alogênico	TMO	10	A
pacientes	foram	não	10	NA
pacientes	segundo	Tabela	10	NA
pacientes	apenas	foi	9	NA
pacientes	casos	foi	9	NA
pacientes	Entre	não	9	NA
pacientes	Entre	p	9	A

3.2.3.1. pacientes/masculino/sexo

⁵⁴ Excluímos da análise os concgrams da palavra *Tabela*, o substantivo mais frequente da sessão Resultados, pois a maioria encontrava-se no título de legendas. Por exemplo: *Tabela 1 - Comparação da incidência de...*

1 61% do **sexo masculino**. Dois terços (65%) dos **pacientes** apresentavam doença prévia e 87% das admissões
 2 discreto do **sexo masculino** (56,1%). Vinte e um **pacientes** foram procedentes de Porto Alegre e da Grande
 3 dois **pacientes** (60%) eram do **sexo masculino**. Dez **pacientes** preencheram os critérios clínicos e ou
 4 148 restantes, 84 (56,7%) eram provenientes de **pacientes** do **sexo masculino**; a idade variou de um dia a
 5 apresentavam idade média de 19,2 meses, sendo 19 **pacientes** do **sexo masculino** e 21 do **sexo feminino**. Foram
 6 Nossa amostra, portanto, constitui-se de 74 **pacientes**, 38 do **sexo feminino** e 36 do **sexo masculino** com
 7 <Vol80Iss3_197> Oitenta e oito (52,7%) **pacientes** eram do **sexo masculino** e 79 (47,3%) eram do **sexo**
 8 3 e 22 anos (mediana de 13,5 anos), sendo 34 **pacientes** (68%) do **sexo masculino** e 16 do feminino (32%);
 9 das crianças. <Vol79Iss1_63> Trinta e seis **pacientes**, sendo 14 do **sexo masculino**, entre 3,3-15,4
 10 idade < 2 anos e < 5 anos; 57,1% dos **pacientes** pertenciam ao **sexo masculino**. As cepas foram
 11 e mediana de 10,6 anos). Quarenta e dois **pacientes** (60%) eram do **sexo masculino**. Dez **pacientes**
 12 meses e 5.050 (3.800-6.675) g. Trinta e quatro **pacientes** (54%) foram do **sexo masculino**. Os resultados das
 13 da população estudada (n=560). Dos 560 **pacientes**, 61,8% eram do **sexo masculino**, correspondendo a
 14 ou grave. Foram a óbito neste período 18 **pacientes** (17,3%), 10 do **sexo masculino** e oito do feminino,
 15 de maior renda. <Vol80Iss6_511> Entre os 327 **pacientes**, 171 (52,3%) eram do **sexo masculino** e 156 (47,7%)
 16 ao estado nutricional. <Vol79Iss4_337> Dos dez **pacientes** avaliados, seis eram do **sexo masculino**, e quatro
 17 <Vol79Iss5_435> Foram incluídos 111 **pacientes**. Houve leve predomínio do **sexo masculino** (54%) e

Figura 24 – Linhas de concordância do concgram pacientes/masculino/sexo

Nessas linhas, todas as ocorrências, exceto 1, 2, 3 e 17, que são ocasionais, tratam dos indivíduos do estudo. Mais uma vez podemos claramente distinguir a parte fixa (pacientes/masculino/sexo) e as partes variáveis. Nesses exemplos, em termos de variações em torno do *concgram*, podemos ver as diferentes maneiras de apresentar a população do estudo. Usando o verbo ser: *x pacientes eram/foram do sexo masculino* (7, 11, 12), com a variação de incluir *provenientes* (linha 4) ou o verbo *pertencer* (linha 10), usando um total e depois apresentando um recorte, *x pacientes, sendo y do sexo masculino; dos x pacientes, y eram do sexo masculino; entre os x pacientes, y eram do sexo masculino; sendo x pacientes do sexo masculino* (linhas 5, 6, 8, 9, 13, 14, 15, 16). Com base somente nesses trechos das linhas de concordância, teríamos na categoria de associações semânticas, os números e os percentuais.

As linhas 1, 2, 3 e 17 foram excluídas da análise.

3.2.3.2. pacientes/se/observou⁵⁵

1 **pacientes**. Observou-se esofagite grau I em 66 **pacientes**, esofagite grau II em quatro (2,4%) e esofagite
 2 Observou-se, ainda, que 80,4% dos **pacientes** clínicos e 78,1% dos **pacientes** cirúrgicos
 3 grau II à histologia, observou-se que 56 (93,3%) **pacientes** apresentaram EDA grau I e quatro (6,7%) grau II,
 4 normal e grau I, observou-se que: 65 (58%) **pacientes** apresentaram endoscopia normal e 47 (42%)
 5 prévios à internação, observou-se que 21% dos **pacientes** tinham história de BVA, 14% de pneumonia, 4% de
 6 não apresentaram febre. Observou-se que, nos **pacientes** portadores de malária, as manifestações
 7 aguda, e 2 (1,8%) septicemia. Observou-se que 29 **pacientes** (26,1%) tinham um ou mais diagnósticos de
 8 infecciosas diagnosticadas, observou-se que 23 **pacientes** (20,7%) apresentavam pneumonia, 6 (5,4%) otite
 9 nafazolina e fenoxazolina, observou-se que os **pacientes** expostos à nafazolina apresentaram um maior risco
 10 do refluxo vesíco-ureteral (RVU) dos **pacientes** Observou-se 77,6% de crianças do sexo feminino e
 11 (3%). A EDA foi normal em 96 (57,5%) **pacientes**. Observou-se esofagite grau I em 66 **pacientes**,
 12 em 21 havia fratura patológica (42%). Em 62% dos **pacientes**, observou-se dificuldade para deambular; outros
 13 contrações não inibidas em 62 (68,1%). Em quatro **pacientes** (4,4%), observou-se contração pós-miccional.
 14 da idade, do peso, da estatura e da DMO dos **pacientes** e controles. Observou-se diferença
 15 entre os **pacientes** com ou sem cicatriz renal Nos **pacientes** com cicatriz renal, observou-se maior e

Figura 25 – Linhas de concordância do concgram pacientes/se/observou

⁵⁵ Excluímos da análise os concgrams *pacientes/foi/grupo* pois eles constituíam-se na maioria em associações ocasionais

Nessas linhas, devemos considerar praticamente *observou-se* como um item só e como se combina com *pacientes*. Novamente, observamos como essas duas unidades podem se combinar com outras porções de texto com uma mesma função. Aqui temos, por exemplo, *observou-se*, seguido do pronome relativo *que* e do número de pacientes ou percentual (exemplo 2, 3, 4, 5, 7, 8). Nesses casos, algumas variáveis foram encontradas. Por exemplo, a posição em que o percentual é apresentado (antes ou depois de *pacientes*). Há ainda as estruturas: *observou-se* seguida do elemento observado e do número de pacientes (exemplo 1), número de pacientes seguido de *observou-se* (exemplos 12, 13). Como preferências semânticas presentes nessas linhas, citamos os números e percentuais. Como colocado, o verbo *apresentar*.

3.2.3.3. pacientes/anos/idade

```

1 + 3,66 anos); respectivamente, 52,9% e 81,4% dos pacientes apresentaram idade < 2 anos e < 5 anos; 57,1% dos
2 2 - Dados clínicos e antropomórficos de 22 pacientes com fibrose cística com menos de 5 anos de idade
3 4 - Dados clínicos e antropomórficos de 20 pacientes com fibrose cística com mais de 10 anos de idade
4 média de idade foi de 9,8±6,2 anos; 485 (69,7%) pacientes eram do gênero masculino e 211 (30,3%) do
5 idade < 2 anos e < 5 anos; 57,1% dos pacientes pertenciam ao sexo masculino. As cepas foram
6 A idade do início dos sintomas de 43 pacientes variou de 3 anos a 16 anos e 4 meses (mediana =
7 1,4% no SINPEL e 1,4% no HCRS. A idade dos 70 pacientes variou de um mês a 19,5 anos, com mediana de 1,92
8 a realizar o exame de urodinâmica. A idade dos pacientes variou de 5 a 17 anos. 45 (49,5%) pertenciam ao
9 pacientes em 54 admissões de UTIP. A idade dos pacientes variou de 1 mês a 13 anos, com mediana de 10,5
10 vitro à penicilina, 16 (69,5%) foram obtidas de pacientes com idade até dois anos. Essa taxa, quando
11 seguidos por pré-escolares (44%); apenas nove pacientes tinham idade superior a 6 anos. Ao diagnóstico,

```

Figura 26 – Linhas de concordância do concgram pacientes/anos/idade

Esse não é um *concgram* muito diferente dos que analisamos na sessão de *Introdução* com *crianças*. Ele associa os sujeitos do estudo e a descrição da idade, que varia em termos de mês e anos, em geral. Como podemos observar, as possíveis combinações dessas associações com as informações de idade são bem variadas: uma usa os símbolos de maior e menor, que não apareceram no *concgram crianças/anos/idade*; a palavra *idade* não necessariamente aparece na expressão *anos de idade*, ela ocorre em outra parte da frase, como nos exemplos 8, 10 e 11. Como colocado para esse *concgram* temos o verbo *variar*. Associação semântica, numerais e percentuais e palavras e símbolos que representam limites: < e >, de x a y, até.

3.2.3.4. pacientes/foram/encaminhados

1 os **pacientes** que foram encaminhados por leigos (**pacientes** da urodinâmica e professora) e considerando-se
 2 foram realizadas em 21 38 (55,3%) **pacientes**. Dez **pacientes** foram encaminhados para avaliação pré-operatória;
 3 maiores foram realizadas em 21 38 (55,3%) **pacientes**. Dez **pacientes** foram encaminhados para avaliação
 4 <Vol180Iss2_147> Foram incluídos na pesquisa 102 **pacientes** encaminhados por enurese ao Ambulatório de
 5 não epilépticas). Monitorização prolongada Dez **pacientes** foram encaminhados para o vídeo-EEG prolongado,
 6 frequência dos episódios de enurese noturna Os **pacientes** foram encaminhados por: pediatras (63,4%),
 7 73 (68,4%) (p < 0,001). Mesmo excluindo-se os **pacientes** que foram encaminhados por leigos (**pacientes** da
 8 síndrome de Down e infecções de repetição Cinco **pacientes** (11,1%) foram encaminhados por sepse nos
 9 horas, o qual foi conclusivo em nove casos. Dos **pacientes** monitorizados, dez foram encaminhados à

Figura 27 – Linhas de concordância do *concgram* pacientes/foram/encaminhados

As combinações desse *concgram* referem-se ao encaminhamento de pacientes a um exame ou serviço por determinado profissional ou leigo devido a uma determinada condição. Essas funções, no entanto, não são todas contempladas em todas as ocorrências. Conforme a função, uma diferente combinação com outras palavras é utilizada. Para indicar por quem o paciente foi encaminhado, há uma coligação com a preposição *por* (linhas 1 e 6), para indicar porque foi encaminhado, há uma coligação também com a preposição *por* (linhas 4 e 8), e indicando para onde ou para que foram encaminhados, há uma coligação com a preposição *para* ou *a* (linhas 2, 5 e 9).

Dentro de cada uma dessas 3 "categorias" há também variações que podem ser observadas, bem como outras regularidades.

3.2.3.5. pacientes/etária/faixa

1 **faixa etária** Ocorreram 16 (2,3%) óbitos entre os **pacientes** hospitalizados por injúrias; 12 foram devidos a
 2 a **faixa etária** Tabela 2 - Distribuição dos **pacientes** segundo a **faixa etária** e a região Tabela 3 -
 3 prevalências de prescrições por **faixa etária** de **pacientes** nas três classes de medicamentos A distribuição
 4 prevalências de prescrição por **faixa etária** de **pacientes**, nas três classes de medicamentos, é mostrada na
 5 e 12 anos = 24,2%. Tabela 1 - Distribuição dos **pacientes** segundo **faixa etária** e região A presença de IgE
 6 1. Na Tabela 2, observamos a distribuição dos **pacientes** alérgicos segundo a **faixa etária** e a região de
 7 para espaçador (Tabela 1). Em oito dos nove **pacientes**, independentemente da **faixa etária**, houve uma
 8 3 - Dados clínicos e antropomórficos de 32 **pacientes** com fibrose cística na **faixa etária** de 5 a 10
 9 feminino e intencionais. A distribuição dos **pacientes** por tipo de injúria e segundo a **faixa etária** e o
 10 na Tabela 4. Tabela 4 - Distribuição dos **pacientes** quanto aos tipos de injúrias segundo **faixa etária**

Figura 28 – Linhas de concordância do *concgram* pacientes/faixa/etária

Nessa combinação de palavras, agrupamos as linhas 2, 5, 6, 9, 10, pois elas tratam, evidentemente, de esclarecer que os pacientes foram distribuídos por faixa etária. Nessas linhas, encontramos variações no entorno do *concgram* como, por exemplo, a inclusão do artigo definido *a* (linha 2), de um qualificador para os pacientes (*alérgicos*, na linha 6), e uma expressão *por tipo de injúria* (linha 9). Não há variações de ordem do *concgram* nessas linhas.

As linhas 3, 4, 7 e 8 também são modos de apresentar uma distribuição por faixa etária, porém com estruturas bem diferentes.

Como colocados identificamos *distribuição* e *segundo*.

3.2.3.6. pacientes/foram/incluídos

1 **Foram incluídos** no estudo 127 RNPT, sendo 53 **pacientes** no grupo A (controle), que apresentaram média de
 2 30,7% **foram incluídos** por apresentarem OMEC. Os **pacientes** com OMR tinham uma média de 5,3±1,4 episódios de
 3 <Vol180Iss2_147> **Foram incluídos** na pesquisa 102 **pacientes** encaminhados por enurese ao Ambulatório de
 4 <Vol179Iss5_435> **Foram incluídos** 111 **pacientes**. Houve leve predomínio do sexo masculino (54%) e
 5 homozigota e heterozigota, **foram incluídos** 78 **pacientes**, sendo 23% (18) homozigotos e 77% (60)
 6 II Eventos clínicos **foram** registrados nos oito **pacientes incluídos** neste grupo, que apresentavam períodos
 7 nasofaríngea. Em função de duas perdas, 13 **pacientes foram incluídos** em nossa amostra. O estudo
 8 com peso de nascimento = 1.500g. Trinta (39%) **pacientes** não **foram incluídos** na pesquisa por óbito antes
 9 as características clínicas e laboratoriais dos **pacientes**. Para esta análise, **foram incluídos** 83 sujeitos,
 10 da análise pela PCR. Um total de 69,3% dos **pacientes** apresentou OMR, e 30,7% **foram incluídos** por
 11 no momento da internação, restando 49 **pacientes incluídos** na pesquisa. No CTIP-HMV, **foram**

Figura 29 – Linhas de concordância do conccram pacientes/foram/incluídos

Nessas linhas, excluímos da análise as linhas 2 e 9. Nas demais, observamos uma flexibilidade de combinações para a apresentação de quantos ou quais pacientes foram incluídos no estudo. Temos a estrutura das linhas 1 e 5, que faz um recorte do total: *Foram incluídos x, sendo...*. Nas linhas 3, 4, 7 e 8, são usados os mesmos itens, mas há uma mudança na ordem dos *conccrams*: *foram incluídos x pacientes* e *x pacientes foram incluídos*. Além disso, *pesquisa* foi incluído na linha 3 e, na linha 8, foi incluída uma negação.

Como colocados para esse *conccram* poderíamos citar palavras como *análise* e *pesquisa*.

3.2.3.7. pacientes/alogênico/TMO

1 **TMO alogênico** era negativa em apenas 20,0% dos **pacientes** (IC95%: 6,7 a 41,5%). No **TMO** autogênico, a
 2 óbitos tardios no **TMO alogênico**. Em três **pacientes**, a DECH foi a causa principal do óbito e, nos
 3 do procedimento. No **TMO alogênico**, seis **pacientes** (30%, IC 95%: 13,2 a 52,3) faleceram antes do D
 4 três, a causa principal foi a infecção. Três **pacientes** do grupo do **TMO alogênico** recidivaram com uma
 5 com os mesmos tempos. Considerações sobre os **pacientes** submetidos ao **TMO alogênico** com sangue de cordão
 6 por esta causa. Considerando-se todo o grupo de **pacientes** submetidos ao **TMO alogênico**, a DECH aguda
 7 (4/20) deles (IC 95%: 6,7 a 41,5%). Nos 14 **pacientes** submetidos ao **TMO alogênico**, que tiveram uma
 8 repetição e fibrose pulmonar. Observamos que os **pacientes** que receberam o **TMO alogênico** faleceram em uma
 9 entretanto analisamos a sobrevida dos 11 **pacientes** com leucemias agudas submetidos ao **TMO alogênico**,

Figura 30 – Linhas de concordância do conccram pacientes/alogênico/TMO

Observando as linhas de 5 a 9, podemos observar que há uma variação no verbo usado entre pacientes e TMO alogênico. Ainda, na linha 9, foi acrescentado um qualificador para *pacientes*. Na linha 4, observamos que foi incluído um grupo nominal onde antes havia verbos: *pacientes do grupo do TMO alogênico*. Nas linhas 1 e 3, podemos observar que o *conccram* tem uma associação com dados do Intervalo de Confiança (IC), informação que apareceu também próxima na linha 7.

3.2.3.8. pacientes/entre/p

1 significativa **entre** os dois grupos de **pacientes** ($p = 0,637$). Figura 1 - Curva de sobrevida
 2 enquanto que o fentanil foi o preferido **entre** os **pacientes** clínicos ($p < 0,01$) (Tabela 1). Chama a atenção
 3 na região hipofisária visualizadas à TC **entre** os **pacientes** dos Grupos A e B ($p = 0,55$); entretanto, os
 4 significativa nos níveis de IL-2 **entre** os **pacientes** com ou sem necessidade de VM ($p = 0,48$). A IL-2,
 5 diferença estatisticamente significante **entre** os **pacientes** crianças e controles crianças ($p = 0,595$). Tabela
 6 significativa ($p = 0,015$) da DMO **entre** os **pacientes** adolescentes e controles adolescentes. Não houve
 7 e terceira aferição ($p = 0,59$) **entre** os 47 **pacientes** com os dados disponíveis (Figura 4). Figura 4 -
 8 significativa ($p = 0,69$) **entre** os 43 **pacientes** com o dado disponível. O mesmo ocorrendo quando
 9 significativamente maior ($p = 0,002$) **entre** os **pacientes** clínicos ($6,8 \pm 5,0$ dias) do que nos **pacientes**

Figura 31 – Linhas de concordância do concgram pacientes/entre/p

Nessas linhas, identificamos facilmente a flexibilidade dos itens que coocorrem com o *concgram*. A primeira observação é que alguma análise ou comparação é feita, pois é apresentado um valor de p , que pode ser anterior ou posterior à expressão *entre os pacientes*. Esta, por sua vez, também contempla variações: *entre os dois grupos de*, *entre os x pacientes*. Percebe-se também que as palavras posteriores a *pacientes* tratam de qualificá-los ou de situá-los, podendo ser agrupadas em um grupo semântico de descrição de pacientes.

3.2.4. Crianças na seção Discussão

A palavra da qual pesquisamos os *concgrams* na seção *Métodos* foi *crianças*. A lista dos primeiros *concgrams* de 3 palavras para essa palavra-nodo pode ser vista no quadro a seguir, que também mostra a frequência de ocorrência das associações e se elas foram consideradas para análise (A) ou não (NA). A lista completa encontra-se no Anexo III.

Quadro 26 – Concgrams de 3 palavras para a palavra-nodo crianças

Palavra-nodo			Frequência	Analizados
crianças	al	et	37	NA
crianças	anos	idade	27	A
crianças	anos	entre	16	A
crianças	estudo	presente	16	A
crianças	entre	idade	14	NA
crianças	anos	até	13	NA
crianças	estudo	foi	13	NA
crianças	foi	não	13	NA
crianças	estudo	neste	12	A
crianças	até	idade	11	NA
crianças	dieta	leite	11	A
crianças	entre	mais	11	NA
crianças	foi	idade	11	NA
crianças	idade	meses	11	A
crianças	leite	vaca	11	NA
crianças	anos	foi	10	NA
crianças	dois	idade	10	NA
crianças	foi	mais	10	NA
crianças	meses	vida	10	A

crianças	não	se	10	NA
crianças	adolescentes	TVP	9	A

3.2.4.1. crianças/anos/idade

de EA diagnosticado por médico apenas em crianças de 6-7 anos de idade, essa redução não foi e da estimativa de que virtualmente todas as crianças até os 5 anos de idade já foram infectadas por mostra valores diminuídos de HDL-C em 13,5% de crianças de 2 a 12 anos de idade e em 14,2% de para se realizar o perfil lipídico em crianças acima de 2 anos de idade (9,10). Está bem Salbutamol não tem dose estabelecida para uso em crianças até quatro anos de idade, ainda que seja amostras coletadas até 1999, mas restritas às crianças de até seis anos de idade, evidenciou uma taxa relatados por Bruni (11), 74,9% ocorreram em crianças com até dois anos de idade. Mahieu et al. (9) após 3 horas. Não foi observada melhora em crianças abaixo dos 2 anos, mas acima dessa idade, os (6), a nafazolina não deveria ser usada em crianças abaixo de seis anos de idade. O autor recomenda o que endocardite infecciosa foi freqüente entre crianças abaixo de dois anos de idade e naquelas com 19 crianças. Dos 261 casos, 89,6% ocorreram em crianças com até quatro anos de idade, sendo 75% devidos à de endocardite infecciosa, principalmente entre crianças abaixo de dois anos de idade, em decorrência do (12,18). Stevenson et al. (19) demonstraram que crianças menores de cinco anos de idade, somente com dessas medicações, principalmente em crianças com menos de dois anos de idade (5,7,9-14,16). imidazolínicos, principalmente à nafazolina e em crianças com menos de três anos de idade, ocasiona, na isenta de glúten, enquanto que a minoria das crianças iniciou a dieta após dois anos de idade. Esse potencial da vacina conjugada 7-valente em crianças brasileiras com até os cinco anos de idade em crianças por ano (13,25,30,31), em seguimento de crianças a partir do nascimento até 2 a 3 anos de idade. e a idade. Os dois primeiros autores estudam crianças de três anos para o estudo normativo e concluem de ferritina associada ao aumento de idade das crianças haitianas entre 2 a 5 anos. A correlação entre predomínio de cepas resistentes à penicilina em crianças com idade até dois anos. Frente à resposta A cobertura da vacina conjugada 7-valente, em crianças com idade até os cinco anos, seria de 46,7% dos de sensibilização ao leite de vaca entre as crianças com idade inferior a 2 anos, enquanto que a das infecções pneumocócicas invasivas entre crianças com idade inferior a cinco anos, a maioria delas foi o caso. Mahu et al. (23), ao estudarem 384 crianças de Réunion com idade inferior a seis anos, também et al. (21), em estudo prospectivo no qual 679 crianças foram acompanhadas até a idade de 23 anos, com de cálcio a 800 mg dia por 18 meses para crianças de ambos os sexos com média de idade de 8,5 anos.

Figura 32 – Linhas de concordância do conagram crianças/idade/anos

Nesse *congram* podemos ver, novamente, a variação na posição dos *congrams* em trechos do texto que tratam de informar a idade das crianças do estudo. A variação que observamos é que entre *crianças* e *anos de idade* há diferentes maneiras de apresentar a idade ou a faixa etária: *abaixo de* (8, 9, 10, 12), *com até* (7), *com menos de* (14, 15), *a partir do nascimento até* (18), *de x a y anos* (1, 3), entre outras. Além disso, há uma outra maneira de organizar o *congram* que não segue o trigramma *anos de idade*, mas que segue a estrutura *crianças com idade inferior a* (23, 24, 25), *com idade até* (21, 22).

Da mesma forma que em outros casos, é possível inserir qualificadores (outros, que não os de idade) para *crianças*: *brasileiras* (17), *haitianas* (20), *de Reunión* (25).

3.2.4.2. crianças/anos/entre

doenças na faixa etária de 5 a 14 anos. Entre as crianças menores de 1 ano, as quedas foram o principal 12,4% em Recife e 17,7% em Porto Alegre para as crianças de 6-7 anos, e entre 10% em Curitiba e 14% em São entre 1,1% no Irã e 18,4% na Suécia para as crianças de 6-7 anos, e entre 0,8% na Albânia e 17,7% na entre 1992 e 1996, em estudo realizado em crianças carentes de três a cinco anos. Estes índices são Os autores verificaram, entre adolescentes e crianças de 6-7 anos, uma relação direta e inversa entre dados, as prevalências foram maiores entre as crianças de 6-7 anos de vida. Neste estudo, avaliamos a relacionados ao EA foi mais elevada entre as crianças de 6-7 anos. O preenchimento do QE pelos pais e de sensibilização ao leite de vaca entre as crianças com idade inferior a 2 anos, enquanto que a que endocardite infecciosa foi freqüente entre crianças abaixo de dois anos de idade e naquelas com de endocardite infecciosa, principalmente entre crianças abaixo de dois anos de idade, em decorrência do das infecções pneumocócicas invasivas entre crianças com idade inferior a cinco anos, a maioria delas como descongestionantes tópicos nasais para crianças entre dois e seis anos, não indicando nenhuma de ferritina associada ao aumento de idade das crianças haitianas entre 2 a 5 anos. A correlação entre

Figura 33 – Linhas de concordância do concgram crianças/anos/entre

Nesse caso, encontramos ocorrências em que um dos itens do *concgram* (*entre*) não apresenta uma relação com os outros dois itens (*crianças*, *anos*), como nas linhas 2, 3 e 4. Nas linhas de 5 a 11, a função é delimitar “em quem” determinado evento acontece ou foi encontrado. Com essa função, podemos ressaltar como itens flexíveis o uso do artigo definido *as* e as maneiras de apresentar a delimitação da idade: *de x a y*, *abaixo de*, *com idade inferior a*. Nas linhas 12 e 13, os *concgrams* estão associados, desempenhando a função de delimitar uma faixa etária. Observamos a possibilidade de acrescentar um qualificador para crianças na linha 13.

3.2.4.3. crianças/estudo/presente

1 o presente estudo Um dos primeiros estudos com crianças avaliando-se a resposta à vacina contra a
 2 No presente estudo, foi verificado que 71,1% das crianças e adolescentes tinham a obesidade presente na
 3 No presente estudo, que englobou 65 daquelas crianças, acrescentamos a avaliação nutricional,
 4 ao presente estudo. Em investigação com 207 crianças e adolescentes caucasianos de ambos os sexos
 5 no presente estudo. O achado de que 50% das crianças e adolescentes já haviam realizado dieta para
 6 como demonstrado no presente estudo, no qual as crianças submetidas à ventilação mecânica utilizaram, em
 7 das variáveis feita no presente estudo, que as crianças cujas mães receberam corticosteróide antenatal
 8 aguda (7,24). No presente estudo, as crianças que apresentaram febre provavelmente apresentaram
 9 próximas às detectadas no presente estudo em crianças indianas na faixa etária de 1 mês a 2 anos de
 10 inalatórios, para uso de broncodilatador, por crianças O presente estudo usou uma amostra de
 11 de soja no tratamento de constipação em crianças. Os resultados do presente estudo mostraram que,
 12 desse critério diagnóstico para identificar crianças com possível RGEF. No presente estudo, a
 13 não aprovados ou não padronizados para crianças, indistintamente (17). O presente estudo avaliou
 14 uma prevalência de 29% de hipovitaminose D entre crianças com colestase crônica. O presente estudo
 15 retardo no diagnóstico, principalmente em crianças menores. <Vol180Iss4_326> No presente estudo, foi

Figura 34 – Linhas de concordância do concgram crianças/estudo/presente

Nesses *concgrams*, muitas coocorrências são aleatórias, pois *presente estudo* está em uma frase e *crianças* em outra (linhas 1, 4, 5, 10-15). Restam as ocorrências 2, 3, 6-9. Em todas elas *presente estudo* aparece como um bigrama. A associação entre *crianças* e *presente estudo* parece ser mais por uma questão de pragmática do que de semântica, nos parece que *presente estudo* delimita o ambiente de realização da pesquisa, e *crianças* coocorre nessa mesma frase porque a maioria dos estudos é feita com crianças. Um ponto em comum entre as linhas de 6 a 9 é que a palavra *crianças* sempre é seguida de qualificadores. Isso parece estar associado com essa seção de *Discussão*, em que se procura esclarecer e delimitar o que se obteve no estudo com os sujeitos que foram estudados.

3.2.4.4. crianças/estudo/neste

1 neste estudo, entretanto, observou-se que as crianças soropositivas tiveram contato com maior número de
 2 (5). Neste estudo, documentamos que 67,6% das crianças com doenças alérgicas e 25,8% das não-alérgicas
 3 considerado neste estudo. Apesar de predominarem crianças da região Sudeste, a distribuição delas, segundo
 4 (22). Neste estudo, 71,4% das famílias das crianças soropositivas alcançaram, no máximo, meio salário
 5 Concluindo, neste estudo, verificou-se que as crianças em dieta isenta de leite de vaca e derivados
 6 (6-10%) (17-20). Neste estudo, 79% das crianças tinham como principal queixa a leucocoria,
 7 se correlacionam com os achados histológicos em crianças (15). Neste estudo, foi utilizada a classificação
 8 encontrada para AME aos quatro meses nas crianças seguidas neste estudo. O pediatra treinado foi
 9 internacional e ausente em nosso meio. As 20 crianças incluídas neste estudo, 10 meninos e 10 meninas,
 10 de linfócitos a apresentar-se reduzida em crianças de 9 a 12 anos (06). Neste estudo, observamos
 11 dados, as prevalências foram maiores entre as crianças de 6-7 anos de vida. Neste estudo, avaliamos a
 12 são desnecessárias, na maioria das vezes, para crianças sadias como as incluídas neste estudo, e

Figura 35 – Linhas de concordância do concgram crianças/estudo/neste

Este caso é semelhante ao anterior, e podemos considerar que o bigrama *Neste estudo* cumpriria o mesmo papel de *No presente estudo*, porém aqui percebemos o *concgram* assumindo duas funções diferentes no texto. Na primeira função, *neste estudo* indica o ambiente de realização do estudo (linhas 1, 2, 4, 5, 6) e percebe-se uma associação com percentuais (linhas 2, 4 e 6) e com verbo impessoal (*observou-se, verificou-se*) ou na primeira pessoa plural (*documentamos*). Na segunda função, o bigrama tem uma característica de qualificar as crianças. Nas frases, há uma variação do verbo usado: *crianças seguidas neste estudo* (linha 8), *crianças incluídas neste estudo* (9 e 12); a frase 12 ainda inclui um qualificador (*sadias*).

3.2.4.5. crianças/dieta/leite

1 permitiram constatar que a alimentação dessas crianças em dieta isenta de leite de vaca e seus derivados
 2 na Europa, que avaliaram o consumo alimentar de crianças em dieta isenta de leite de vaca e idade média em
 3 Outro estudo realizado na Finlândia avaliou 18 crianças em dieta isenta de leite de vaca e derivados (13
 4 o volume de fórmula consumido diariamente por crianças em dieta isenta de leite de vaca e derivados é
 5 Concluindo, neste estudo, verificou-se que as crianças em dieta isenta de leite de vaca e derivados
 6 que não recebiam fórmula (n = 10) e controles (crianças cuja dieta excluía ovos, mas não leite e
 7 da ingestão alimentar e do estado nutricional de crianças submetidas à dieta isenta de leite de vaca e
 8 de 2 anos (1,3-5). A ingestão alimentar de 16 crianças norueguesas em dieta isenta de leite de vaca e
 9 do leite de vaca. Em nosso estudo, duas das crianças em dieta isenta de leite de vaca e derivados
 10 menor do que a ingestão de leite por crianças com dieta normal (1,4). E, em nosso meio, há

Figura 36 – Linhas de concordância do concgram crianças/dieta/leite

Neste caso, podemos claramente ver que existe uma maior frequência de uma sequência fixa *crianças em dieta isenta de leite de vaca*. Porém, há variações como a inclusão do qualificador *crianças norueguesas* (linha 8), o uso do verbo *submetidas* (linha 7). Não há variações na ordem sintática dos elementos. As linhas 6 e 10 foram excluídas porque a associação é ocasional.

3.2.4.6. crianças/idade/meses

1 inicial. No nosso estudo, a idade média das crianças foi de 25 meses, concordando com a literatura,
 2 (8). Observou-se que a mediana de idade das crianças admitidas foi de 86 meses, apesar de o início das
 3 há relato de três casos de raquitismo em crianças belgas com idade entre 15 e 18 meses associado a
 4 da falta de apetite para refeições de sal em crianças clinicamente normais com idade de 8 meses a 5
 5 do HIV, Ramilo et al. (1996) (30) estudaram 16 crianças infectadas pelo vírus, com idade entre seis meses
 6 meses de idade corrigida entre os dois grupos de crianças que utilizaram oxigênio prolongadamente (uso de
 7 aos dois meses de idade, comparando-se com as crianças que estavam em aleitamento exclusivo (19).
 8 e terceiro meses de vida (9), e de 55,8% em crianças com idade entre 35 e 53 meses (15). Um estudo
 9 de cálcio a 800 mg dia por 18 meses para crianças de ambos os sexos com média de idade de 8,5 anos.
 10 diagnóstica de RGE. Provavelmente, algumas das crianças menores de 6 meses de idade identificadas como

Figura 37 – Linhas de concordância do concgram crianças/idade/meses

Nessas linhas, encontramos algumas regularidades e diferenças nas estruturas que relacionam as palavras crianças/idade/meses. Poderíamos agrupar as expressões em dois grupos. No primeiro, o enfoque é na idade das crianças:

Idade média das crianças foi de x meses (linha 1)

Mediana de idade das crianças admitidas foi de x meses (linha 2)

No segundo, grupo, o das frases 3, 4, 5, 8 e 10 temos a descrição da idade das crianças, mas o tema da frase não é a idade:

Crianças com idade entre/de 3, 4, 5, 8

Crianças menores de x meses de idade (10)

Podemos observar, também, que, no primeiro grupo, a ordem dos *concgrams* é mantida nas duas frases *idade crianças meses*. No segundo grupo, há uma flexibilidade.

3.2.4.7. crianças/meses/vida

1 6 meses de vida e identificaram que 20% das 185 crianças apresentavam síbilos, porém o tempo do seguimento
 2 18 meses e aos 5 anos de vida (14). Dentre as crianças com exame positivo aos 6 meses de vida, 75%
 3 e terceiro meses de vida (9), e de 55,8% em crianças com idade entre 35 e 53 meses (15). Um estudo
 4 antes dos 6 meses de vida, tanto para as crianças amamentadas, quanto para as alimentadas
 5 materno na alta e aos 3 meses de vida em crianças alimentadas por copo ou mamadeira. No entanto,
 6 de o início das crises ter ocorrido em 55,8% das crianças antes de 12 meses de vida, terem 92,9% delas sido
 7 as prevalências de 65 e 68%, respectivamente, em crianças no primeiro e terceiro meses de vida (9), e de
 8 <Vol80Iss4_291> Participaram do estudo 798 crianças na faixa etária de 1 a 12 meses de vida. Essa
 9 justificando a ocorrência de um maior número de crianças sintomáticas nos primeiros meses de vida. Outro

Figura 38 – Linhas de concordância do concgram crianças/meses/vida

Nesse *concgram*, percebemos que a associação *meses de vida* cumpre uma função diferente da que encontramos em *anos de vida*. Se observarmos, nas linhas 4 e 5, *meses de vida* não descreve, na frase, a idade das crianças, mas o momento em que algo aconteceu ou foi avaliado. Nessa função, observamos as variações *aos x meses de vida* (linha 5) e *antes dos x meses de vida* (linha 4). Nas frases de 6 a 9, *meses de vida* parece estar mais próximo de

cumprir a mesma função que *anos de vida*, com variações na apresentação da idade similares às descritas em outros *concgrams*.

3.2.4.8. crianças/adolescentes/TVP

1 adolescentes com TVP. Apesar de que a metade das crianças investigadas apresentou deficiência de uma ou
 2 ou mais distúrbios trombofílicos em 9 14 (64%) crianças e adolescentes com TVP. Apesar de que a metade
 3 de controles pareados por idade com as crianças e adolescentes com TVP foi necessária para a
 4 inibidoras da coagulação, na maioria das crianças e adolescentes com TVP, é consequência da
 5 das proteínas inibidoras da coagulação nas crianças e adolescentes com TVP é comparável àquela
 6 que, em um destes estudos, foram investigadas 49 crianças e adolescentes com TVP (14). Resultados
 7 a presença da mutação G20210A da protrombina em crianças e adolescentes com TVP, o nosso estudo foi o
 8 da proteína C é frequentemente encontrada nas crianças e adolescentes com TVP, mas não parece ser de
 9 um papel importante em relação à trombose nas crianças e adolescentes com TVP estudadas. Agradecimentos

Figura 39 – Linhas de concordância do concgram crianças/com/TVP

Fica evidente nessas linhas que o *concgram crianças/adolescentes/TVP* não tem uma flexibilidade em sua ocorrência nas frases, figurando como uma estrutura mais fixa (a linha 1 deve ser excluída da análise). Nesse caso, também não observamos palavras que indicam ser colocados; aparentemente, as estruturas que antecedem o concgram são bem específicas do estudo em questão: *distúrbios trombofílicos*, *trombose*, *protrombina*, *proteína C*. Há, no entanto, uma associação semântica com itens relacionados a vasos e sangue: *coagulação*, *protrombina*, *trombofílicos*, *trombose*.

3.3. Discussão sobre os resultados da Etapa 1

Primeiramente, gostaríamos de salientar que essa foi uma análise de cunho qualitativo, portanto não houve uma sistematização absoluta ou exaustiva dos dados. O tema da flexibilidade das associações foi abordado de maneira a englobar os resultados fornecidos pelo corpus na máxima extensão possível, excluindo-se apenas os casos em que as coocorrências eram acidentais ou repetiam-se muito e, por isso, já haviam sido analisadas.

Esperamos que nosso ponto de vista tenha ficado claro, principalmente a ideia de que, para que uma associação desempenhe uma mesma função, seus componentes textuais não precisam estar necessariamente numa mesma ordem ou ter uma estrutura fixa. Quisemos mostrar que, mesmo as palavras assumindo posições diferentes no texto, um mesmo significado ideacional pode ser transmitido. A descrição da idade dos pacientes demonstra isso com bastante clareza. Com a associação *crianças/idade/meses*, há uma série de possibilidades de formação das frases (ver item 3.2.1.8).

Também observamos que, na maioria dos casos, as palavras de um *concgram* estão associadas a um mesmo significado ideacional. Porém, há também casos em que mais de um significado ideacional pode ser inferido. Nesses casos, os itens do entorno do *concgram* e seus colocados ajudam a definir o contexto e o sentido.

O que estamos apresentando aqui, também, é que, mesmo a noção ampla de matrizes fraseológicas, como proposta por Gouadec e discutida anteriormente, não poderia dar conta da distribuição das unidades que investigamos aqui. Na associação *crianças/anos/idade* (3.2.1.1), por exemplo, seriam necessárias duas matrizes para incorporar as possíveis fraseologias:

crianças [] anos de idade, em que a variável poderia ser *abaixo de*, *menores de*, *na faixa etária de*, entre outras, e:

crianças com idade [] anos, em que a variável seria *inferior a*.

Poderíamos também afirmar que, no primeiro caso, a variável seria preenchida por elementos relacionados à delimitação de faixa etária e numerais. Seriam preferências semânticas para essa associação.

Os exemplos analisados acima mostram que, em vez de procurar definir um padrão fraseológico sintático, podemos pensar em grupos de palavras que se associam, que seriam itens fixos, e outras partes flexíveis, como seus colocados, coligações, associações semânticas. Quando se pensa na aplicação desse formato em um software de auxílio ao tradutor, que trataremos na Parte II desta tese, a possibilidade de identificar trechos semelhantes aos armazenados em uma memória de tradução ou banco de dados aumenta. Assim, buscando por *crianças/anos/idade* encontraríamos, no caso do exemplo 3.2.1.1, nove frases que poderiam auxiliar na tradução, ao invés de apenas uma. Além disso, com o uso de *concgrams* foi possível encontrar exemplos em que o espaço entre a palavra-nodo e as associações é variável.

Os tipos de variações encontrados na análise dos *concgrams* e seus contextos foram:

- inclusão de classificadores: (crianças) saudáveis, haitianas, normais;
- diferentes proposições: (em) crianças, (das) crianças;
- palavras de mesmo grupo semântico: *medir*, *controlar*, *avaliar a pressão*;
- alteração de ordem na frase; e
- inclusão de um número ou percentual.

Comparando-se também alguns *congrams* entre as seções, apenas a título de “curiosidade” a respeito do princípio idiomático, chama a atenção que as variáveis do *congram* *crianças/anos/idade*, por exemplo, na maior parte das vezes, usa delimitadores como *até x anos de idade*, *abaixo de x anos de idade*, *menores de*; enquanto *pacientes/anos/idade*, na seção *Resultados*, usa *a idade dos pacientes variou*, e lendo os contextos temos a impressão de que não se trata de crianças pequenas (menores de 6 anos). Pelos nossos exemplos, essa parece ser uma preferência semântica das palavras-nodo, ou dos próprios *congrams*.

Com essa análise, foi possível identificar as coocorrências de duas palavras a uma palavra-nodo independentemente da variação de posição e constituição. Assim, entendemos, podemos observar melhor as variações fraseológicas e encontrar dados brutos que permitam identificar outras formas de regularidade em textos especializados.

4. Etapa 2: associações semânticas e pragmáticas

De acordo com a teoria dos lexical *primings* (Hoey, 2005), as palavras podem estar associadas por diferentes “motivações” ou “forças”: semânticas, pragmáticas, gramaticais, entre outras. Segundo o autor, sempre que encontramos uma palavra (uma sílaba ou uma combinação de palavras), subconscientemente percebemos palavras que coocorrem com elas (colocados), significados com os quais está associada (associações semânticas), funções pragmáticas (associações pragmáticas) e padrões gramaticais (coligações) com as quais está associada. No desdobramento de nossa hipótese, imaginamos que entre duas ou mais palavras também se pode descrever uma “força” que une as palavras.

Entendemos, no caso das unidades que estamos estudando, em artigos científicos, ser possível identificar se o que motiva a associação é de ordem predominantemente semântica ou pragmática. Sendo assim, essa seria mais uma característica das unidades que estamos estudando e que, como já discutimos, podem ser consideradas unidades de tradução. Nosso intuito é demonstrar que as associações não são aleatórias e que elas são unidas, no eixo sintagmático, por essas motivações semânticas, pragmáticas e gramaticais. Consideramos, então, associações com atração semântica aquelas que contêm palavras que pertencem a um mesmo ambiente semântico que, no nosso caso, tem a ver com a Pediatria; e as associações com atração pragmática contêm palavras que se associam a determinadas funções

pragmáticas, em nosso caso, a seção do artigo em que se encontram ajuda a identificar essas funções.

4.1. Metodologia

4.1.1. Seleção das palavras-nodo

De acordo com estudo-piloto previamente realizado, havíamos detectado palavras-chave exclusivas de cada seção, além de outras que eram comuns a todas as seções. Utilizamos algumas dessas palavras para demonstrar associações pragmáticas e semânticas. A seguir, listamos todas elas novamente.

- Palavras que só ocorrem na seção *Introdução*:

alterações, amamentação, aumento, avaliar, clínica, clínicos, desenvolvimento, il, infância, lesão, mortalidade, neonatal, nutricional, obesidade, objetivo, óssea, pré, saúde, terapia, termo

- Palavras que só ocorrem na seção *Métodos*:

amostras, análise, após, através, c, coleta, comitê, critérios, ética, exame, g, kg, período, pesquisa, protocolo, quadrado, qui, realizada, realizado, significância, teste, utilizado, utilizando

- Palavras que só ocorrem na seção *Resultados*:

apresentaram, apresentavam, características, curva, diferença, distribuição, estatisticamente, feminino, figura, grau, houve, i, masculino, média, mediana, médias, meninas, meses, n, nascimento, p, respectivamente, tabela, valores

- Palavras que só ocorrem na seção *Discussão*:

achados, associação, autores, dieta, literatura, maior, menor, neste, níveis, pode, presente, resultados

- Palavras comuns a todas as seções:

crianças, diagnóstico, estudo, idade, nascidos, pacientes, peso, recém, uso.

Observando-se as palavras que são chave em cada seção, já podemos perceber que elas têm relação com a seção em que se encontram (um fator mais pragmático) ou, obviamente, com o domínio em questão (um fator mais semântico). Por exemplo, na seção *Métodos* temos palavras como *amostra, análise, coleta*; na seção *Resultados*, a primeira palavra é um verbo no passado *apresentaram*, na *Discussão*, palavras como *autores e literatura*.

4.1.2. Processamento para obter *concgrams*

O próximo passo foi processar os *córpus* para obter os *concgrams* de cada palavra-nodo selecionada. Novamente, usamos uma lista de palavras a serem excluídas da busca, com isso evitando que o sistema apresentasse como *concgram* associações com palavras que são muito frequentes na Língua Portuguesa, como artigos e preposições (a lista pode ser vista no Anexo II). Como nesta etapa nosso enfoque é demonstrar exemplos de como as nossas unidades de análise podem se associar no nível semântico ou pragmático, nós selecionamos algumas das palavras-chave acima descritas, conforme julgamos que elas poderiam evidenciar melhor nosso pressuposto. Também não nos restringimos aos *concgrams* formados por 3 palavras e apresentamos tanto exemplos de *concgrams* de 2 palavras como de 4 e 5.

4.2. Resultados

A seguir apresentamos 4 quadros com alguns *concgrams* de palavras selecionadas: *amamentação, avaliar e objetivo* para *Introdução*; *amostras, análise, comitê e pacientes* para *Métodos*; *apresentaram, características, curva e feminino* para *Resultados*; *achados, associação, autores*, para *Discussão*. Apresentamos os *concgrams* em quadros e, ao lado de cada grupo de linhas de contexto, classificamos que tipo de *priming* acreditamos que está sendo predominantemente ativado: semântico (S) ou pragmático (P).

Para determinar cada classificação, entendemos que as associações semânticas são aquelas em que prevalece uma coocorrência em que os itens são mais relacionados à Medicina, à Pediatria, às crianças, ou seja, ao domínio de nosso *córpus*. As associações pragmáticas são aquelas em que percebemos a “força” da seção em que as palavras coocorrem. Para seguir um critério nessa classificação, já que muitas vezes ficaria difícil determinar qual a motivação predominante, nos fizemos a pergunta: "Essas palavras estariam

associadas se não estivessem em um artigo científico de Pediatria?". Se a resposta fosse "sim", seria uma indicação de que estão associadas por motivos semânticos. Se a resposta fosse "não", seria porque o gênero as estaria determinando. Por vezes também utilizamos o Google para ajudar em nossa intuição.

4.2.1. Seção *Introdução*

Quadro 27 – Concgrams de palavras-nodo da seção *Introdução*

amamentação/duração		
1 2 3 4 5	positiva ou negativamente o início da amamentação e sua duração . Em particular, as enfermeiras e os aquém das recomendações da OMS, com mediana de amamentação exclusiva de apenas um mês, duração total de de 4 a 6 meses, e a mediana de duração da amamentação exclusiva foi de apenas 1 mês (5,6). Assim, fica profissionais, ocasionando maior duração da amamentação nas comunidades por eles assistidas (7,8). Este exclusiva de apenas um mês, duração total de amamentação de sete meses, e o aleitamento continuado um ano	S
Assoc. semântica: motivada pelo contexto de cuidados pediátricos		
amamentação/exclusiva/meses e amamentação/exclusiva/mês		
1 2 3 4 5	exclusiva de apenas um mês, duração total de amamentação de sete meses, e o aleitamento continuado um ano a política nacional coincidem na recomendação da amamentação exclusiva nos 6 primeiros meses de vida (1,5). Na infantil. Atualmente, a OMS recomenda: (1) amamentação exclusiva durante os seis primeiros meses de Declaração Innocenti (8), em 1990, preconizou a amamentação exclusiva nos primeiros quatro a seis meses de de 4 a 6 meses, e a mediana de duração da amamentação exclusiva foi de apenas 1 mês (5,6). Assim, fica	S
1 2 3 4	exclusiva de apenas um mês, duração total de amamentação de sete meses, e o aleitamento continuado um ano preconizada pela Organização Mundial da Saúde: amamentação exclusiva até o sexto mês e aleitamento materno aquém das recomendações da OMS, com mediana de amamentação exclusiva de apenas um mês, duração total de de 4 a 6 meses, e a mediana de duração da amamentação exclusiva foi de apenas 1 mês (5,6). Assim, fica	S
Assoc. semântica: motivada pelo contexto de cuidados pediátricos		
avaliar/estudo/objetivo		
1 2 3 4 5 6 7 8 9 10 11 12 13	um estudo transversal com o objetivo de avaliar o grau de proteção do aleitamento materno contra literatura. O presente estudo tem por objetivo avaliar as características epidemiológicas, clínicas e adultos (6). O objetivo deste estudo consiste em avaliar aspectos clínicos e histológicos da úlcera (16-18). O objetivo do presente estudo foi avaliar os efeitos da estimulação imunológica induzida (11). O objetivo do presente estudo foi avaliar o consumo de nutrientes e o estado nutricional de (12-14). O objetivo do presente estudo é avaliar a frequência da deficiência de proteína C (PC), em crianças (12). O objetivo deste estudo foi avaliar a extensão do uso de medicamentos não apropriados segundo caso (7,8). O objetivo deste estudo foi avaliar a eficácia de budesonida na forma de suspensão, psicomotoras. O objetivo deste estudo foi avaliar o desenvolvimento cognitivo na idade pré-escolar diversas (3,6,16). O objetivo deste estudo foi avaliar a prevalência das alterações auditivas em (2,3,7-12). O objetivo deste estudo é avaliar o impacto da vacina contra Hib sobre a situação de IL-6 (12). O objetivo deste estudo é avaliar os diferentes achados perinatais e o tempo da dos recém-nascidos. O objetivo deste estudo é avaliar os níveis de leptina do cordão umbilical em	P
Assoc. pragmática: motivada pelo gênero/seção em que se encontra		
objetivo/estudo/foi		

1	de radiação (1). Este estudo foi realizado com o objetivo de comparar a DMO de crianças e adolescentes com	P
2	regulares e se vêem como adultos fumantes. O objetivo do estudo foi determinar a prevalência do	
3	análise das características dos fibrocísticos. O objetivo deste estudo foi descrever as características	
4	de surdez e síndromes diversas (3,6,16). O objetivo deste estudo foi avaliar a prevalência das	
5	de utilidade na avaliação desses pacientes. O objetivo deste estudo foi verificar a incidência de	
6	medicamentosas para uso em crianças (12). O objetivo deste estudo foi avaliar a extensão do uso de	
7	de prevenção e controle. Nesse contexto, o objetivo deste estudo foi investigar os fatores	
8	de parceiros e DST nos últimos seis meses. O objetivo deste estudo foi identificar a ocorrência de	
9	(1,5,7-9-11). Diante dessas considerações, o objetivo deste estudo foi efetuar uma análise	
10	número de reagudizações no segundo caso (7,8). O objetivo deste estudo foi avaliar a eficácia de budesonida	
11	(UTIPs) e neonatais (UTINs). Assim, o objetivo deste estudo foi analisar o procedimento de	
12	associadas a instabilidades psicomotoras. O objetivo deste estudo foi avaliar o desenvolvimento	
13	de prematuros de muito baixo peso. O objetivo deste estudo foi verificar a taxa de incidência	
14	maior dos índices de obesidade observados. O objetivo deste estudo foi, então, comparar um programa de	
15	comportamento alimentar adequado ou não (11). O objetivo do presente estudo foi descrever os hábitos	
16	sejam detectadas e corrigidas (11). O objetivo do presente estudo foi avaliar o consumo de	
17	do Ministério da Saúde do Brasil (16-18). O objetivo do presente estudo foi avaliar os efeitos da	
18	foi realizado um estudo transversal com o objetivo de avaliar o grau de proteção do aleitamento	
Assoc. pragmática: motivada pelo gênero/seção em que se encontra		
Objetivo/estudo/teve/como		
1	Nesse contexto, o presente estudo teve como objetivo analisar os conhecimentos dos pediatras que atuam	P
2	de prevenção. O presente estudo teve como objetivo determinar a prevalência de sobrepeso e obesidade	
3	base nesses fundamentos, este estudo teve como objetivo investigar o perfil lipídico e a peroxidação de	
4	nacionais sobre o tema. O estudo teve como objetivo verificar a prevalência do RGEp em um grupo de	
Assoc. pragmática: motivada pelo gênero/seção em que se encontra		

4.2.2. Seção Métodos

Quadro 28 – Concgrams de palavras-nodo da seção Métodos

amostras/Student/t		
1	t de Student para igualdade de médias em duas amostras independentes para as variáveis contínuas. Os	S
2	estatística O teste t de Student para duas amostras independentes foi utilizado para comparar o peso	
3	grupos através do teste t de Student para duas amostras independentes. Nos casos de amostras com	
4	foi possível usar o teste t de Student para as amostras independentes. O teste de Mann-Whitney foi usado	
5	qui-quadrado (X ²) ou o teste t de Student, para amostras pareadas. Na interpretação dos testes	
6	das correlações; e o teste t de Student, para amostras independentes, utilizado nas comparações. P-value	
7	foi realizada pelo teste t de Student, para amostras independentes, com estimativa de diferença média	
8	foram comparadas pelo teste t de Student para amostras independentes e, na situação de três ou mais	
9	testadas. Utilizou-se o teste t de Student para amostras independentes, para comparar os resultados médios	
10	das médias por meio do teste t de Student para amostras independentes e pela comparação de proporções por	
Assoc. semântica: motivada pelo contexto de cálculos estatísticos		
amostras/coletadas/foram		
1	ativas e clinicamente estáveis. Foram coletadas amostras de todos os leites humanos (LHC e LHP) oferecidos	P
2	padronizado pela Universidade de Taubaté. As amostras foram coletadas pelo pesquisador, previamente	
3	volume total) e em seguida congelado a -70°C. As amostras foram coletadas nas primeiras 48 horas de vida do	
4	por 10 minutos (leite intermediário). Essas amostras foram coletadas diretamente do seio, após	
5	interferências na adesão aos esquemas. As amostras foram coletadas em tubos contendo anticoagulante	
6	de mães presentes na UTI neonatal do IFF. As amostras de LHP foram coletadas dos recipientes enviados à	
7	segundo o recomendado por Wang et al. (4). As amostras do LHC foram coletadas por meio de expressão	
8	espessa, sem porcionamento da amostra. Todas as amostras (LHC e LHP) foram coletadas em triplicata	
9	Determinação dos níveis plasmáticos de IL-6 As amostras de sangue (100 l) foram coletadas em tubos com	
Assoc. pragmática: motivada pelo gênero/seção em que se encontra		
análise/estatística/foi		

1	CTC NCI) (9). Análise estatística Foi realizada análise descritiva, utilizando-se a média, o desvio	P
2	desses programas. A estratégia adotada para a análise estatística foi inicialmente de comparações	
3	resultados normais para a faixa etária. A análise estatística foi realizada através do teste de	
4	inquéritos foram digitados em dupla entrada. A análise estatística foi feita nos programas Epi-Info 200	
5	em 0,8% das amostras (22). Análise dos dados A análise estatística foi descritiva, utilizando-se o	
6	para a discriminação dos grupos. Toda a análise estatística foi realizada através do SPSS(r)	
7	do Hospital Infantil Nossa Senhora da Glória. Análise estatística Foi feita com a utilização do	
8	para coeficiente de correlação (r) de 0,7. A análise estatística foi realizada através dos programas	
9	National Health Institute CTC NCI) (9). Análise estatística Foi realizada análise descritiva,	
10	ao uso de opióides no período neonatal. A análise estatística foi descritiva. A comparação das	
11	possíveis dificuldades no preenchimento. Para a análise estatística foi utilizado o cálculo das médias	
12	utilizou-se o teste exato de Fisher. Toda a análise estatística foi realizada utilizando o software	
13	dados através do programa Epi-Info versão 6, e a análise estatística foi feita pelo programa SAS versão	
14	múltiplas sem diagnóstico síndrome). A análise estatística dos dados foi realizada no programa	
15	com o padrão de peso molecular utilizado. A análise estatística dos resultados foi realizada por meio	
16	Spectramax no comprimento de onda de 570 nm. Análise estatística O tamanho da amostra foi calculado	
17	o pneumococo isolado em 0,8% das amostras (22). Análise dos dados A análise estatística foi descritiva,	
18	foi usado o programa Scion Image (17). Para a análise estatística , os dados quantitativos obtidos fora	
19	Esse escore foi utilizado para processamento da análise estatística , que foi realizada através do teste	
20	do paciente não foi objeto deste trabalho. A análise estatística foi realizada utilizando-se o	
21	do paciente não foi objeto deste trabalho. A análise estatística foi realizada utilizando-se o	
Assoc. pragmática: motivada pelo gênero/seção em que se encontra		
Análise/estatística/realizada		
1	CTC NCI) (9). Análise estatística Foi realizada análise descritiva, utilizando-se a média, o desvio	P
2	para a discriminação dos grupos. Toda a análise estatística foi realizada através do SPSS(r)	
3	do paciente não foi objeto deste trabalho. A análise estatística foi realizada utilizando-se o	
4	National Health Institute CTC NCI) (9). Análise estatística Foi realizada análise descritiva,	
5	para coeficiente de correlação (r) de 0,7. A análise estatística foi realizada através dos programas	
6	utilizou-se o teste exato de Fisher. Toda a análise estatística foi realizada utilizando o software	
7	do paciente não foi objeto deste trabalho. A análise estatística foi realizada utilizando-se o	
8	resultados normais para a faixa etária. A análise estatística foi realizada através do teste de	
9	Esse escore foi utilizado para processamento da análise estatística , que foi realizada através de U	
10	múltiplas sem diagnóstico síndrome). A análise estatística dos dados foi realizada no programa	
11	com o padrão de peso molecular utilizado. A análise estatística dos resultados foi realizada por meio	
Assoc. pragmática: motivada pelo gênero/seção em que se encontra		
Análise/descritiva/foi/realizada		
1	realizada em planilha Excel(r). Foi realizada análise descritiva dos resultados obtidos, bem como	P
2	Package for Social Science (SPSS). Foi realizada análise descritiva, para caracterizar a população	
3	CTC NCI) (9). Análise estatística Foi realizada análise descritiva, utilizando-se a média, o desvio	
4	do aluno. Foi realizada, preliminarmente, uma análise descritiva da variável dependente (consumo de	
5	National Health Institute CTC NCI) (9). Análise estatística Foi realizada análise descritiva,	
Assoc. pragmática: motivada pelo gênero/seção em que se encontra		
análise/variância/foram/ANOVA/não		
1	dependentes do tempo foram avaliadas pela análise de variância (ANOVA) para dados não-balanceados	P
2	dependentes do tempo foram avaliadas pela análise de variância (ANOVA) para dados não balanceados	
3	foi de 30 crianças. Foram utilizados a análise de variância (ANOVA) e os testes não-paramétricos	
Assoc. pragmática: motivada pelo gênero/seção em que se encontra		
comitê/aprovado/pesquisa		

1	informado. O projeto foi aprovado pelo Comitê de Ética e Pesquisa do Hospital de Clínicas de	S
2	a participar do estudo, que foi aprovado pelo Comitê de Ética e Pesquisa de nossa instituição, tendo	
3	populacional, do tipo transversal, aprovado pelo Comitê de Ética e Pesquisa da Universidade Federal da	
4	O protocolo de investigação foi aprovado pelo Comitê de Ética em Pesquisa da instituição. A escolha do	
5	e esclarecido. Este trabalho foi aprovado pelo Comitê de Ética em Pesquisa da UFMG. Foram excluídos	
6	Stata 5.0. O presente trabalho foi aprovado pelo Comitê de Ética em Pesquisa da FMTM. Tabela 1 -	
7	O presente estudo foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal do	
8	(19), e teve seu início após ser aprovado pelo Comitê de Ética em Pesquisa do Instituto Fernandes	
9	Aspectos éticos O projeto foi aprovado pelo Comitê de Ética em Pesquisa do Centro Geral de Pediatria	
10	de Tanner (10). O estudo foi aprovado pelo Comitê de Ética em Pesquisa da UFRN. O consentimento	
11	diretor da escola. O estudo foi aprovado pelo Comitê de Ética em Pesquisa do Instituto	
12	O protocolo do estudo foi aprovado pelo Comitê de Ética em Pesquisa do Centro de Ciências da	
13	a= 5% e β = 20%. O estudo foi aprovado pelo Comitê de Ética em Pesquisa da Faculdade de Ciências	
14	(13,17). O estudo foi submetido e aprovado pelo Comitê de Ética em Pesquisa com Seres Humanos da UFSC,	
15	de rotina do serviço. O estudo foi aprovado pelo Comitê de Ética em Pesquisa do Hospital de Clínicas.	
16	O projeto deste artigo foi aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de	
17	do Hospital Pequeno Príncipe e aprovado pelo Comitê de Ética em Pesquisa da instituição. Foram	
18	multivariada (17). O estudo foi aprovado pelo Comitê de Ética em Pesquisa do Hospital Universitário da	
19	O projeto foi submetido e aprovado pelo Comitê de Ética em Pesquisa da Faculdade de Medicina da	
20	foi $\alpha = 0,05$. O projeto foi aprovado pelo Comitê de Ética em Pesquisa da Faculdade de Ciências	
21	estatística. O projeto foi aprovado pelo Comitê de Ética em Pesquisa da FCM-Unicamp, parecer 461	
22	0,20. O protocolo do estudo foi aprovado pelo Comitê de Ética em pesquisa do Complexo Hospitalar Santa	
23	e não-hepatopatas. O estudo foi aprovado pelo Comitê de Ética em Pesquisa Médica da UFMG. Foi obtido	
24	de nulidade. Este estudo foi aprovado pelo Comitê de Ética em Pesquisa da UNIFESP-EPM.	
25	de pesquisa foi aprovado previamente pelo Comitê de Ética do Hospital Universitário da	
26	de pesquisa foi analisado e aprovado pelo Comitê de Ética em Pesquisa da Universidade Federal de	
27	0,05. O protocolo de pesquisa foi aprovado pelo Comitê de Ética em Pesquisa do Grupo de Pesquisa e	
Assoc. semântica: motivada pelo contexto de aprovação por parte de um comitê de ética		
pacientes/exame		
1	excluído, devido à presença de artefato no seu exame densitométrico. Dos 23 pacientes estudados, nove	S
2	eliminação do material radioativo inalado no exame anterior. Em cada procedimento, os pacientes eram	
3	no início do estudo, uma vez que esse era uma exame de sangue rotineiro anual. Apenas três pacientes	
4	o RVU primário, porém 58 pacientes realizaram o exame de cintilografia renal com DMSA, no início do	
5	a extubação, os pacientes foram submetidos a exame endoscópico da via aérea, conduzido em todos os	
6	auto-imune. Os pacientes foram submetidos ao exame ultra-sonográfico do fígado e vias biliares no	
7	estudo 1.261 pacientes que haviam realizado exame sorológico para toxoplasmose no laboratório	
8	Cada grupo de pacientes foi submetido tanto ao exame para pesquisa de plasmódio quanto ao dos	
9	os pacientes foram submetidos a novo exame pela PDC. A metodologia de análise incluiu	
10	grupo de pacientes foi submetido à anamnese e exame físico detalhado, que não evidenciaram sintomas e	
11	dos pacientes , obtidas no questionário e no exame de cada grupo, para avaliar se os dois grupos	
Assoc. semântica: motivada pelo contexto de cuidados médicos		

4.2.3. Seção Resultados

Quadro 29 – Concgrams de palavras-nodo da seção Resultados

apresentaram/crianças	
------------------------------	--

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27	<p>que, independentemente do sexo, as crianças apresentaram uma redução em torno de 50% do ganho ponderal e o sexo No Alto Xingu, oito (8,4%) crianças apresentaram baixa estatura, condição evidenciada em 26 ou estatura dentro da normalidade, oito crianças apresentaram peso abaixo do percentil 5 e 7 e estatura abaixo de 2SD. Cerca de 40% das crianças apresentaram prolongado tempo de internação no berçário (28,6%) (Figura 1). Treze (46,4%) crianças apresentaram sepse concomitante, e três (10,7%) desenvolveram (Aerochamber(r) e Inal-Air(r)). As crianças apresentaram uma diferença significativa na deposição pulmonar e significativa proporção de crianças que apresentaram febre. Nas crianças com RVU dilatado, 33 a 34 semanas (n = 97) Entre as 45 crianças que apresentaram síndrome de desconforto respiratório, 32 <Vol79Iss6_504> Das 190 crianças que se apresentaram para o seguimento de puericultura no ambulatório para a pressão diastólica; as crianças obesas apresentaram níveis menores de pressão diastólica. Figur de linfócitos CD8+, seis crianças (18,8%) apresentaram valores acima do percentil 95, e três (9,4%) Nos casos de hepatite aguda, 22 crianças (73,3%) apresentaram febre no início da doença, e destas, apenas 13 (RVU) No DMSA inicial, 32 crianças (55,2 %) apresentaram cicatrizes renais. Duas crianças do sexo notamos que as crianças do Alto Xingu apresentaram valores significativamente maiores do que os <Vol79Iss6_455> Das 1.090 crianças avaliadas, 68 apresentaram escore z de peso para estatura acima de 2, 4 mostra que 12,8% das crianças e adolescentes apresentaram um único fator de risco para aterosclerose e 72,5, 20,2 e 7,3% das crianças e adolescentes apresentaram, respectivamente, valores séricos de CT sem asma. Porém, as crianças portadoras de asma apresentaram, significativamente ($p < 0,0001$), frequência 12,5% das crianças da população Ikpeng apresentaram peso abaixo do mínimo esperado para a idade. Cerca de 6% (três crianças) dos recém-nascidos apresentaram escores ruins na avaliação da sucção, isto é febre. Nas crianças com RVU dilatado, 33 apresentaram valores significativamente maiores do que os da idade. As crianças com IMC elevado somente apresentaram médias mais elevadas de pressão diastólica. Todas as crianças com cirrose Child-Pugh B e C apresentaram deficiência de uma ou mais proteínas inibidoras em 39 crianças (76,5%). Entre as crianças que apresentaram alguma sintomatologia, relatou-se, mais As crianças que nasceram de cesariana também apresentaram uma maior chance de ter cólica (RO 1,55 - IC 26 crianças avaliadas quanto às subclasses de IgG apresentaram deficiência de IgG2. As dosagens de subclasses crianças que fizeram uso de surfactante exógeno apresentaram, durante o primeiro ano de vida,</p>	P
Assoc. pragmática: motivada pelo gênero/seção em que se encontra		
Características/gerais/grupos		
1 2 3 4	<p>Pediátrica da UFRN Na Tabela 2, apresentamos as características gerais dos grupos estudados. O IMC diferiu Não foram demonstradas diferenças entre as características gerais dos animais dos dois grupos (aspect em relação ao grupo controle. Tabela 2 - Características gerais das crianças e adolescentes dos gru de coloração alaranjada ou vermelha (n = 1). As características gerais e os índices antropométricos dos gr</p>	P
Assoc. pragmática: motivada pelo gênero/seção em que se encontra		
curva/recém/apresentam/semana		
1 2 3 4	<p>até a 12a semana. Os recém-nascidos maiores (curva 6 e 7) apresentam, durante todo o período de semana na 12a semana. Os recém-nascidos maiores (curva 5 e 6) apresentam um discreto aumento na taxa de de crescimento. Assim, os recém-nascidos da curva 1, que, na primeira semana, apresentam taxa de de ganho de comprimento. Os recém-nascidos da curva 5, na primeira semana, apresentam valores de 0,71</p>	P
Assoc. pragmática: motivada pelo gênero/seção em que se encontra		
Feminino/masculino/sexo		
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36	<p>do sexo masculino e 44,60% (356 798) do sexo feminino. Foi encontrado um maior percentual de crianças do sexo masculino (76,5 %), e quatro do sexo feminino (23,5 %). A mediana de idade foi 4,3 anos, com do sexo masculino e 3.139 g para as do sexo feminino. Para que nossa avaliação não fosse errônea, eram do sexo masculino e 79 (47,3%) eram do sexo feminino. A idade variou de 38 a 364 dias, com média de do sexo feminino, sendo a relação masculino feminino de 1,1:1. Houve um predomínio de crianças de pele eram do sexo masculino, e 53,5% (122 228), do feminino. Quanto à procedência, 43,4% (99 228) da amostra eram do sexo masculino e 44,12% (45) do sexo feminino. Ao compararmos as médias de peso segundo o sexo, 203 do sexo masculino (44,9%) e 249 do sexo feminino (55,1%). Do total de meninos examinados, 85 eram do sexo masculino e 156 (47,7%) do sexo feminino, sendo a relação masculino feminino de 1,1:1. eram do sexo masculino e 920 (50,9%) do sexo feminino, com idade entre 7 e 19 anos, média de idade (n = 385) do sexo masculino e 48,1% (n = 357) do feminino. Como mostra a Tabela 1, a faixa etária (53,7%) do sexo masculino e 19 (46,3%) do sexo feminino. A média do escore total do NIPPSI-R encontrado (36,8%) do sexo masculino e 12 (63,2%) do sexo feminino, todos acima de 10 anos de idade. De acordo com a 9,5% (52) no sexo masculino e 9,3% (66) no sexo feminino, todos acima de 10 anos de idade. De acordo com a 15 seis eram do sexo masculino, e quatro do sexo feminino. A idade variou entre 5,4 anos e 10,3 anos (média 16 47,7% eram do sexo masculino e 52,3% do sexo feminino. A idade do conjunto variou de 2,1 a 19,6 anos, 17 21 eram do sexo masculino e 19 do sexo feminino. Vinte e seis recém-nascidos foram classificados 18 19 pacientes do sexo masculino e 21 do sexo feminino. Foram colhidas 380 amostras desse grupo durante 19 sendo 20 crianças do sexo masculino e 20 do sexo feminino. Foram coletadas 386 amostras fecais no grupo 20 (17,3%), 10 do sexo masculino e oito do sexo feminino, 17 caucasóides e um negróide, todos por 21 34 pacientes (68%) do sexo masculino e 16 do feminino (32%); 48 eram da raça branca (96%) e dois da 22 de 13-14 anos, segundo o sexo masculino (M), feminino (F) e número total (T), e respostas afirmativas 23 de 6-7 anos, segundo o sexo masculino (M), feminino (F) e número total (T), e respostas afirmativas 24 entre recém-nascidos do sexo masculino e feminino (Tabela 1). A leptina do cordão umbilical se 25 diferença na incidência em relação ao sexo feminino e no masculino. O encaminhamento das meninas para 26 e 5 meses e 4,8 e 6 meses, em geral, no sexo feminino e 22,4 % do masculino, e idade menor ou igual a 27 pacientes Observou-se 77,6 % de crianças do sexo feminino e 22,4 % do masculino, e idade menor ou igual a 28 constitui-se de 74 pacientes, 38 do sexo feminino e 36 do sexo masculino com idades entre 6 meses e 29 de parecer mais magros, 69% pertencem ao sexo feminino e 31% ao sexo masculino (p = 0,000). Entre os 30 de 5 a 17 anos. 45 (49,5%) pertenciam ao sexo feminino e 46 (50,5%) ao masculino (p > 0,882). A 31 de idade, sendo 52% (366) das crianças do sexo feminino, e 48% (333) do sexo masculino. Levando em conta 32 Foram avaliados 528 escolares, sendo 273 do sexo feminino (51,7%) e 255 do sexo masculino (48,3%). Esse 33 45 crianças incluídas no estudo, 17 eram do sexo feminino (37,8%), e 28 do sexo masculino (62,2%), com 34 g. Existiam 10 prematuros no grupo do sexo feminino e 12 prematuros no grupo masculino (p = 0,728). 35 com colostomia crônica, sendo 13 do sexo feminino (59,1 %), e nove do sexo masculino (40,9 %). A 36 observada uma maior proporção de alunos do sexo feminino (58,2 versus 41,8% do sexo masculino). O maior</p>	S

Assoc. semântica: motivada pelo contexto de descrição de gênero sexual

4.2.4. Seção *Discussão*

Quadro 30 – Concgrams de palavras-nodo da seção Resultados

achados/et/al		
1	ligeiramente no período pós-termo. Esses achados foram igualmente relatados por Einspieler et	P
2	da FC. Patriquin et al. (11) compararam os achados ultra-sonográficos com as alterações da	
3	(35,7%) e feminino (50,0%), corroborando os achados de Valverde et al. (5), que, investigando 74	
4	deposição pulmonar superior aos demais. Esses achados estão de acordo com Tal et al., que estudaram	
5	onde concentram-se os mais pobres. Esses achados estão de acordo com os de Chew et al. (28),	
6	relação à deficiência da subclasse IgG2, nossos achados foram similares aos de Avanzini et al. (09),	
7	com MG anormais apresentam maior índice de achados anormais na USC no período neonatal Bos et al.	
8	observando-se relação apenas com o PN. Estes achados estão de acordo com o estudo de Losonsky et	
Assoc. pragmática: motivada pelo gênero/seção em que se encontra		
achados/acordo/estão		
1	inoculação de urina na cavidade peritoneal. Tais achados estão de acordo com algumas hipóteses já	P
2	observando-se relação apenas com o PN. Estes achados estão de acordo com o estudo de Losonsky et	
3	deposição pulmonar superior aos demais. Esses achados estão de acordo com Tal et al., que estudaram	
4	onde concentram-se os mais pobres. Esses achados estão de acordo com os de Chew et al. (28),	
5	e os perímetros da cintura e do quadril. Esses achados estão de pleno acordo com outras observações	
Assoc. pragmática: motivada pelo gênero/seção em que se encontra		
associação/entre/uso		
1	precoce. Victora et al. (7) observaram que a associação entre uso de chupeta e duração do aleitamento	P
2	O delineamento do estudo não permite avaliar a associação entre uso de casca de banana e infecção. Novo	
3	bem respondida mediante estudos etnográficos. A associação entre uso de chupeta e menor duração do	
4	dados de estudo etnográfico, concluíram que a associação entre uso de chupeta e desmame precoce é comp	
5	também não mostraram efeito modificador na associação entre uso de chupeta e desmame precoce. Embor	
6	precoce. Embora já esteja bem estabelecida a associação entre uso de chupeta e menor duração do	
7	689 estudantes de segundo grau, demonstrou uma associação entre uso de álcool e maconha e comportamento	
8	adulto (22). Na população estudada, notou-se uma associação entre o uso de chupeta e o desmame precoce.	
9	foi demonstrado em alguns artigos que relatam a associação entre o uso de células progenitoras	
10	a possibilidade da oferta do corticosteróide. A associação entre o uso de tocolíticos e do corticosteróide	
11	tratamento de fissuras mamilares. No entanto, a associação entre alergia na mama e o uso de casca de ban	
12	precoce. Diversos estudos têm mostrado uma associação inversa entre o uso de chupeta e o aleitamento	
13	o mal, indicando desconhecimento da possível associação negativa entre uso de chupeta e aleitamento	
14	O presente estudo também mostrou uma associação mais forte entre uso de chupeta e desmame pre	
15	para evitar o seu uso, tendo sido confirmada a associação entre o hábito de usar chupeta e o desmame	
Assoc. pragmática: motivada pelo gênero/seção em que se encontra		
autores/como/outros		
1	como de vegetais (19). À semelhança de outros autores, a prevalência de sintomas relacionados ao E	P
2	a infecções, como já destacado por outros autores (20-22). O envolvimento hepático tem sido	
3	nesta série (18), como já relatado por outros autores (12,20,29). A doença viral prévia esteve pre	
4	inicial. Assim como descrito por outros autores (1,3,4,7,8), também não observamos diabetes	
5	também de ascaridíase, como descrito por outros autores (6,7). Como doença multifatorial, a asma é	
6	significativa. Entendemos, assim como outros autores, que as pequenas diferenças observadas podem	
7	Crowley (2) não aponta para maior risco, outros autores sugerem que, como demonstrado aqui, o risco	
8	com espaçador foram apontadas por outros autores, porém não investigadas neste estudo, como:	
Assoc. pragmática: motivada pelo gênero/seção em que se encontra		

4.3. Discussão sobre os resultados da Etapa 2

Primeiramente, queremos relatar a dificuldade que tivemos para encontrar uma maneira coerente e consistente de fazer essa "rotulagem" da força que motiva as associações. É preciso dizer que, mesmo com a grande disponibilidade de autores que tratam do tema de associações, de exemplos e explicações, quando se trata de aplicar a teoria em casos reais, encontramos dificuldades e as dúvidas surgem. Tomemos por exemplo os exemplos usados por Hoey (2007) para diferenciar associação semântica de pragmática. O autor diz que a expressão *a word against* tem uma associação semântica com *envio e recebimento de informação* e *send/receive a word against* tem uma associação pragmática com negação.

Em nosso entendimento, considerando que prosódia semântica pode incluir outros aspectos que não apenas negação e afirmação, negação seria um item da prosódia semântica da associação, desempenhando também uma função pragmática; Garretson, que revisa o livro de Michael Hoey (2007) afirma que associação pragmática, na referida teoria, é sinônimo de preferência semântica. Ainda dentro de estudos de padrões, na tradição britânica, temos outra definição de Hunston and Francis (Hunston & Francis, 2000, p.137): "pode-se dizer que uma palavra tem uma prosódia semântica particular se ela demonstrar coocorrer tipicamente com outras palavras que pertencem a um mesmo grupo semântico específico". Tendo em vista esses e outros pontos fluidos na definição do que é uma associação semântica e pragmática, optamos por aplicar um critério para esta tese do que seria uma associação motivada por critérios semânticos ou pragmáticos.

Conforme explicamos acima, pensamos primeiramente em aplicar o "rótulo" de associação semântica ou pragmática com base em uma predominância de um sobre o outro, não que sejam excludentes. Além disso, pensamos em adotar o critério de responder a pergunta: "Essas palavras estariam associadas se não estivessem em um artigo científico de Pediatria?". Se a resposta fosse "sim", seria uma indicação de que estão associadas por motivos semânticos, isto é, é uma associação que ocorre também em outros ambientes, portanto é mais motivada por critérios semânticos. Se a resposta fosse "não", seria porque o gênero as estaria determinando, e seria o motivador mais influente. Por vezes também utilizamos o Google para confirmar nossa intuição e, em muitos casos, o resultado das buscas surpreendeu, demonstrando que algumas associações que parecem semânticas são muito mais determinadas pelo gênero do que parecem. É interessante também lembrar que Hoey (2007) afirma que quem ativa o *priming* é o falante, não a palavra, e os córpus podem apenas indiretamente nos fornecer quais são os prováveis *primings* que as pessoas ativam.

A visão que temos do que pertence ao domínio da Medicina também torna difícil determinar que, por exemplo, a associação *análise/estatística/foi* é motivada mais por motivos pragmáticos do que semânticos, uma vez que em termos de texto, *estatística* é também uma palavra do domínio da Medicina quando inserida em um todo que é um artigo médico. No entanto, em uma rápida pesquisa no Google, os resultados apontam que essa é uma associação que não ocorre muito em outros ambientes textuais que não o de artigos científicos. Assim também com *crianças/apresentaram* que, embora pareça muito pouco especializada, dos dez primeiros resultados do Google, nove são de artigos acadêmicos.

Nos parece que essas duas forças, semântica e pragmática, vão atuando sempre em conjunto e, em muitos casos, não há como dissociá-las. Há casos também que temos coligações, ou associações gramaticais, como quando temos um verbo de ligação: *análise/estatística/foi*. Como em todos os casos analisados tínhamos associações de pelo menos duas palavras não-gramaticais, não evidenciamos as coligações.

Concluimos, assim, que as associações sintagmáticas que observamos são, sim, motivadas por critérios semânticos e pragmáticos, mas não separadamente, como acreditávamos a princípio, mas sim em conjunto. No momento que temos uma associação como *teste/t/Student* como uma associação que ocorre em contextos de estudos estatísticos, não necessariamente artigos de Medicina, mas ao mesmo tempo também ela se configura como uma associação em nosso corpus, também entendemos que ela pertence a esse ambiente, a esse todo do texto do artigo médico, sendo, no contexto em que se encontra, formadora do texto especializado em que se encontra.

Isso dá ainda mais evidências de que, em um texto especializado, há várias forças operando: do gênero, do domínio, da gramática da língua, do princípio idiomático, e que é preciso cada vez mais olhar para tais particularidades para ter um entendimento do que lhes é característico e peculiar.

5. Etapa 3: Associações como particularidades do texto especializado

O objetivo da investigação desta etapa é mostrar como as associações entre palavras exibem um perfil diferente em textos especializados em comparação com textos da linguagem cotidiana, não-científica, em termos de modos e de frequência de repetições.

Nosso intuito é demonstrar como, nos textos especializados, a frequência de ocorrência de agrupamentos de palavras é maior, construindo, assim, a linguagem de determinada área de especialidade e imprimindo marcas ao seu texto. Desse modo, esperamos, abre-se uma perspectiva de análise de textos especializados que ultrapassa termos ou fraseologias especializadas *stricto sensu*.

5.1. Metodologia

Usamos aqui, como palavra-nodo, os substantivos mais frequentes de cada um dos *córpus*: *criança* (Jornal de Pediatria), *anos* (Lácio) e *ácido* (Química). Buscamos os *concgrams* de cada palavra-nodo em cada um dos *córpus*. Assim foi possível comparar e verificar se haveria alguma relação entre a palavra mais frequente de cada *córpus* e a formação ou não de associações em torno dela. É importante ressaltar que, nesta etapa, usamos o *córpus* do Jornal de Pediatria completo, não segmentado nas seções *Introduções*, *Métodos*, *Resultados* e *Discussão*. Para manter uma consistência com as etapas anteriores, usamos os *concgrams* de 3 palavras também nesta etapa.

A seguir, temos um quadro com as frequências absolutas e relativas de ocorrência no *córpus* dos substantivos mais frequentes de cada um. Os sublinhados indicam os mais frequentes de cada *córpus*.

Quadro 31 – Frequências absolutas e relativas de ocorrência no *córpus* dos substantivos mais frequentes

	JPed		Lácio		Química	
	FA	FR	FA	FR	FA	FR
crianças	<u>1.959</u>	<u>56,15</u>	778	4,62	258	2,04
anos	906	25,97	<u>3047</u>	<u>18,10</u>	283	2,23
ácido	29	0,83	35	0,21	<u>642</u>	<u>15,80</u>

Com o processamento dos textos com o software *concgram* para obtermos as combinações para cada palavra-nodo, objetivamos comparar a frequência de ocorrência do primeiro *concgram* de cada *córpus* e a frequência do número de *concgrams* com recorrência até 2 em cada *córpus*.

Visto que as listas completas dos resultados são muito extensas (mais de 200 páginas para cada palavra-nodo), apresentamos aqui apenas os resultados comparativos.

5.2. Resultados e discussão

O primeiro dado quantitativo que chama a atenção na análise comparativa das frequências dos *concgrams* é a diferença na frequência nos *concgrams* de 3 palavras normalizada para 10 mil palavras entre os *córpus*, utilizando-se o substantivo mais frequente do *córpus* do Jornal de Pediatria: *crianças*. O que vemos é que, para o *córpus* de artigos médicos, o primeiro agrupamento tem uma frequência de 21,64 palavras para cada 10 mil, contra 1,70 do *córpus* geral, e 0,29 do *córpus* de contraste, como sintetiza o quadro 32.

Quadro 32 – Frequência relativa e absoluta dos *concgrams* de 3 palavras, utilizando a palavra-nodo *crianças* sem o uso de uma *stoplist*

Córpus	Concgrams de 3 palavras	Frequência relativa	Frequência absoluta
Córpus especializado (Jornal de Pediatria)	crianças a de	21,64	755
Córpus geral (Lácio)	crianças a de	1,70	286
Córpus de contraste (Química Nova)	crianças a de	0,29	12

Quadro 33 – frequência relativa e absoluta dos *concgrams* de 3 palavras utilizando a palavra-nodo *crianças* com o uso de uma *stoplist*

Córpus	Concgrams de 3 palavras	Frequência relativa	Frequência absoluta
Córpus especializado (Jornal de Pediatria)	crianças com idade	2,32	81
Córpus geral (Lácio)	crianças anos com	0,10	17
Córpus de contraste (Química Nova)	crianças adultos chumbo	0,04	2

Esse primeiro dado quantitativo demonstra que, no texto especializado, a recorrência do *concgram* mais frequente que tem como palavra-nodo *crianças* é, evidentemente, maior que no *córpus* geral e no de contraste.

Podemos fazer duas leituras desse dado. A primeira leitura é que *crianças* não é normalmente reconhecida como um termo técnica em Medicina ou Pediatria, mas como uma palavra da “língua geral” e, sendo assim, em um *córpus* de notícias e linguagem cotidiana, a

presença dessa palavra, formando associações, surpreende por ter sido tão discrepante do *córpus* especializado. Perguntamos, então, que palavra é essa, *criança*, que é tão frequente no *córpus* especializado mas que, se olhada fora do contexto, é um substantivo mais do que comum. Nesse ponto, pensamos que a reflexão sobre a construção de um sentido-funcionamento especializado deve ser feita.

A segunda leitura que fazemos desse dado é o questionamento sobre o fato de *crianças* ser a palavra mais frequente do *córpus* de Pediatria, mas não dos outros. Por isso, comparamos o *concgram* mais frequente composto pelas palavras de maior frequência dos outros dois *córpus*: *anos*, no *córpus* geral, e *ácido*, no *córpus* de Química, conforme mostram os quadros a seguir.

Quadro 34 – Concgrams mais frequentes das palavras mais frequentes de cada *córpus* sem uma stoplist ativada

Córpus	Concgrams de 3 palavras	Frequência no <i>córpus</i>	Frequência relativa
Córpus especializado Pediatria (crianças)	Crianças em e	219	6,28
Córpus geral Lácio (anos)	anos a de	1199	7,12
Córpus de contraste Química (ácido)	ácido a de	263	6,47

Quadro 35 – Concgrams mais frequentes das palavras mais frequentes de cada *córpus* com uma stoplist ativada

Córpus	Concgrams de 3 palavras	Frequência no <i>córpus</i>	Frequência relativa
Córpus especializado Pediatria (crianças)	crianças com idade	81	2,32
Córpus geral Lácio (anos)	anos há mais	76	0,45
Córpus de contraste Química (ácido)	ácido nítrico v	42	1,0

Notamos, então, uma diferença importante nos resultados com uma *stoplist* ativada e naqueles sem essa ativação. No Quadro 34, em que não foram excluídas certas palavras gramaticais, notamos que, em termos de frequência relativa, o *córpus* geral tem o maior número de ocorrências. No Quadro 35, em que foram excluídas essas palavras, a situação é invertida, e o *córpus* geral é o que tem a menor frequência de associações.

Esses dados mostram que os textos especializados são mais densos, contendo mais palavras "cheias", enquanto os textos chamados de linguagem cotidiana, nesse caso os textos jornalísticos, associam-se mais com palavras gramaticais. Diante desse dado, pouco referido na bibliografia conforme sabemos, colhe-se mais uma evidência de que as associações entre as palavras são também um ponto importante de caracterização e diferenciação dos textos especializados. Evidentemente que, para obter mais informações a esse respeito, seria interessante consultar também outros *córpus* de textos especializados.

Seguindo a observação quantitativa, a lista de *concgrams* de 3 palavras do *córpus* especializado, com a *stoplist* ativada, contém 6.000 *concgrams*, que vão de *concgrams* com frequência de ocorrência de 81 até 2 vezes no *córpus*; a lista de *concgrams* do *córpus* geral contém 8.287 *concgrams*, com frequência de ocorrência de 76 a 2 vezes no *córpus*; e o *córpus* de contraste contém 2.128 *concgrams*, que têm frequência de ocorrência de 42 a 2 vezes no *córpus*. Normalizando-se esses dados para 10 mil palavras, novamente se percebe que o texto especializado apresenta um potencial maior de conter grupos de palavras que se associam e de associações que se repetem.

Quadro 36 – Número de *concgrams* de cada *córpus*, com a *stoplist* ativada

Córpus	Número de <i>concgrams</i> de 3 palavras com freq. até 2	Frequência Relativa
Córpus especializado Pediatria (crianças)	6.000	171,99
Córpus geral Lácio (anos)	8.287	49,22
Córpus de contraste Química (ácido)	2.128	52,36

Em nosso entendimento, novamente a frequência mais alta no número de *concgrams* dos *córpus* especializados demonstra que esses textos têm maior probabilidade de conterem

palavras que se associam do modo como desenhamos. As associações refletem, reiteramos, como as áreas de especialidade constroem a sua linguagem, os modos como apresentam seu conteúdo. Esses “modos de dizer” podem ser mais ou menos fixos e ajudam a delimitar e definir o escopo dos significados das palavras. Sua apreensão, entretanto, vai além da junção entre termos e outros elementos ou de ocorrência contígua. Por isso, nosso pressuposto de que, quando um texto é especializado, uma união de fatores opera para que ele exista como tal. Entre os fatores em ação está a união sintagmática que algumas palavras estabelecem com outras. Assim, mesmo sem conter um termo em sua constituição, elas adquirem um status de especialidade (Finatto e Possamai, 2006).

Entendemos que o fato das palavras se atualizarem como termos em certos contextos não é uma novidade. Mas o que nos parece novo, a partir dos dados aqui investigados, é pensar, não no termo, ou no que faz um termo ser um termo, no que é ou seria mais “terminológico” em meio aos artigos de Pediatria, mas, sim, no que faz um texto especializado constituir-se como tal.

E, o que desejamos iluminar são justamente outros recursos e características do texto especializado, não como secundários, coadjuvantes ou “meramente discursivos”, mas como peças determinantes na instauração do fator “especialidade” de um texto. Entendemos que o texto especializado só passa a existir como tal no momento que algumas associações entre determinados itens ocorre, independentemente de serem palavras mais ou menos portadoras de conceitos (termos), palavras gramaticais, adjetivos, verbos etc. Assim, podemos afirmar, amparados nos dados dos *cópus* observados, que não é a categoria da palavra (termo, não-termo) que contribuir para desenhar a especialidade, mas o fato de ela estar associada a alguns itens e não a outros.

Em função do acima exposto, sublinhamos a já antiga proposta de estudo do texto especializado realizada por Hoffmann (2004), que traz um enfoque do texto e de seus componentes. Segundo esse autor, o caráter de especialidade de um texto não se deve apenas ao critério de tema (área científica ou técnica), mas por “(...) particularidades de sua macroestrutura (articulação), por relações de coerência entre seus elementos e pela utilização de unidades sintáticas, lexicais, morfológicas e gráfico-fonéticas” (p. 81).

6. Etapa 4: As associações sintagmáticas como unidades de tradução especializadas

Nossa hipótese é a de que as associações sintagmáticas com um determinado perfil (conforme etapas 1 e 2 acima) se configuram como unidades de tradução especializadas. Nas etapas anteriores, tratamos de demonstrar como palavras que, aparentemente, não formam uma fraseologia especializada de caráter mais fixo ou um sintagma especializado, podem ser frequentes e se associar, motivadas por uma “força” predominantemente semântica ou pragmática. Esta última etapa de investigação demonstra como essas associações foram traduzidas. Com isso, pretendemos ilustrar por que elas podem ser tratadas como unidades de tradução.

6.1. Metodologia

Usaremos uma associação identificada na Etapa 3 para cada seção dos artigos e buscaremos suas traduções no cópulus de textos traduzidos de Pediatria. Investigando as traduções, pretendemos chamar a atenção para as associações e para como os tradutores lidam com elas. Os resultados são apresentados em forma de quadros, que mostram quais as palavras que se associam, os contextos das associações conforme resultado do *ConcGram*. Em seguida, são mostrados os contextos ampliados em português e a tradução correspondente. **Não** faremos julgamento das traduções e nem procuraremos dar sugestões “mais adequadas” de tradução, pois justamente estamos tratando da problemática de traduzi-las. Assim, acreditamos que estamos respeitando a dedicação dos profissionais que as fizeram em situações que não conhecemos para poder julgar.

6.2. Resultados e discussão

Quadro 37 – Linhas de concordância e traduções de concgrams da Introdução

amamentação/duração	
<p>1 positiva ou negativamente o início da amamentação e sua duração. Em particular, as enfermeiras e os 2 aquém das recomendações da OMS, com mediana de amamentação exclusiva de apenas um mês, duração total de 3 de 4 a 6 meses, e a mediana de duração da amamentação exclusiva foi de apenas 1 mês (5,6). Assim, fica 4 profissionais, ocasionando maior duração da amamentação nas comunidades por eles assistidas (7,8). Este 5 exclusiva de apenas um mês, duração total de amamentação de sete meses, e o aleitamento continuado[um ano</p>	
<i>Texto fonte - Português</i>	<i>Texto alvo - Inglês</i>
<p>1. Os profissionais de saúde, por meio de suas atitudes e práticas, podem influenciar positiva ou negativamente o início da amamentação e sua duração.</p>	<p>Health professionals, by means of their attitudes and practices, may positively or negatively influence the beginning of breastfeeding and its length.</p>
<p>3. e a mediana de duração da amamentação exclusiva foi de apenas 1 mês</p>	<p>and that the median of exclusive breastfeeding duration was only 1 month</p>

4. ocasionando maior duração da amamentação nas comunidades por eles assistidas	resulting in longer duration breastfeeding in the communities they serve
5. com mediana de amamentação exclusiva de apenas um mês, duração total de amamentação de sete meses, e o aleitamento continuado um ano em 41,0% das crianças	with the median for exclusive breastfeeding at just one month, total breastfeeding at seven months and breastfeeding continuing for one year for 41.0 % of crianças

Examinando as traduções para essa associação, percebemos que *amamentação* não pareceu causar maiores problemas aos tradutores, uma vez que não há maiores controvérsias com relação a essa palavra-termo e sua opção de equivalência. Já, a associação *duração* e *amamentação* mostra-se mais complexa para a tradução. Nenhuma das traduções optou pelo mesmo “equivalente”. Os tradutores escolheram dois termos para traduzir *duração* (*length* e *duration*) e, na linha 5, não há um equivalente direto para essa palavra. Além disso, quando optaram pela palavra *duration*, duas posições foram escolhidas: *breastfeeding duration* e *longer duration breastfeeding*.

Quadro 38 – Linhas de concordância e traduções de concgrams dos Métodos

Análise/estatística/realizada	
1 CTC NCI) (9). Análise estatística Foi realizada análise descritiva, utilizando-se a média, o desvio 2 para a discriminação dos grupos. Toda a análise estatística foi realizada através do SPSS(x) 3 do paciente não foi objeto deste trabalho. A análise estatística foi realizada utilizando-se o 4 National Health Institute CTC NCI) (9). Análise estatística Foi realizada análise descritiva, 5 para coeficiente de correlação (r) de 0,7. A análise estatística foi realizada através dos programas 6 utilizou-se o teste exato de Fisher. Toda a análise estatística foi realizada utilizando o software 7 do paciente não foi objeto deste trabalho. A análise estatística foi realizada utilizando-se o 8 resultados normais para a faixa etária. A análise estatística foi realizada através do teste de 9 Esse escore foi utilizado para processamento da análise estatística , que foi realizada através do teste U 10 múltiplas sem diagnóstico síndrome). A análise estatística dos dados foi realizada no programa 11 com o padrão de peso molecular utilizado. A análise estatística dos resultados foi realizada por meio	
Texto fonte - Português	Texto alvo - Inglês
2. Toda a análise estatística foi realizada através do SPSS® (versão 8.0).	All the statistical analysis was carried out through SPSS® (version 8.0)
3. A análise estatística foi realizada utilizando-se o programa Epi Info 6.04 do CDC (<i>Centers for Disease Control and Prevention</i>)	The statistical analysis was made using Epi Info 6.04 of CDC (<i>Centers for Disease Control & Prevention</i>)
5. A análise estatística foi realizada através dos programas SPSS (<i>Statistical Package for Social Sciences</i>), versão 6.0, e Epi-Info 6, versão 6.04d	Statistical analysis was performed using SPSS (<i>Statistical Package for Social Sciences</i>), version 6.0, and EpiInfo 6, version 6.04d
6. Toda a análise estatística foi realizada utilizando o software SAS, versão 8.2.	All statistical analysis was performed using the software SAS, version 8.2.
8. A análise estatística foi realizada através do teste de McNemar para avaliação da concordância entre a qualidade e os tipos de anormalidade dos MG observados	Statistical analysis was performed using the McNemar test in order to study the concordance between the quality and the abnormality types of the GM observed
9. Esse escore foi utilizado para processamento da análise estatística , que foi realizada através	The scores were used for the statistical analysis that was performed using the "U" test (Wilcoxon -

do teste U (Wilcoxon-Mann-Whitney).	Mann - Whitney test).
10. A análise estatística dos dados foi realizada no programa Epi-Info 6.0 (17), com utilização dos testes do qui-quadrado e exato de Fisher, de acordo com a indicação.	The statistical analysis of data was performed with the Epi-Info 6.0 program (17) with the CHI-squared and Fisher exact tests being used where indicated.
11. A análise estatística dos resultados foi realizada por meio do teste exato de Fisher, sendo fixado em 5% (alfa igual a 0,05) o nível de rejeição da hipótese de nulidade, com intervalo de confiança de 95%.	Statistical analysis was performed using Fisher's exact test, with the level of rejection of the null hypothesis set at 5% (alpha error 0.05), with a 95% confidence interval

Novamente, não houve diferenças para a tradução do termo *análise estatística*. Em se tratando de variações nas escolhas de equivalentes, notamos que predominou a escolha por *perform* como tradução para *realizar*, mas foram usados também os verbos *make* e *carry out*. É interessante notar que, quando o texto fonte é *análise estatística dos resultados*, na linha 11, o tradutor optou por não traduzir *resultados*. A tradução da linha 2 usa equivalentes bem diferentes das outras sete: enquanto estas usam basicamente a forma *performed using* (tradução literal: realizada usando), aquela usa *carried out through* (tradução literal: conduzida através). Percebe-se também uma terceira opção, *performed with* (tradução literal: realizada com).

Quadro 39 – Linhas de concordância e traduções de concgrams dos Resultados

recém/curva/apresentam	
1 até a 12a semana. Os recém-nascidos maiores (curva 6 e 7) apresentam, durante todo o período de semana na 12a semana. Os recém-nascidos maiores (curva 5 e 6) apresentam um discreto aumento na taxa de de crescimento. Assim, os recém-nascidos da curva 1, que, na primeira semana, apresentam taxa de de ganho de comprimento. Os recém-nascidos da curva 5, na primeira semana, apresentam valores de 0,71	
<i>Texto fonte - Português</i>	<i>Texto alvo - Inglês</i>
1. Os recém-nascidos maiores (curva 6 e 7) apresentam , durante todo o período de estudo, uma velocidade constante de crescimento de perímetro cefálico, com taxas de 0,62 e 0,57 cm/semana, respectivamente.	The larger newborns (curves 6 and 7) exhibit a constant head circumference growth velocity, with rates of 0.62 and 0.57 cm/week, respectively.
2. Os recém-nascidos maiores (curva 5 e 6) apresentam um discreto aumento na taxa de ganho de comprimento.	The larger babies (curve 5 and 6) present a discrete increase in length gain rate.
3. Assim, os recém-nascidos da curva 1, que, na primeira semana, apresentam taxa de 0,04 cm/semana, aumentam gradativamente essa taxa e alcançam 1 cm/semana na 12a semana,	Thus, the newborn babies on curve 1 who, in the first week, presented a rate of 0.04 cm/week, gradually increased this rate and reached 1 cm/week at the 12th week
4. Os recém-nascidos da curva 5, na primeira semana, apresentam valores de 0,71 cm/semana, aumentando para 0,99 cm/semana	Those newborns on 5, exhibited first-week values of 0.71 cm/week, increasing to 0.99 cm/week by the 12th week

na 12a semana;	
----------------	--

Nesse exemplo, primeiramente temos três possibilidades de tradução para *recém-nascidos*: *newborns*, *babies* e *newborn babies*, as quais o tradutor deve avaliar para fazer sua escolha. No que concerne à *curva*, podemos ver que, na informação apresentada entre parênteses, um dos profissionais usou o plural e o outro não, e, na frase 4, *curve* foi eliminado. O verbo, que em português é *apresentam*, foi traduzido por *present* e *exhibit*. Essas opções de tradução mostram como podemos ter, para uma mesma associação em português, várias possibilidades em inglês. Ainda sem considerar se todas são apropriadas, entendemos que a variação evidencia diferentes possibilidades de tradução para uma mesma associação.

Quadro 40 – Linhas de concordância e traduções de concgrams autores/outras/como

Autores/outras/como	
1 como de vegetais (19). À semelhança de outros autores , a prevalência de sintomas relacionados ao EA foi 2 a infecções, como já destacado por outros autores (20-22). O envolvimento hepático tem sido 3 nesta série (18), como já relatado por outros autores (12,20,29). A doença viral prévia esteve presente 4 inicial. Assim como descrito por outros autores (1,3,4,7,8), também não observamos diabetes 5 também de ascaridíase, como descrito por outros autores (6,7). Como doença multifatorial, a asma é 6 significativa. Entendemos, assim como outros autores , que as pequenas diferenças observadas podem ser 7 Crowley (2) não aponta para maior risco, outros autores sugerem que, como demonstrado aqui, o risco de 8 com espaçador foram apontadas por outros autores , porém não investigadas neste estudo, como :	
Texto fonte - Português	Texto alvo - Inglês
2. Nesta casuística, vários fatores, isolados ou associados, merecendo destaque os valores médios de hemoglobinas, leucócitos e neutrófilos no início do diagnóstico, poderiam estar associados a infecções, como já destacado por outros autores (20-22)	As has been explained by other authors (20-22), in this sample, a number of different factors, in association or isolation, may have been associated with infection. Of these factors, average hemoglobin, leukocyte and neutrophil levels deserve special attention.
3. O sexo não influenciou na mortalidade nesta série (18), como já relatado por outros autores (12,20,29).	Sex did not influence mortality in this series (18), which is in agreement with other authors (12,20,29).
4. Assim como descrito por outros autores (1,3,4,7,8), também não observamos diabetes insípido nos nossos pacientes que apresentavam ectopia da neuro-hipófise.	In common with descriptions by other authors (1,3,4,7,8), we did not observe diabetes insipidus among our patients with ectopic posterior pituitaries.
5. Nestas condições, seria até surpreendente a ocorrência maior de casos de asma entre aqueles portadores também de ascaridíase, como descrito por outros autores (6,7).	Under these conditions, it would even be surprising to observe greater occurrence of asthma cases amongst those who are also suffering from ascariasis, as described by other authors (6,7).
6. Entendemos, assim como outros autores , que as pequenas diferenças observadas podem ser explicadas facilmente pela variabilidade das amostras (16).	We understand, and so do other authors , that the small differences observed might be easily explained by the variability of the samples (16).

A associação *como... por outros autores* foi traduzida de diferentes maneiras. Como podemos perceber, foi traduzida por *as... other authors, in agreement...other authors, in common with... other authors* e *and so do other authors*. Mesmo desconsiderando os verbos que motivam essas alterações, pensamos que o fato de elas serem traduzidas de maneira diferente também as justifica como unidades, pois mostra que, em cada texto, se escolheu formas diferentes de reconstruí-las. O tradutor, conhecendo-as, pode ampliar seu repertório de escolhas quando se depara com essa associação. Assim, essa variabilidade de opções para essas construções, conforme entendemos, revela o seu caráter de unidades de tradução especializadas, uma vez que para fazer escolhas o tradutor precisará refletir sobre elas, pesquisá-las, certificar-se de seu uso.

6.3. Discussão sobre os achados

Essas associações podem ser unidades de tradução porque, como vimos, são itens recorrentes, colaboram para a naturalidade do texto por meio do encadeamento sintagmático do texto e constroem “modos de dizer” de determinadas especialidades e gêneros. Não há como negar que o tradutor precisa trabalhar todos esses elementos de um texto, entre outros e, portanto, é uma vantagem saber que existem, saber quais são e saber o risco de não conhecê-las.

Além disso, a identificação de palavras coocorrentes como essas, em um sistema de apoio ao tradutor, representará um ganho quando usado o modelo de memórias de tradução. Isso demonstraremos mais adiante na parte que descreve nossa aplicação prática.

Enquanto um tradutor experiente já sabe reconhecer essas associações que se repetem, os menos experientes ainda não têm como as identificar. Ao aprender como lidar com elas, o tradutor ganha conhecimento na área e, com isso, rapidez e segurança em suas traduções. É por esses motivos que, ao fim e ao cabo, achamos que essas associações, quando são identificadas automaticamente para o tradutor, podem ajudá-lo a ter melhor domínio do texto médico. Por isso, desenvolvemos a ferramenta que será apresentada na Parte II desta tese.

7. Considerações sobre este capítulo

O estudo dos *concgrams* permitiu fazer uma série de observações importantes sobre a hipótese que nos propusemos a investigar, delineada no início deste Capítulo⁵⁶.

Em quatro diferentes etapas procuramos isolar aspectos diferentes de nossas unidades de análise. Na primeira etapa, demonstramos como as associações poderiam ter variações posicionais e constitutivas, além de outras regularidades nos padrões de associações, como colocações e preferências semânticas. Na segunda etapa, procuramos demonstrar que “força” motiva as associações e enfocamos as duas forças que consideramos mais fortes, a semântica e a pragmática. Cabe dizer que, em muitos exemplos, vemos que a associação é motivada por um aspecto gramatical, mas entendemos que as forças semântica e pragmática são capazes de englobar os itens gramaticais também. Na terceira etapa, o objetivo era demonstrar como a noção de termo pode passar ao largo de um estudo sobre o texto especializado. Usamos dados quantitativos para demonstrar como as associações são representativas em textos especializados e diferem dos textos de língua geral.

Na quarta etapa, quisemos demonstrar como foram traduzidas algumas das associações para ilustrar, através das possibilidades de tradução, por que elas podem ser consideradas como unidades de tradução. Na verdade, como já ficou definido no Capítulo 2, as unidades de tradução são itens flexíveis, e qualquer unidade sobre a qual o tradutor se debruça em um determinado momento pode ser considerada uma unidade de tradução.

Pensamos que, se apresentarmos possibilidades de tradução para essas associações, o tradutor vai ter um leque maior de exemplos nos quais se embasar para fazer a sua tradução, além de, aos poucos, ir compreendendo e apreendendo o domínio da linguagem em questão. Vimos diversos exemplos em que as mesmas associações foram traduzidas de modos diferentes. É esse tipo de ganho que esperamos que o nosso modelo de ferramenta de apoio ao tradutor possa apresentar: para uma determinada associação de palavras, diversas possibilidades são apresentadas, ao invés de apenas aquelas com um alto grau de semelhança com o texto fonte, que é o que os sistemas de memórias de tradução atuais utilizam.

⁵⁶ Discutiremos mais sobre a hipótese no Capítulo VI.

CAPÍTULO VI

CONSIDERAÇÕES FINAIS – PARTE I

Esta primeira parte de nossa tese está ancorada na investigação da primeira questão de pesquisa, que consistia em descobrir a possibilidade de identificar um determinado tipo de associações sintagmáticas. Depois disso, como hipótese, nos propomos a investigar um possível perfil para essas unidades, que viriam a ser também recortadas, pelo tradutor, como Unidades de Tradução Especializadas.

Primeiramente, para delimitar as unidades de análise que nos propomos a investigar, foi necessário fazer uma revisão de conceitos relacionados, como o de *colocações* e o de *fraseologia*. Foi nessa fase da investigação que nos deparamos com outros conceitos que antes desconhecíamos, como o *prosódia* e *associação semântica*, *lexical primings* e *congrams*, entre outros.

Não é novidade que, quando se começa a investigar essa área, há uma grande possibilidade de entrar em um "estado de choque" com tantas abordagens diferentes sobre um mesmo tema. A própria classificação dos tipos de abordagem varia. Sabemos que isso tudo é porque em cada movimento que foi dado na investigação das associações do eixo sintagmático foram enfocados aspectos diferentes, com distintas motivações, sejam elas de ordem lexicográfica, de descrição do sistema linguístico, psicológicas, de encadeamento textual. Não nos propomos a fazer mais uma longa revisão e uma nova classificação dos estudos, pois esse não era nosso objetivo. Apenas resgatamos os que estavam mais próximos ao nosso estudo e que, em sua natureza, tinham mais pontos em comum com a nossa investigação.

Desses estudos, sublinhamos a ideia dos *lexical primings* em que uma palavra-nodo se associa a outra motivada por diferentes “forças”: semântica, gramatical, pragmática etc. Cabe destacar que as forças não são exclusivas em uma determinada associação, pois uma colocação pode ocorrer por mais de um *priming* ao mesmo tempo. Essa é uma teoria que dá conta de tentar “explicar” o porquê das associações. Já a noção de *concgrams* é mais voltada à identificação dessas associações. Não se constitui uma teoria e nem invalida que os *concgrams* sejam tratados sob a luz dos *lexical primings*, que foi uma abordagem pela qual nós acabamos optando.

Como estamos interessados em demonstrar que a visão do texto especializado sob a ótica do tradutor pode trazer contribuições interessantes para os estudos desse tipo de texto, cabia trazer a tradução, tanto como produto quanto como processo, à discussão. No Capítulo 2, tratamos de descrever como entendemos que as associações podem estar envolvidas no produto e no processo da tradução, buscando embasamento nos conceitos de risco e vantagem, usados por Gouadec, e também na Teoria Interpretativa da Tradução. Em seguida, o Capítulo III faz uma associação entre a sintagmatização e a possibilidade dessas unidades se conformarem como unidades de tradução especializadas.

Após definir os aspectos mais teóricos do trabalho, partimos para a investigação do córpus. Optamos por manter na tese as principais investigações iniciais à questão de pesquisa, pois realmente não era uma questão que tinha uma resposta no momento que a elaboramos. Não sabíamos da existência de um software no modelo do *ConcGram*, e os pesquisadores de Linguística de Córpus e Linguística Computacional também não conheciam nada semelhantes. Não encontramos nenhum estudo que abordasse associações sintagmáticas com variações de posição e sem um número predefinido de palavras entre elas. Assim, começamos avaliando a noção de matrizes e depois estudos com ferramentas da Linguística de Córpus. Os estudos que fizemos usando as listas de colocados e cálculo de Informação Mútua haviam se mostrado bastante interessantes, mas em grande escala demandariam um grande esforço manual. Esses estudos nos permitiram ver, por exemplo, como eram diferentes as associações para uma mesma palavra conforme a seção do artigo em que se encontravam.

Identificamos, estudando as associações para *crianças* nas quatro seções, que essa palavra do léxico geral, ao estar em combinação com outras, nos contextos aqui analisados, assumiam também um caráter especializado. E esse não é um caso isolado, há outras palavras da língua geral que se encontram nos textos especializados, basta pensar que, geralmente, as palavras muito técnicas costumam ter baixa frequência nos córpus especializados. Isso tudo

nos leva a confirmar que a terminologia de uma área pode ser vista dentro de um contínuo, e que apenas observando as unidades léxicas em associação é que temos uma dimensão de sua possível localização em um espectro que vai de léxico ao termo ou vice-versa. Assim, palavras como *presença*, *intervalos*, *controle* podem ser aceitas como pertencentes ao campo da Medicina quando estão relacionadas a determinadas outras. Foi importante detectar também que há essa relatividade na “rotulagem” de um item léxico como “termo”, pois evidenciamos ainda mais como as relações textuais são fundamentais para tal determinação. É por esse viés que apresentamos também algo que pode não ser uma novidade, mas que talvez não seja tão lembrado quando se trata de descrever os textos especializados, que, assim como a língua se constrói nas relações de oposições, também o texto, e o caráter de especialidade do texto, só se constroem nessas condições de relações e oposições.

Durante a pesquisa sobre *colocações* e *fraseologias*, conhecemos o trabalho sobre os *congrams*, com o qual nos identificamos imediatamente. Após termos encontrado, então, um método que permitiria investigar as associações em maior escala, segmentamos nossa análise em quatro etapas, para assim isolar e perceber melhor a característica que estávamos enfocando (sua descontinuidade e variação posicional e forças semântica e pragmática), o que foi apresentado no Capítulo V.

Nossa primeira questão de pesquisa visava descobrir a possibilidade de identificar, em larga escala e com base em corpus, associações sintagmáticas que não se encontram em uma sequência contínua. Buscamos respondê-la procurando identificar qual seria o método mais indicado para tal. Inicialmente, usamos apenas o aplicativo *WordSmith Tools* e Cálculo da Tabela de Colocados, disponibilizado pelo Prof. Tony Berber-Sardinha. Os resultados eram satisfatórios, mas queríamos algo mais robusto. Graças à indicação do artigo *Unconverging the extend of the phraseological tendency: towards a systematic analysis of congrams* (Cheng, 2008), do qual John Sinclair é coautor, que nos foi feita pela orientadora do estágio de doutorado sanduíche na Universidade de Birmingham, a Profa. Gabriela Saldanha, chegamos ao software *ConcGram*.

Com o software, concentramos nas 4 etapas de investigação da hipótese, apresentada novamente a seguir:

Existem nos textos especializados associações no nível sintagmático que:

- a. são combinações mais ou menos fixas e frequentes de palavras;
- b. não são necessariamente contínuas;

- c. estão associadas no nível sintagmático por uma atração semântica e/ou pragmática e/ou gramatical;
- d. não necessariamente contêm um termo em sua composição, aqui termo entendido como uma expressão lexical associada a um conceito.

Essas associações se configuram como Unidades de Tradução Especializadas porque são recorrentes, fazem parte dos “modos de dizer” dos gêneros e do domínio em questão e apresentam riscos e vantagens ao tradutor, visto que implicam soluções complexas de tradução.

Acreditamos que, através das Etapas de 1 a 4, no Capítulo IV, conseguimos demonstrar cada uma dessas crenças. A flexibilidade posicional, que significa que há variações na ordem dos itens na frase, e a flexibilidade composicional, que significa que pode haver diferentes números de palavras entre as palavras associadas, ficaram evidentes nos exemplos que mostramos. É importante considerar que tratamos como associações aqueles itens que, entre as frases que os continham, transmitiam um mesmo significado ideacional, nos termos da GSF. Eles cumpriam uma mesma função na frase e é isso que demonstra que, para um mesmo sentido ou função, temos uma série de possibilidades de composições textuais, utilizando palavras recorrentes em combinações também recorrentes.

Também investigamos como se comportavam as associações em textos especializados e na linguagem não-científica e constatamos, ainda, que em uma amostra de certa forma limitada, que as associações entre palavras “cheias” é mais frequente em textos especializados. Esse achado pode continuar levando a caminhos de investigação sobre o texto especializado que enfoquem aspectos do texto como um todo, para além da aparência de que termos e fraseologias especializadas devam ser, como sempre tem sido, os principais focos de atenção. Logicamente, que não é de agora que outros elementos do texto especializado são enfatizados. Já citamos anteriormente Hoffman (2004) e, dentro da tradição de linguistas alemães, citamos ainda, por exemplo, o trabalho de Weinrich (2005 *apud* Zilio, 2009), que assinala, por exemplo, que “é característico que as linguagens especializadas dêem preferência a uma expressão nominal em vez de uma expressão verbal”. cremos, assim, que a inovação e maior utilidade estarão na produção de materiais de consulta e pesquisa que contemplem essas outras características e elementos do texto especializado, pois assim podem ser mais úteis ao tradutor.

É por esse viés, também, que consideramos que as associações que detectamos e descrevemos nas Etapas 1, 2 e 3 são também Unidades de Tradução Especializadas. Ao considerarmos que são unidades de tradução, estamos dando atenção para itens recorrentes, que fazem parte do texto especializado e que ajudam a construir seu encadeamento sintagmático e, portanto, as formas como as informações especializadas são transmitidas. Não haveria texto especializado se não houvesse a combinação entre certas palavras e a presença de outras em seu entorno. Como dissemos, talvez para aqueles que não estão acostumados à prática e ao estudo da tradução seja mais complicado entender nossas análises na Etapa 4 e o porquê, afinal, tais associações são Unidades de Tradução Especializadas. Procuramos demonstrar que uma associação foi traduzida de diferentes formas, e nesse sentido, pensamos haver um risco para a tradução, pois algumas opções podem ser acertadas e outras nem tanto. O acerto, claro, dependerá do que é mais ou menos convencionalizado em termos de expressão pela comunidade discursiva da área.

Demonstramos, ainda, que alguns itens das frases com as associações não foram traduzidos, evidenciando também que apresentam certa dificuldade para o tradutor. Mas, talvez, o mais importante seja destacar justamente que essas associações são apenas um dos recortes possíveis para Unidades de Tradução Especializadas, e que é possível aprender a traduzir Medicina observando e sabendo como traduzir essas associações, que se repetem e que, ao serem “dominadas” pelo tradutor, podem representar ganhos de tempo e qualidade.

Depois de conduzir um trabalho cujos questionamentos foram sobre a importância de determinadas associações sintagmáticas para a tradução de artigos de Pediatria, queremos também expor como entendemos que uma abordagem voltada para a prática da tradução gera também questionamentos sobre estudos e práticas em Terminologia e Terminografia. Temos o tradutor — típico usuário em potencial de obras Terminográficas — como um também potencial contribuinte para o desenvolvimento de tais obras. O tradutor pode levar sua experiência no manuseio de textos especializados para o âmbito das reflexões em Terminologia, em comparação a quando ela é olhada por si mesma, isto é, a Terminologia (e a terminologia)⁵⁷ vista por terminólogos, com fins e propósitos terminológicos e terminográficos. O fato de que tradutores são usuários diretos de produtos terminográficos nos leva a entender que seria um contra-senso a Terminologia negligenciar as contribuições

⁵⁷ Aqui fazemos uma distinção entre Terminologia, com inicial maiúscula, como ramo da lingüística; e terminologia, com inicial minúscula, enquanto denominação genérica para termos/unidades de significação especializada. (Krieger e Finatto, 2004).

e críticas — quando bem fundamentadas — que esses profissionais podem vir a ter. E, caso a Terminologia tradicional prefira, então, não contribuir com essas necessidades do tradutor, por considerar que estaria se desvirtuando de seu objeto de estudo, temos que optar por seguir outros caminhos, quem sabe o proposto por Teixeira (2008), do desenvolvimento de uma Terminologia Tradutiva, ou ainda, dar mais foco à Terminologia de base textual.

Nosso foco se dá sobre a combinação de itens justamente por considerarmos que essa combinação é um dos fatores que também compõem o caráter especializado do texto, em “oposição” a um pensamento de que os termos seriam sempre os maiores responsáveis pelo caráter especializado de um texto. Trata-se, assim, de defender que também esses elementos são importantes e que não podem ser negligenciados em meio a teorias ou métodos descritivos que pretendem dialogar com tradutores e com a Tradução.

Na Terminologia, estamos em um estágio que considera que todas as palavras têm potencial para *atualizarem-se* como termos, dados os seus contextos e as suas condições de ocorrência. Saussure já havia feito uma consideração semelhante ao dizer que as palavras definem-se em relação umas às outras e que são como moedas, ajustando seu *valor* de acordo com a posição no sistema como um todo.

Obviamente, tudo o que se discutiu sobre Terminologia (entendida como disciplina ou área de estudos) até os dias atuais não se resume a essa observação sobre palavra *versus* termo, mas acreditamos que considerar a Terminologia como um ramo da Linguística não é tão “pejorativo” como foi exposto por Cabré (2004, p. 14), ainda que há 6 anos. Na ocasião, a autora colocava que considerar a posição da Terminologia como parte de outra disciplina, como a Linguística, “significaria que a Terminologia não teria nenhuma autonomia, mas que seria um mero apêndice de outra disciplina”. Se assim fosse, tantos os outros importantes ramos da Linguística (Estilística, Estudos do Texto e do Discurso, Sociolinguística etc.) seriam “meros apêndices”, o que não é verdadeiro. São, sim, apenas subáreas da Linguística construídas conforme os fins que almejam, mas que, sobretudo, tratam de linguagem, assim como a Terminologia (que além da face linguística pode também ser reconhecida por outros ângulos: o cognitivo e o comunicativo).

Também nesse ponto, Cabré (2004) considera que a concepção de linguagem da Terminologia e da Linguística diferem, justificando que a “Linguística parte da linguagem através de uma amostra idealizada” e que “os termos interessam por si mesmos, sem nenhum interesse por sua flexão (...) ou por sua sintaxe”. Acreditamos que a linguística não parte de

uma amostra idealizada, a não ser a linguística gerativa de Chomsky, que é apenas “uma das” linguísticas possíveis e, como deixamos subentendido anteriormente, não é a mais indicada para trabalhos voltados para terminologia.

A Linguística, do modo como nela nos inserimos, parte de observações de realizações da linguagem, e não de uma linguagem idealizada. Quanto aos termos interessarem por si mesmos, acreditamos que, em uma visão “filosófica” da Terminologia, essa consideração parece ser verdadeira, porém, ao fazer Terminologia, não há, hoje, como começar a definir e identificar o que é termo em meio a um isolamento epistemológico.

Ainda, segundo Cabré (2004, p. 14), “os aspectos pragmáticos são os que melhor permitem diferenciar os termos das palavras (...) i) por seus usuários, ii) pelas situações em que se utilizam, iii) pela temática que veiculam e iv) pelo tipo de discurso em que costumam aparecer”. Acreditamos que há evidências suficientes para demonstrar que, se assumirmos que os termos são palavras que fazem parte do léxico e que se atualizam como termos, as quatro justificativas acima não permitem diferenciar termos de palavras. Como explicaríamos, por exemplo, que “termos” como *reverso*, *arremeter*, *grooving*, relacionados à aviação, em questão de dias, deixaram de ser usados apenas por pilotos e controladores, em situações de manutenção, pouso e decolagem, para serem usadas na imprensa e em discussões populares sobre as causas de um acidente aéreo? Entendemos que essas palavras não deixaram de ser termos quando “adentraram” a esfera do léxico geral, assim como não deixavam de ser palavras quando eram usadas só em um tipo de discurso, por uma certa comunidade de usuários.

Em nosso ponto de vista, no momento em que uma sequência de caracteres como *com média de x meses*, que não apresenta nenhuma palavra que seja considerada termo técnico, se mostra frequente em um texto especializado de Medicina, ela deixa de ser apenas um fenômeno do texto especializado e passa a configurar-se como pertencente ao âmbito da Medicina pela sua relação com as palavras. Se buscarmos por essa mesma expressão em um cópulo de artigos de Informática, como o fizemos, não encontramos uma ocorrência da expressão. Assim, o que a princípio parece ser uma expressão do âmbito da ciência em geral já não se comprova em algumas áreas. Poder-se-ia argumentar que é uma expressão típica da Estatística, o que provavelmente se confirma. No entanto, devemos pensar onde estão os limites entre Medicina e Estatística atualmente? Além do que, existe também uma terceira área, a Bioestatística. Entendemos que a palavra *média* já avança, então, no espectro que vai

de palavras a termos e se estabelece numa posição que se aproxima mais do caráter de termo, dado o contexto em que se realiza.

Por esses motivos expostos, pensamos que a abordagem das linguagens científicas com uma perspectiva comunicativo-textual, que extrapole terminologias, poderá gerar novos e enriquecidos tipos de produtos terminográficos, principalmente quando se pretenda atingir um público alvo como os tradutores.

Após a investigação da hipótese e de nossa primeira questão de pesquisa, este trabalho segue com a identificação de dados para alimentar e incrementar a eficiência de um sistema de auxílio ao tradutor, baseado em *cópus*, nos moldes de um sistema de tradução baseado em exemplos.

PARTE II: FERRAMENTA DE APOIO AO TRADUTOR

CAPÍTULO VII

LINGUÍSTICA COMPUTACIONAL E A TECNOLOGIA EM TRADUÇÃO

Nossa segunda questão de pesquisa implica a construção de um recurso computacional que auxilia o tradutor com base nas associações descritas no capítulo anterior. Neste capítulo, apresentamos uma breve contextualização da área de Linguística Computacional bem como de alguns trabalhos desenvolvidos que fazem associação entre o processamento computacional da linguagem humana e a produção/tradução/pesquisa de texto. O objetivo dessa breve revisão também foi o de tentar identificar que tecnologia poderíamos aplicar para o desenvolvimento dessa ferramenta que pretendemos desenvolver. Passamos por algumas áreas, como Sistemas de Recomendação e Recuperação de Informações e Memórias de Tradução.

1. Linguística Computacional

Tomando a definição de Vieira e Lima (2001):

A linguística computacional é a área de conhecimento que explora as relações entre linguística e informática, tornando possível a construção de sistemas com capacidade de reconhecer e produzir informação apresentada em linguagem natural.(p. 1)

Percebemos que o desenvolvimento inicial da área procurou seguir os diferentes níveis de conhecimento linguístico: da fonologia à pragmática, passando pela morfologia, sintaxe e semântica. Observamos também que tem sido dado destaque, além desse processamento simbólico, que acompanha a tradição da Linguística Gerativa, aos métodos estatísticos ou aos métodos híbridos (Vieira e Lima, 2001). Conforme apontam Vieira e Lima (2001), as

“abordagens quantitativas passaram a adicionar robustez e abrangência a sistemas simbólicos de processamento de linguagem natural.” Um processamento estatístico da linguagem natural utiliza essas abordagens para o processamento automático de textos, o que permite levantar probabilidades de ocorrência de palavras, sequências de palavras, categorias de palavras, etc.

Essas abordagens eram usadas anteriormente por correntes distintas de pesquisa, e os métodos estatísticos eram usados mais pela área de processamento da fala. Posteriormente, nas décadas de 80 e 90, começaram a aparecer trabalhos sobre resolução de ambiguidade de sintagmas preposicionais com base nos mesmos métodos estatísticos já tradicionais em trabalhos de processamento de fala. Tais estudos foram bem sucedidos e o conhecimento necessário para a solução de problemas começou a ser induzido pela análise de grandes corpú, ao invés de ser construído manualmente na forma de regras simbólicas. Conforme esclarecem Viera e Lima (2001), “desde então, cada uma das comunidades, a simbólica e a estatística, tem aceitado melhor a ideia de que, para se atingir os objetivos de cada área, pode ser necessário o conhecimento amadurecido pela outra.” (p. 3)

Aplicações decorrentes do estudo e do desenvolvimento da área de Linguística Computacional são: reconhecedores e sintetizadores de fala, corretores ortográficos e gramaticais, tradutores automáticos, geradores de textos e resumo, interfaces em linguagem natural (ou sistemas de perguntas e respostas), recuperação de informação, extração de informação, sistemas de recomendação, entre outras. Em cada uma dessas aplicações, utilizam-se métodos mais ou menos estatísticos e combinações deles.

Vamos nos deter a seguir em algumas áreas da Linguística Computacional que estão de alguma forma mais relacionadas com esta pesquisa, devido a seus objetivos e/ou métodos, e, em seguida, percorreremos áreas mais relacionadas à tradução.

1.1. Recuperação de informações

Conforme definição de Vieira e Lima (2001)

A recuperação de informação é a área de aplicação envolvida com a obtenção de documentos relevantes dado um determinado tema, e não está diretamente envolvida com a obtenção de uma informação específica ou com a obtenção de resposta a uma dada pergunta. Recuperação de informação pode, então, ser definida como sendo o conjunto de técnicas que servem ao propósito de encontrar documentos relevantes de acordo com uma necessidade de informação. (p. 33)

Os sistemas de recuperação automatizada da informação foram usados inicialmente para gerenciar a explosão da informação na literatura científica. Muitas universidades e bibliotecas públicas usam estes sistemas para prover acesso a livros, jornais, periódicos e outros documentos. Acreditamos que suas sistemáticas podem ser aproveitadas para a configuração da nossa ferramenta informatizada para uso de tradutores.

Teufel (1999), em sua tese de doutorado, aponta que há uma série de estudos empíricos sobre a experiência de realizar buscas, sobre os tipos de perguntas nas buscas e sobre os objetivos dos usuários que consultam bases de informação. A autora considera que alguns dos poucos estudos que contemplam usuários pouco experientes concluem que estes estão em desvantagem, pois as perguntas que elaboram aos sistemas e as suas necessidades no contexto da busca não são bem definidas.

Isso ocorre porque esses usuários menos experientes não conhecem nem os conceitos básicos, nem a terminologia de uma área, o que faz com que suas pesquisas por palavras-chave também não funcionem. Os resultados de suas buscas são pouco específicos e retornam ou muitos resultados, ou resultados em que o termo tem algum outro significado. Essa constatação, ao render uma analogia com buscas de informação sobre usos de língua por parte de tradutores, faz lembrar que tradutores inexperientes ficariam em uma situação parecida quando em busca de referências para suas escolhas de tradução. Acreditamos que oferecer informação qualificada para esse tipo de usuário seria um diferencial de nossa ferramenta, pois ela terá armazenado em sua base uma representação de um conhecimento: o conhecimento sobre as associações sintagmáticas mais relevantes de Pediatría. Isso será disponibilizado para quem faz a pesquisa, um tradutor iniciante ou estudante de tradução.

No âmbito da recuperação de informações, encontramos a *Cross-language Information Retrieval*, ou Recuperação de informações multilíngues, que abordamos a seguir.

1.1.1. Recuperação de informações multilíngues

Esta área se ocupa de desenvolver métodos para encontrar documentos em qualquer idioma, usando uma consulta escrita em uma única língua. Esses sistemas valem-se de técnicas de tradução automática, ontologias, dicionários e técnicas baseadas em córpus. As

buscas, de um modo ideal, devem apontar, para o consulente, textos de conteúdo e gênero semelhantes em uma ou mais línguas.

Em um trabalho apresentado na Conferência LREC 2006 (Language Resources and Evaluation, cujos autores são da Universidade de Leeds (Sharoff *et al.*, 2006), vemos um objetivo que se intercrusa com o nosso. Seus autores apresentam uma ferramenta para a busca de equivalentes apropriados de tradução para palavras do léxico geral, utilizando um córpis comparado. De acordo com Sharoff *et al.* (2006), a maior parte dos trabalhos sobre detecção automática de equivalência é feito para encontrar equivalentes terminológicos. Procuramos manter contato com o referido pesquisador para saber da implementação e desenvolvimento do sistema, mas não obtivemos retorno.

1.1.2. Sistemas de recomendação

Segundo Resnick (1997, *apud* Ávila), os Sistemas de Recomendação ajudam e incrementam o processo natural e social de auxílio na tomada de decisões do dia a dia”. Acrescenta ainda que, em um sistema de recomendação típico, pessoas fornecem recomendações como entrada que o sistema então agrega e direciona para a audiência apropriada. Mais recentemente, Burke (2006, p.1) definiu sistemas de recomendação sob um ponto de vista mais amplo, como “qualquer sistema que produza recomendações individualizadas como saída, ou que tenha o efeito de guiar o usuário de forma personalizada a objetos interessantes e úteis, diante de uma grande variedade de opções”.

Um ponto importante a esclarecer é que, conforme esclarece Burke (2006), o critério de “recomendações individualizadas” e “interessantes e úteis” é que separa um sistema de recomendação de sistemas de recuperação de informações do sistema de motores de busca.

De maneira geral, os sistemas de recomendação são formados por:

- dados prévios ou armazenados (*background data*): correspondem a toda informação que o sistema armazena para utilizar no processo de recomendação;
- dados de entrada ou do usuário: trata-se de informação fornecida pelo usuário para que o processo de recomendação possa ser iniciado;

- algoritmo de recomendação: é responsável por combinar dados prévios e dados de entrada para que sejam fornecidas as devidas recomendações;
- os dados prévios compreendem a relação de itens a serem recomendados (catálogo de produtos, documentos, páginas web, conteúdo multimídia etc.), podendo englobar também o conjunto de usuários já cadastrados e suas avaliações.

Exemplos de recomendações desse tipo de sistema podem ser vistos em *sites* de comércio eletrônico, como o da livraria virtual Amazon.com.

Os dados que o usuário fornece servem de base para modelar seu perfil e, segundo a literatura, podem ser coletados de forma explícita ou implícita. As avaliações explícitas se referem àquelas em que as preferências são expressas conscientemente, seja por meio de questionários ou de voto declarado em itens; as avaliações implícitas são inferidas, automaticamente, com base na interpretação das preferências do usuário, em sua navegação, em seu histórico de compras, etc.

1.2. Alguns trabalhos aplicados em Linguística Computacional

Apresentamos aqui alguns trabalhos desenvolvidos em Linguística Computacional com os quais tivemos contato durante a elaboração desta tese.

Em Acosta (2006) é apresentada uma ferramenta destinada a gerar grafos a partir de textos. Essa ferramenta tem como objetivo facilitar a compreensão do conteúdo das informações de um texto. As funções da ferramenta são analisar o texto e descobrir a relação entre as palavras através da proximidade entre elas: palavras mais próximas têm um peso maior do que as mais distantes. O grafo gerado a partir daí mostra todas as relações entre as palavras de um texto e também pode demonstrar o número de palavras que existe entre as palavras que têm algum relacionamento. Imaginamos que com essa ferramenta seria possível, no caso da palavra usada como exemplo (*prevalência*), calcular as palavras que mais se associam a ela e que, portanto, devem fazer parte de uma sequência notável de caracteres. (Neste caso, estaríamos partindo da porção fixa da matriz/sequência).

O SciPo, resultado do estudo de doutorado de Valéria Delisandra Feltrim (2002), desenvolvido no NILC, é um “conjunto de ferramentas integradas com o objetivo de auxiliar

o estudante na escrita de resumos e introduções de textos acadêmicos” (<http://www.nilc.icmc.usp.br/nilc/projects/scipo.htm>). O SciPo é realizado sobre uma base de textos na área de Ciência da computação. Já o Scipo-Farmácia, desenvolvido no âmbito do mesmo grupo, é um “conjunto de ferramentas computacionais construídas para auxiliar o estudante a escrever artigos científicos em inglês (...) e baseia-se em análise de textos da área das Ciências Farmacêuticas.” O SciPo-Farmácia foi um projeto realizado em parceria com a Faculdade de Ciências Farmacêuticas da USP-São Paulo. Apesar das duas áreas declaradas (Computação e Farmácia), estas estão presentes na ferramenta apenas através dos exemplos, pois as unidades apresentadas são constituintes do gênero “artigo científico” e não componentes típicos da área em questão. Além disso, os dois SciPos não são voltados para o tradutor.

A tese de doutorado de Simone Teufel (1999) também tratou de artigos científicos com uma abordagem computacional. É um estudo voltado para a recuperação de informações. A tese tinha como objetivo contribuir para a geração automática de elementos para a recuperação de documentos. A parte prática da tese envolveu descobrir como elementos de busca de informações baseados em zonas argumentativas podem ajudar pesquisadores em atividades de busca de informações científicas, especialmente o pesquisador que não têm conhecimento em uma dada área. Esse ponto da referida tese nos interessa, pois, assim como um profissional ou estudante que é novo em uma área, ao tradutor também é apresentado o desafio de ter que entender sobre determinado assunto para produzir sua tradução.

Um sistema de recomendação que tem uma interface com texto do qual tivemos conhecimento foi desenvolvido na Universidade Católica de Pelotas, o SISRECAC. Segundo (Ávila, 2008), “o objetivo desse sistema é ser um repositório de documentos que recomenda artigos científicos baseados nos próprios documentos armazenados” (p. 1). O sistema utiliza o Google Acadêmico (<http://scholar.google.com/>) para indicar artigos científicos que sejam similares a documentos armazenados pelos usuários da ferramenta.

O SISRECAC utiliza como palavras-chave as cinco palavras mais frequentes nos artigos, excluindo-se as chamadas *stopwords*. Vale pontuar aqui que utilizamos um método de recorte semelhante na análise das associações que fizemos na Parte I.

2. A tecnologia e a tradução

Em geral, quando se menciona tecnologia e tradução juntas, em ambientes mais “leigos” ou afastados do mundo da tradução, lembra-se logo dos sistemas de tradução automática. Como mencionamos na introdução, a tradução automática hoje já conhece melhor seus limites e possibilidades de aplicação, mas isso não impede que haja um outro papel para a tecnologia no processo de tradução. Conforme aponta Bowker (2002), o foco agora não está mais em como a máquina pode substituir o tradutor, mas em como ela pode ajudá-lo. A principal diferença entre tradução automática e tradução assistida por computador está em quem é o responsável por realizar a tradução. Na tradução automática, quem traduz é a máquina, embora o resultado possa ser depois editado por um tradutor humano. Na tradução assistida por computador, os tradutores humanos são responsáveis pela tradução, mas utilizam meios computadorizados para ajudá-los a completar suas tarefas e a aumentar a produtividade (Bowker, 2002). Além desses sistemas, conhecidos como CAT (*Computer-Aided Translation*) tools existem outros recursos que são auxiliares da tradução humana, como os editores de texto, verificadores ortográficos, a Internet, os portais de colaboração entre tradutores.

Antes de falarmos mais extensamente nas seções a seguir a respeito da tradução automática e das CAT tools, queremos apenas trazer um exemplo de “ferramentas tecnológicas” que não são nem de tradução automática e nem CAT tools. Queremos apenas demonstrar outros recursos atuais de pesquisa que vem sendo usados e que diferem dos glossários e dicionários tradicionais.

Teixeira (2008) já apontou que:

Os dicionários sempre foram e ainda são uma das principais ferramentas da tarefa tradutória. No entanto, a terminografia parece não ter se beneficiado ainda de forma sistemática, pelo menos no Brasil, da estreita relação entre dicionários técnicos e esse público-alvo específico e cada vez mais expressivo: os tradutores técnicos. (p. 29)

Nesse sentido, hoje acreditamos que já podemos ver, aos poucos, que os tradutores estão se envolvendo mais na tarefa e se preocupando em como contribuir para o desenvolvimento de materiais de referência que lhes sejam úteis. Ainda, dada a multiplicidade de informações que são úteis ao tradutor, o uso da tecnologia só tem a contribuir para a divulgação e viabilidade desse tipo de material. Vemos isso não só em trabalhos e pesquisas acadêmicas, como na tese de Teixeira (2008) e nos trabalhos elaborados no âmbito dos projetos TextQuim (www.ufrgs.br/textquim), TextTecc (www6.ufrgs.br/textecc/), Termisul

(www.ufrgs.br/termisul) e CoMET (<http://www.fflch.usp.br/dlm/comet/>), mas também em materiais patrocinados por empresas privadas, como o portal do ProZ (www.proz.com).

Talvez, à primeira vista, para um terminólogo/lexicólogo, os recursos de glossários do referido portal possam parecer uma grande confusão, mas o fato é que nesse portal o tradutor encontra respostas para suas dúvidas com base em material publicado no site, anteriormente postado também como uma dúvida de algum colega e respondida por outros. Nele, o tradutor encontra sugestões de tradução para palavras e expressões que, geralmente, não são encontradas em dicionários.

Vamos citar como exemplo a tradução para *Ato de Concentração*, num contexto jurídico. Caso o tradutor precisasse traduzir esse termo para o inglês e quisesse buscar opções de tradução no portal do proz.com, ele acessaria a página <http://ksearch.proz.com/search/>, onde encontraria a seguinte tela:

Figura 40 – Tela em branco na página de acesso dos glossários do proz.com

Depois de digitar a expressão ou palavra a ser buscada no campo *Term or phrase*, as opções seguintes são: direção da tradução, em *Languages*; área, em *Field*; e Nível de dificuldade em *Difficulty level*. A seguir, o usuário, se desejar, seleciona em quais bases deseja que *Ato de concentração* seja buscado.

Term or phrase: (Required)
 Non-allowed characters: * (Search operators)
 Whole words only Match exact phrase

Languages: to
 [Show more languages]
 Also search reverse language pair (target into source)
 Search for all likely character encodings

Field:

Difficulty level:

Resources to search:
 KOG (KudoZ Open Glossary) [Browse]
 KudoZ archive [Browse]
 Personal glossaries [Browse]
 Glosspost (database of online glossaries) [Browse]
 Wikiwords [Browse]
 Open search results in a new window

Figura 41 – Preenchimento da consulta

São apresentados todos os resultados encontrados e, para acessar a sugestão de tradução, é preciso selecionar uma das opções. Selecionamos a que está marcada com um círculo:

KudoZ archive terms matching your search criteria			
28 results found. Showing results 1 - 28			
	Field	Term	Asked by
<input checked="" type="checkbox"/>	Business/Commerce (general) (Bus/Financial)	concentração horizontal (1)	Michael Powers (PhD)
<input checked="" type="checkbox"/>	Financial Markets = Finance (Bus/Financial)	Concentração dos concorrentes (3)	Tina Delia
<input checked="" type="checkbox"/>	Law: Contract(s) (Law/Patents)	a ato (2)	Paul Sadur
<input checked="" type="checkbox"/>	Law (general) (Law/Patents)	a origem do ato (1)	Non-ProZ.com
<input checked="" type="checkbox"/>	Psychology (Science)	ato falho (2)	Non-ProZ.com
<input checked="" type="checkbox"/>	(Law/Patents)	ato declaratório (3)	Auristela Webster
<input checked="" type="checkbox"/>	(Law/Patents)	DO ATO DE CONCENTRAÇÃO (2)	Sara Salvador
<input checked="" type="checkbox"/>	Law: Contract(s) (Law/Patents)	ato ou fato (2)	Karen Vincent-Jones
<input checked="" type="checkbox"/>	(Law/Patents)	ato do princípio (1)	Lars Bradley
<input checked="" type="checkbox"/>	Law: Contract(s) (Law/Patents)	neste ato (2)	gususi
<input checked="" type="checkbox"/>	Law: Contract(s) (Law/Patents)	no ato da assinatura (4)	Pedrina Godofredo
<input checked="" type="checkbox"/>	Law: Contract(s) (Law/Patents)	ato publico (2)	Adriana Portas
<input checked="" type="checkbox"/>	(Bus/Financial)	ato convocatorio (5)	Jeanne Zang
<input checked="" type="checkbox"/>	(Other)	ato simbólico (1)	Sidnei Arruda
<input checked="" type="checkbox"/>	Law: Contract(s) (Law/Patents)	ato de ciencia (4)	maisa
<input checked="" type="checkbox"/>	(Law/Patents)	ato próprio de mandatário (3)	xxxHoqqelen

Figura 42 – Resultado da busca

Ao acessar o link DO ATO DE CONCENTRAÇÃO, temos acesso à seção de definição/comentários. Primeiramente, é apresentado o termo em português e sua definição, e depois o comentário do tradutor que respondeu, em outro momento, à dúvida postada:

6:10pm Oct 15, 2002 Login or register (free) for more options.

Portuguese to English translations [PRO]
Law/Patents - unspecified / Trade Market

Portuguese term or phrase: **DO ATO DE CONCENTRAÇÃO** **Sara Salvador**
Brazil

DO ATO DE CONCENTRAÇÃO:
Entende-se por ato de concentração todo e qualquer ato que possa limitar ou prejudicar a livre concorrência ou resultar na dominação de mercados relevantes de bens ou serviços, o qual deverá ser submetido à apreciação do CADE, conforme disposto no artigo 54 da lei No. 8884, de 11 de junho de 1994.

antitrust filing ou concentration act **Selected response from:** 

é uma questão um pouco controversa...
Veja bem. O ato de concentração em si é o ato que possa prejudicar a concorrência. É também o documento que vc. tem de apresentar ao CADE p/ que ele aprove o ato [fusão, incorporação, etc.]

O antitrust filing é um "filing" [uma petição, processo, etc.] apresentado contra uma operação considerada prejudicial à concorrência...[como no caso da Microsoft]...Tenho visto em alguns documentos o termo "Concentration act" no caso do ato e "antitrust filing" no caso do documento... É preciso analisar qual o real sentido do termo no contexto... Espero que isso ajude...

Note from asker to answerer

Thanks for your help!
Sara
4 KudoZ points were awarded for this answer

Figura 43 – Comentário acerca do termo e sugestão de uso

Como fica evidente, a resposta é completa e, mesmo que não se possa confiar 100% na definição/sugestão apresentada por um tradutor, já é um ponto de partida para que o tradutor pesquise as opções dentro de seu contexto.

Um outro exemplo (Figura 44), buscando-se a estrutura fraseológica “E por estarem assim justas e contratadas, as partes assinam o presente” que dificilmente seria encontrada em um dicionário.

KOG (KudoZ Open Glossary) terms matching your search criteria			
298 results found. Showing results 1 - 50			
Field	Term → translation	Entered by	
(Law/Patents)	POR ESTAREM ASSIM JUSTAS E CONTRATADAS → Therefore, by being fairly binded and subjected to the provisions of		
(Law/Patents)	estarem assim justos e contratados → In witness whereof...	BrazBiz	
Law (general) (Law/Patents)	partes entre si justas e contratadas, a saber → the contracting parties, which are..	airmailrpl	
Business/Commerce (general) (Bus/Financial)	Por terem assim ajustado → In witness whereof, the Parties ...	rhandler	
Pedagogy = Education (Other)	E, por ser verdade, firmo a presente declaração → I vouch for the truth of the foregoing.foregoing.	REAREA	
Certificates, Diplomas, Licenses, CVs (Other)	mandei passar a presente certidão que vai por mim assinada → I ordered the issuance of this certificate, which is signed by me	rhandler	
(Law/Patents)	justas e contratadas → in witness whereof	ana_brum	
Linguistics (Science)	assim como os de 'hemisférico' → as well as the ones that express the idea/concept of hemisphere	Claudio Mazotti	
Law: Contract(s) (Law/Patents)	Assim o disse e outorguei → Thus I declare and do grant	Arlete Moraes	
Linguistics (Science)	Vê-se assim que o verbo abarca → Thus it can be seen that the verb encompasses	Matt Coler	
Law (general) (Law/Patents)	Ainda que assim se não entendesse → Even if it wasn't understood like this	Vania de Souza	
(Other)	Raciocinar assim é uma ótima maneira... → This kind of reasoning is a good way...	Rodolfo Beneyto	
Labor = Human Resources (Bus/Financial)	e assim nos pareceu, de facto, no início → and that was our initial impression as well	Cynthia Lopes	
Law (general) (Law/Patents)	Assim fará Vossa Excelência JUSTIÇA! → There shall be justice, your Honor	tania nogueira	

Figura 44 – Exemplo de consulta de uma estrutura fraseológica

Como se percebe, e ressaltando o que já mencionamos na introdução desta tese e nesta seção, a tecnologia auxilia o tradutor de uma maneira muito rica e, praticamente, indispensável atualmente. Portanto, a associação entre tecnologia e tradução deve estar livre de preconceitos originados de resultados iniciais de sistemas de tradução automática ou de delimitações de fundo metodológico que impeçam a inserção das unidades diferentes de termos e fraseologias especializadas. Nas seções a seguir, nos deteremos na tradução automática e nos sistemas de apoio ao tradutor.

2.1. Tradução automática

O interesse pela tradução automática (TA) não é recente, intensificou-se a partir de 1946, com o início da guerra fria, principalmente por americanos e ingleses. No final dos anos 50, houve o interesse de outros países europeus, mas os primeiros projetos na área não alcançaram as expectativas, e o desenvolvimento dos sistemas acabou sendo mais lento do que se esperava (Specia & Rino, 2008). O sentimento já era muito mais de frustração do que de otimismo, e os problemas apresentados já não pareciam simples de resolver. No ano de 1964, foi criado um comitê para avaliar o status da área, e o relatório final, conhecido como relatório ALPAC, não foi favorável à continuidade de pesquisas e de desenvolvimentos na área. É preciso lembrar, no entanto, que muitos sistemas bem sucedidos utilizados nos dias de

hoje tomaram como ponto de partida os problemas detectados em programas anteriores, das décadas de 40 e 50. Depois de um período de quase ausência de pesquisas na área, houve um *revival*, nos anos 70, da tradução automática, a qual se tornou muito mais realista. Percebeu-se que os resultados poderiam ser úteis mesmo se não apresentassem como saída textos muito bons.

Desde então, a área tem reconhecido seus limites e tem buscado aperfeiçoar-se para atender a objetivos específicos com mais “modéstia”. Não existe mais a meta de tomar o lugar de tradutores e nem de apresentar textos perfeitos prontos para serem publicados. Para fins de leitura e de entendimento de informações, por exemplo, mesmo uma tradução de baixa qualidade pode ser suficiente, o que hoje já pode ser obtido num tempo mínimo. Ou então, como apresentamos na introdução, mediante textos já preparados para a tradução automática para fins de entendimento de informações, como em extensos manuais de software, a intervenção humana já pode ser mínima.

Hoje, para aqueles que lidam com essa área – sejam pesquisadores ou clientes – um sistema de alta qualidade é aquele que gera um texto que permita uma revisão sem grandes problemas e cuja operação completa (incluindo essa revisão) ofereça uma boa relação custo-benefício (Slocum, J. 1985).

A Microsoft, por exemplo, exibe o seguinte aviso⁵⁸ nas páginas do site de suporte que foram feitas por tradução automática,:

IMPORTANTE: Este artigo foi traduzido por um sistema de tradução automática (também designado por Machine Translation ou MT), não tendo sido portanto traduzido ou revisto por pessoas. A Microsoft possui artigos traduzidos por aplicações (MT) e artigos traduzidos por tradutores profissionais, com o objetivo de oferecer em português a totalidade dos artigos existentes na base de dados de suporte. No entanto, a tradução automática não é sempre perfeita, podendo conter erros de vocabulário, sintaxe ou gramática. A Microsoft não é responsável por incoerências, erros ou prejuízos ocorridos em decorrência da utilização dos artigos MT por parte dos nossos clientes. A Microsoft realiza atualizações frequentes ao software de tradução automática (MT). Caso venha a encontrar erros neste artigo e queira colaborar no processo de aperfeiçoamento desta ferramenta, por favor preencha o formulário existente na parte inferior desta página. Obrigado.

A tradução oferecida, obviamente, contém erros e inadequações, mas deve-se reconhecer que, talvez, seja mais importante o técnico ou o usuário ter acesso a essa informação rudimentar em português do que em inglês, ao menos enquanto a tradução em

⁵⁸ <http://support.microsoft.com/kb/189139/pt-br>

português não está disponível. Um exemplo do texto exibido quando se acessa a página em que esse aviso é exibido está a seguir:

É necessário um sistema hospedar Windows NT para compilar e depuração o CEPC sistema operacional e aplicativos. Você pode baixar e depuração um aplicativo com CEPC após compilação um aplicativo usando Visual C++ 5.0 e o Windows CE Embedded Toolkit para Visual C++ 5.0 no sistema hospedar. Para fazer isso, você precisará conectar as máquinas hospedar e destino usando um NULL cabo modem via portas COM das duas máquinas. O procedimento é descrito abaixo.

Rafael Guzmán (2007), coordenador do setor de tradução automática da Symantec Corporation para espanhol, considera que se pode concordar que as pessoas que lêem manuais e guias de usuário podem tolerar um certo nível de língua “artificial”, desde que ela seja inteligível, precisa e gramaticalmente correta. Sendo assim, esse tipo de “mercado” acredita que as correções a serem feitas na fase de edição por tradutores humanos podem ser reduzidas à correção de erros gramaticais e de sintaxe (concordância de número e gênero), erros de pontuação e ortográficos (letras maiúsculas e minúsculas, por exemplo) e traduções consideradas erradas, por serem diferentes da terminologia usada pela empresa ou porque contêm uma interpretação equivocada do texto fonte. A título de exemplificação, apresentamos um trecho de um manual em que uma intervenção maior por parte do editor de uma tradução automática foi necessária (Exemplo 1).

Exemplo 1.:

Texto fonte:	You can connect through the Access Server to any platform that Product® supports, including Windows, Linux, and Mac OS X
Trad. automático:	Você pode conectar com o Servidor de acesso à plataforma qualquer que Product® suportar, incluindo Windows, Linux e ao Mac OS X.
Edição:	Você pode conectar-se com o servidor de acesso a qualquer plataforma que o Product® suportar, incluindo Windows, Linux e Mac OS X.

Esse trabalho de edição tem sido feito por tradutores, que recebem instruções específicas quanto aos tipos de alterações que devem fazer. Estas devem ser reduzidas a um mínimo, e alterações estilísticas são desaconselhadas. Isso porque a principal motivação para esse tipo de tradução é reduzir custos e tempo.

Ainda segundo ele,

O treinamento de editores humanos de traduções automáticas é crucial. É preciso estabelecer orientações claras para que o escopo da edição manual

esperada seja definido. Conseguir o equilíbrio certo é um desafio, pois não se pode esperar que editores de traduções automáticas produzam textos com a mesma qualidade lingüística de tradutores humanos, e nem a má qualidade pode ser aceita (p. 3).

Não entraremos em detalhes acerca dos métodos e das tecnologias de tradução automática. Apenas apresentamos aqui, com fins de contextualização, os métodos levantados por Specia & Rino (2002). Segundo as autoras, que estão embasadas em Dorr *et al.* (2000), dois tipos de informação podem ser utilizados para classificar um sistema de TA: seu método e seu paradigma. “Os métodos referem-se à organização global do processamento e de seus vários módulos, enquanto os paradigmas referem-se aos componentes de representação de conhecimento que auxiliam o projeto de processamento global”. (p. 9)

Os métodos são classificados como diretos e indiretos. Diretos quando não são utilizadas representações intermediárias com o mínimo de processamento lingüístico. É o caso do Systran. Nos métodos indiretos, a análise da língua de partida e a geração da língua de chegada são processos independentes. As sentenças a serem traduzidas são primeiramente transformadas em uma representação em uma linguagem intermediária e, depois, são geradas as sentenças-alvo. No método indireto, a tradução automática pode ser por transferência ou por interlíngua.

Segundo Specia & Rino (2002) os sistemas de tradução automática desenvolvidos de acordo com esses diferentes métodos podem se basear em diferentes tipos de conhecimento, o que determina o tipo de paradigma que seguem. Se basearem-se em conhecimento profundo, fundamental ou lingüístico, considera-se que seguem o *paradigma fundamental*. Isso equivale a dizer que “empregam teorias lingüísticas bem definidas, utilizando restrições sintáticas, lexicais ou semânticas” (p. 16). As referidas autoras apresentam seis tipos de conhecimento que caracterizam modelos que seguem o paradigma fundamental: tradução automática baseada em regras, em conhecimento, em léxico, em restrições, em princípios e o método *shake and bake*, que utiliza regras de transferência.

Já o paradigma empírico merece nossa maior atenção, porque o sistema que pretendemos desenvolver segue uma modalidade desse paradigma. Esses paradigmas (quase) não utilizam teorias lingüísticas no processo de tradução e têm sido bastante explorados dado “o grande avanço de hardware e à disponibilidade crescente de recursos eletrônicos significativos (dicionários, corpora de textos bilíngües e monolíngües, etc.) componentes essenciais para o sucesso da investigação empírica” (Specia & Rino, 2002, p. 19).

No paradigma empírico estão os sistemas de tradução automática baseados em estatísticas (conhecido como SBMT, do inglês *Statistical-Based Machine Translation*), os baseados em exemplos (EBMT, *Example-Based Machine Translation*), os baseados em diálogo (DBMT, do inglês *Dialogue-Based Machine Translation*), e os baseados em redes neurais (NBMT, *Neural-Based Machine Translation*).

2.1.1. A EBMT (*Example Based Machine Translation*, tradução automática baseada em exemplos)

A EBMT é essencialmente uma tradução por analogia: dada uma sentença na língua fonte e uma coleção de pares alinhados de texto fonte/destino, deve ser encontrado o trecho que melhor corresponde à sentença fonte na porção da coleção de textos fonte e aceito o correspondente da porção da coleção de textos de destino como a tradução (Brown, 1996). A tradução é feita por analogia com exemplos coletados em traduções já realizadas, em um cópús bilíngue alinhado. Esse é o ponto em que se assemelha ao sistema que vislumbramos.

Specia & Rino (2002) esclarecem que um algoritmo de unificação é usado para encontrar os exemplos mais próximos da sentença de entrada, a partir do cópús bilíngue. “A proximidade de cada exemplo com a sentença de entrada é determinada pela distância semântica entre as suas palavras, a qual pode ser calculada com base na distância entre essas palavras em uma hierarquia de termos e conceitos provida, em geral, por um *thesaurus* ou uma ontologia” (p. 20)

As autoras esclarecem ainda que a combinação do paradigma EBMT com abordagens linguísticas permite diminuir o tamanho do conjunto de dados necessários para o bom resultado da tradução automática baseada em exemplos.

O *Pangloss* (Brown, 1996) é um exemplo de sistema que utiliza o paradigma EBMT. Ele foi desenvolvido no *Center for Machine Translation* da Carnegie Mellon University (Filadélfia, EUA).

O *Pangloss* é um sistema de tradução automática composto de vários “motores” (*engines*). Esses motores de tradução são executados em paralelo para que, no final, uma tradução para as diferentes porções do material de entrada seja selecionada em um modelo estatístico. O PanEBMT é um desses motores e ele praticamente não necessita conhecimento estrutural das línguas envolvidas, simplesmente um grande cópús paralelo de sentenças, que

servem como exemplos, e um dicionário bilíngue. Os textos a serem traduzidos são segmentados em sequências de palavras que ocorrem no córpus, para os quais são determinadas traduções em um alinhamento de segmentos que constam nos pares das sentenças alinhadas. Essas traduções parciais são combinadas com os resultados de outros mecanismos de tradução para formar o produto final – o texto traduzido.

As três principais fontes de conhecimento que ele utiliza são: um córpus bilíngue paralelo alinhado sentença a sentença, um dicionário bilíngue e uma lista de sinônimos/raiz na língua de destino. Uma quarta fonte de conhecimento (opcional e menos importante) são: informações linguísticas específicas fornecidas no arquivo de configuração, uma lista de tokenizações que equacionam as palavras em classes como dias das semanas, uma lista de palavras que podem ser omitidas durante o alinhamento (como artigos), e uma lista de palavras que podem ser inseridas (Brown 1996).

Segundo Brown (1996), na EBMT, as traduções são produzidas em duas fases: na primeira, agrupamentos (*chunks*) são buscados procurando-se ocorrências de palavras consecutivas no texto de entrada. Na segunda fase, é feito um alinhamento sub-sentencial de cada par de sentenças encontrado na primeira fase para determinar a tradução do agrupamento.

Embora não estejamos entrando em maiores detalhes desse tipo de tradução automática, é importante ressaltar os pontos desse parâmetro que poderiam ser utilizados em nosso sistema:

- o uso de um córpus de sentenças alinhadas em um sistema automático; e
- a noção de trabalho com agrupamentos de palavras.

A diferença é que esses sistemas de tradução automática usam pouco, ou nenhum, uso de conhecimento linguístico. Na ferramenta que propomos, o conhecimento linguístico encontra-se nos princípios de construção do banco de dados, que é embasado na ideia das associações, ainda que, em larga escala, essas associações venham a ser levantadas automaticamente sem análise qualitativa.

Existem trabalhos que fazem uma ligação entre memórias de tradução e EBMT, como o descrito em Carl e Hansen (1999). O trabalho relata experimentos que comparam o resultado de traduções de três sistemas de TA baseados em córpus e os benefícios de vincular

diferentes estratégias de TA, como a baseada em memórias de tradução e EBMT. As tecnologias de memórias comparadas foram de STM (String-Based Translation Memory) e LTM (Lexeme-Based Translation Memory), sendo que o sistema de STM, por ser menos generalizante, demonstrou mais precisão de tradução quando segmentos próximos foram encontrados no banco de dados. No entanto, quando o corpus de referência não continha nenhuma tradução similar, o sistema de EBMT apresentou o melhor desempenho dos três. Esses autores também se propuseram a verificar se os sistemas de TA podem contribuir para um melhor desempenho dos sistemas de memórias, permitindo que o sistema baseado em exemplos (EBMT) forneça sugestões de tradução caso o sistema de memória não apresente nenhum resultado.

2.2. CAT Tools

Bowker (2002) inclui na categoria de CAT tools sistemas como: software de OCR (reconhecimento de caracteres, como os usados para transformar textos escaneados em formato digital); software de reconhecimento de fala; software de análise de corpus; sistemas de memória de tradução e gerenciamento de terminologia; ferramentas de localização e tradução de software; ferramentas de contagem de palavras; ferramentas que estudam a relação custo/benefício.

Nosso foco será nos sistemas de memória de tradução.

2.2.1. Memórias de Tradução

A tecnologia de Memória de Tradução tem se mostrado um grande sucesso na comunidade de tradutores humanos. Ela é inspirada na observação de que o trabalho do tradutor, ao mesmo tempo em que é essencialmente uma atividade criativa, também envolve trabalho repetitivo e entediante. A ideia por trás de um sistema desses é armazenar sentenças como foram traduzidas e depois reutilizá-las para produzir novas traduções. (Pekar *et al*, 2007). Dependendo dos tipos de documentos, da consistência de como foram redigidos e do software utilizado, as ferramentas de memória podem aumentar os níveis de produção em até 50% (Heuberger, 2005 *apud* Bowker, 2002).

Os programas conhecidos como Memória de Tradução dividem o texto em segmentos, que são frases ou partes de frases. O programa oferece ao tradutor a possibilidade de

visualizar um segmento do texto fonte e, logo em seguida, há um espaço onde a tradução deve ser inserida. O programa, então, armazena o par de segmentos em uma espécie de tabela, associando o texto fonte e o texto traduzido. Esses segmentos ficam todos armazenados na memória e, em uma próxima vez que aparecer um segmento semelhante ao já traduzido, o programa identificará a equivalência com a tradução anterior e a indicará como sugestão. O tradutor pode utilizar essa sugestão ou não. Também é possível pesquisar nesse banco de segmentos armazenados como alguma palavra já foi traduzida em trabalhos anteriores. Os programas mais avançados, de segunda e terceira gerações, também permitem reconhecer segmentos aproximadamente iguais.

Esses programas também oferecem recursos de terminologia, pois permitem a criação de glossários e reconhecem nos textos os termos inseridos nesses glossários, fazem análises estatísticas para avaliar quantas palavras precisam ser traduzidas, quantos segmentos podem ser aproveitados da memória, permitem fazer a conversão de diversos formatos de arquivos para sua interface de trabalho, permitindo ao tradutor trabalhar com esses formatos (Word, Excel, PowerPoint, FrameMaker, HTML, Interleaf, AmiPro etc.) sem necessidade do software respectivo, entre outros recursos.

É preciso esclarecer que existem os sistemas que permitem construir e utilizar as memórias, e arquivos independentes que são chamados de *memórias de tradução* (criadas nesses sistemas). As memórias não são mais do que arquivos que armazenam pares bilíngues de sentenças ou segmentos de texto fonte e texto alvo, visando um futuro reaproveitamento de todo o segmento ou de parte dos mesmos; e os sistemas são os aplicativos que permitem armazenar os pares e utilizá-los. O tradutor pode ter diferentes memórias de tradução (ou receber memórias de outros tradutores) conforme cliente ou área a ser traduzida.

Considera-se que esses sistemas, no entanto, ainda têm muito o que melhorar, uma vez que não têm acesso, por exemplo, ao significado do texto traduzido e operam em um nível mais superficial. O que acontece é que não conseguem identificar sentenças com o mesmo significado e já traduzidas, mas com uma estrutura sintática diferente ou expressa por meio de diferentes sinônimos. Conforme nos apresenta Pekar *et al* (2007), o fato de uma mesma semântica poder ser expressa por meio de diferentes formas linguísticas tem algumas consequências importantes para o uso prático das TMs (*Translation Memories*, Memórias de tradução). Os referidos autores resumem essas consequências em cinco diferentes itens:

- Os sistemas de TMs atuais oferecem ganhos perceptíveis apenas quando há uma grande quantidade de traduções prévias disponíveis que permitam a correspondência entre texto fonte e alvo no nível superficial;
- Devido a métodos de correspondência pouco flexíveis, o mecanismo de recuperação não é preciso o suficiente, selecionando ou muitos ou poucos candidatos, que também não são ordenados de maneira apropriada;
- A grande maioria dos sistemas de memórias opera apenas no nível da sentença, e não o ultrapassam. A reutilização de subpartes é parcialmente implementada por alguns sistemas, como o Similis, que veremos na seção de TMs de segunda geração;
- As TMs aplicam-se apenas a tipos limitados de documentos, os que são repetitivos, como documentos técnicos, em que há muitas correspondências literais;
- Os métodos atuais de recuperação que indexam traduções já realizadas exigem muito espaço em disco o que afeta a velocidade da recuperação.

Dados esses problemas apontados, têm sido propostos sistemas conhecidos como Memórias de Tradução de segunda e terceira gerações, que são mais desenvolvidos linguisticamente. A seguir, apresentamos o que se desenvolveu nesse sentido por alguns grupos de estudos e empresas.

2.2.1.1. Memórias de tradução de segunda geração

De acordo com Planas (2005), os sistemas de Memória de Tradução de primeira geração trabalham em primeiro nível: informação. Eles vêem o texto como um código. O SIMILIS⁵⁹, mostrado na Figura 45, é um sistema considerado de segunda geração. Ele oferece a possibilidade de combinar dois tipos de mecanismos para funcionar no nível sintático-léxico-morfológico. O software contém um léxico monolíngue para cada língua processada e algoritmos que permitem que ele analise e identifique categorias gramaticais de palavras individuais (pois o software agora trabalha com palavras e não com código), agrupando essas categorias de palavras em frases. Esse novo recurso oferece uma ferramenta de suporte à

⁵⁹ Desenvolvido pela Lingua et Machina <http://www.lingua-et-machina.com/>

tradução mais eficiente. O exemplo citado é que o sistema pode reconhecer que um grupo de palavras (por exemplo, um grupo nominal) é separado de outro grupo (por exemplo, um grupo verbal), e uma tradução é oferecida para um dos grupos como uma tradução possível.

Esse sistema é baseado na noção de *chunk* (ou agrupamento), e as unidades de tradução são, então, reduzidas a um nível dentro da sentença, e não a sentença inteira. Conforme já comprovado (Planas, 2000), a redundância, que é uma noção fundamental nos sistemas de memória de tradução, é muito maior do que a redundância de frases, o que equivale a dizer que é mais provável que um sistema encontre em sua memória uma similaridade de agrupamentos de palavras do que frases inteiras. Nós também acreditamos essa ser uma verdade, e essa abordagem de trabalharmos com associações de palavras, e não com fraseologias fixas, carrega essa noção também, a de que é mais fácil ajudar o tradutor que usa uma ferramenta trabalhando no nível de segmentos de palavras do que no de frase inteira, como os sistemas de memória mais comercialmente usados.

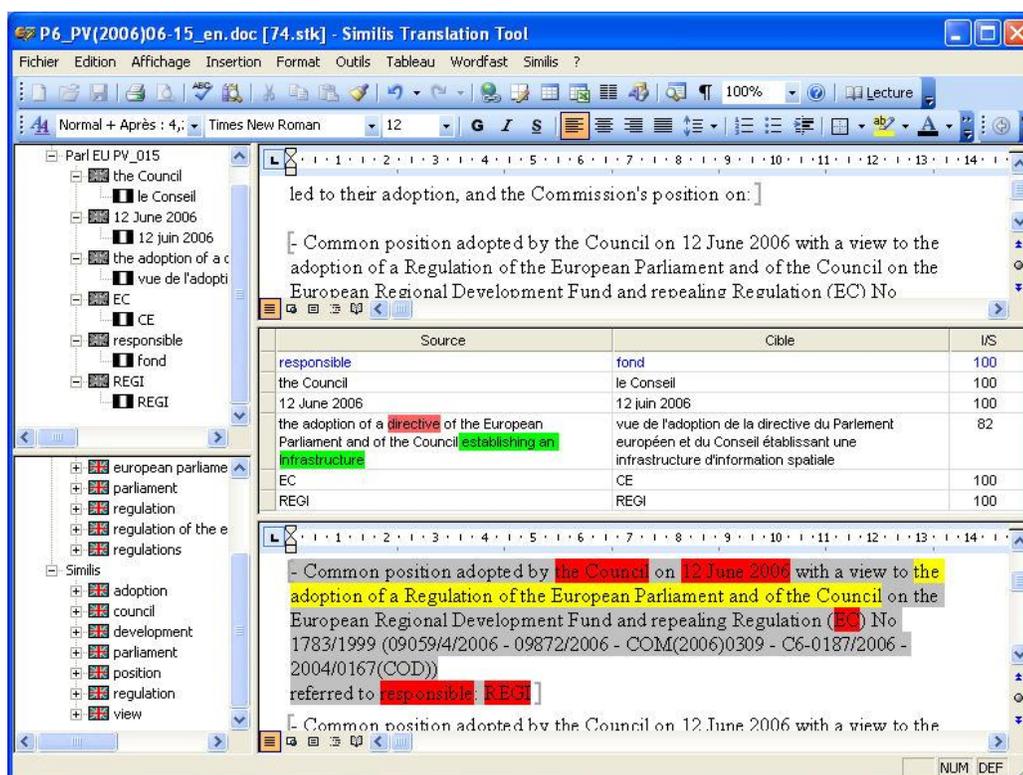


Figura 45 – Tela do Similis

Acreditamos que a tecnologia por trás do Similis possa ser de auxílio para o desenvolvimento da ferramenta pretendida nesta tese. Porém, cabe ressaltar que a ferramenta que vislumbramos não trabalha no nível de identificação de qualquer sub-segmento ou agrupamento de palavras, mas sim apenas daquelas associações específicas, conforme

identificadas no capítulo anterior. A identificação de associações específicas pode ser útil ao tradutor iniciante, pois o ajudará a identificar que tal associação é frequente e típica do tipo de texto com o qual ele está trabalhando.

2.2.1.2. Memórias de tradução de terceira geração

Essa denominação vem sendo usada por pesquisadores da universidade de Wolverhamptom, que têm o objetivo de tentar fazer uma identificação de segmentos no nível semântico. Eles trabalham com o fato de que frases que expressam a mesma semântica podem ter uma ampla variedade de formas. O algoritmo que estão desenvolvendo leva em conta três tipos importantes de variações nas línguas naturais: sintáticas, léxicas e léxico-sintáticas (Pekar, 2007). Não conseguimos detectar como está o desenvolvimento atual desse tipo de ferramenta. A princípio, não houve um maior desenvolvimento desse sistema e ele não foi disponibilizado para uso ou testes.

CAPÍTULO VIII

A FERRAMENTA

A segunda questão de pesquisa que nos colocamos é:

Uma vez identificadas tais associações, é possível transpô-las para uma ferramenta computacional capaz de armazená-las e posteriormente, na situação de uma tradução nova, identificar tais associações para o tradutor e disponibilizar exemplos de como já foram traduzidas?

A nossa intenção é a de buscar soluções alternativas a recursos como gramáticas, glossários e dicionários, usando a tecnologia para apresentar resultados que são mais facilmente encontrados em textos semelhantes ao que se está traduzindo. Uma parte da ideia de construir essa ferramenta surgiu da própria observação de como fazíamos nossas pesquisas no Google, digitando no campo de busca palavras que "andam juntas" e investigando em que posições eram mais frequentes.

Nossa ideia não é apresentar um sistema de memória de tradução de "quarta geração". De uma forma mais modesta, queremos apenas apresentar as bases de um aplicativo que venha a auxiliar os tradutores iniciantes e estudantes em seu ingresso no mercado da tradução de artigos médicos. De acordo com Rodrigues (2002), alguns autores observaram que "tradutores em formação têm um certo medo de enfrentar a situação de traduzir um texto e possuem pouco senso crítico sobre as suas próprias atitudes" (p. 49). Rodrigues (2002) prossegue afirmando que muitas vezes esses tradutores iniciantes procuram uma "confirmação externa" para as soluções que encontram para problemas de tradução, e que tal confirmação é muitas vezes o dicionário, que nem sempre apresenta as melhores

soluções. Nosso intuito com a ferramenta é, justamente, alertar para que estejam atentos a um determinado aspecto da tradução dos artigos, as associações típicas que os textos contêm, e que seja uma alternativa ao dicionário ou à pesquisa aleatória na Web.

Pekar e Mitkov (2007) já constataram que mesmo a identificação de sub-segmentos semelhantes ao texto a ser traduzido e não apenas segmentos completos, quando disponibilizados ao tradutor, proporcionam grande auxílio, uma vez que é mais provável, em determinados tipos de textos, que apenas pequenos segmentos tenham similaridades com outros textos já traduzidos. Com base nessa constatação, a ferramenta que vislumbramos vale-se da crença de que a apresentação de trechos de textos que contêm um mesmo conjunto de palavras ocorrendo no Texto de Entrada (o texto novo a ser traduzido) e no Texto da Base (o texto do córpis alinhado) pode ser de auxílio ao tradutor. A seguir, descrevemos o desenvolvimento inicial desse sistema, que batizamos de “ASSIM” (ASsociações SIntagmáticas em Medicina).

1. Construção dos bancos de dados

O sistema ASSIM é composto de um banco de dados formado por quatro tabelas: uma contém o que foi denominado grupo sintagmático (*concgrams* de duas palavras), a segunda contém as palavras que se associam ao grupo (são as palavras que se encontram na terceira coluna dos *concgrams* de três palavras), a terceira contém o córpis alinhado e a quarta tabela relaciona as frases alinhadas à seção do artigo em que as mesmas se encontram. Para o armazenamento dos dados no banco, foi usado o Microsoft SQL Server Express Edition, e o sistema foi desenvolvido no ambiente Microsoft Visual Studio Express, ambas versões disponibilizadas gratuitamente pela Microsoft. O desenvolvimento computacional do protótipo da ferramenta foi desenvolvido pelo Analista de Sistemas Felipe Spencer Picada.

A primeira tabela é composta pelos *concgrams* de duas palavras (o grupo sintagmático) e a segunda tabela pelas outras palavras que se associam a essas duas. Como se pode ver nos quadros abaixo, as palavras da coluna 4 e 5 são sempre uma recombinação das palavras que estão na coluna 3. Esse dado é importante para entender por que apenas as palavras da coluna 3 foram incluídas na tabela de associações. Mostramos o exemplo a seguir, usando a palavra *avaliar*, que é uma das palavras que se associa a *objetivo*:

Quadro 41 – Exemplo da formação dos *congrams* de 2, 3, 4 e 5 palavras para objetivo em coocorrência com avaliar

1	2
objetivo	avaliar

1	2	3
objetivo	avaliar	estudo
		deste
		foi
		presente
		crianças
		energético
		níveis
		tem
uso		

1	2	3	4
objetivo	avaliar	deste	estudo
			foi
		estudo	foi
			presente

1	2	3	4	5
objetivo	avaliar	crianças	estudo	uso
		deste	estudo	foi
				níveis
			uso	
			foi	uso
		estudo	foi	presente
				uso

Para construir o banco de palavras-nodo e seus *congrams*, fizemos o levantamento dos *congrams* de todas aquelas palavras que detectamos como exclusivas de cada seção, utilizadas também na Etapa 3 das análises. Temos, portanto, um banco de dados com os *congrams* de 2, 3, 4 e 5 palavras das seguintes palavras-nodo:

Quadro 42 – Palavras-nodo para as quais foram processados os *congrams* do banco de dados

INTRODUÇÃO	MÉTODOS	RESULTADOS	DISCUSSÃO
alterações	amostras	apresentaram	achados,
amamentação	análise	apresentavam	associação
aumento	após	características	autores
avaliar	através	curva	dieta
clínica	coleta	diferença	literatura
clínicos	comitê	distribuição	maior

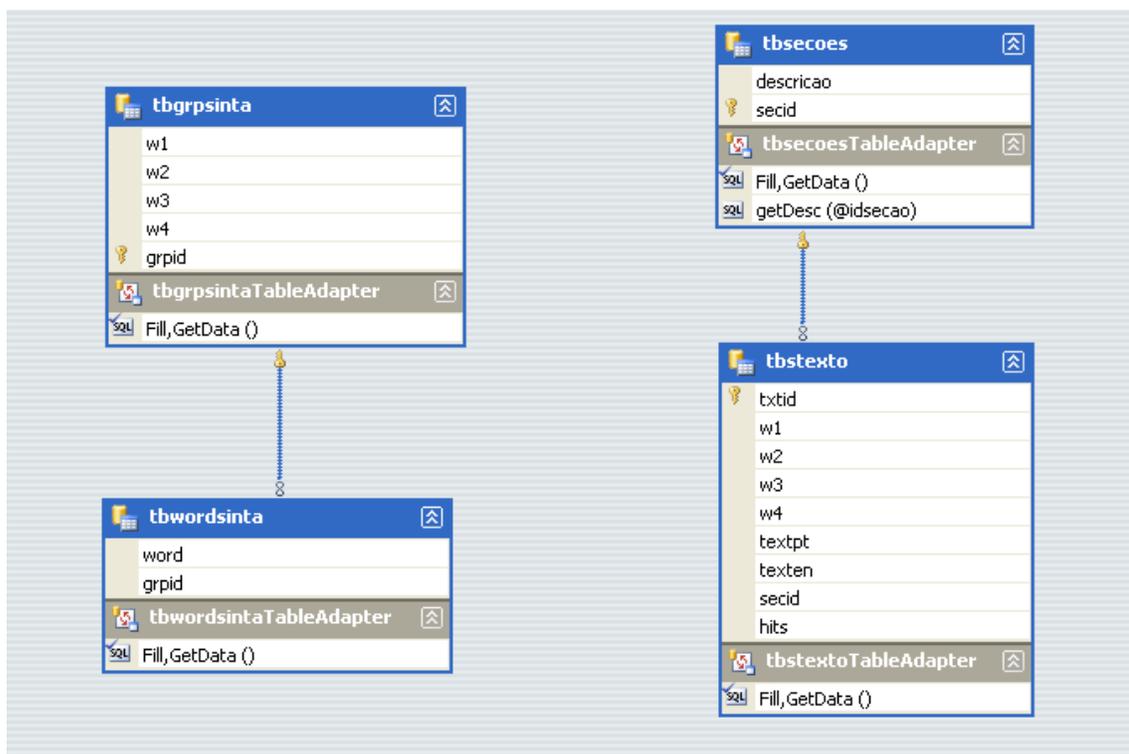
desenvolvimento	critérios	estatisticamente	menor
infância	ética	feminino	neste
lesão	exame	figura	níveis
mortalidade	período	grau	pode
neonatal	pesquisa	houve	presente
nutricional	protocolo	masculino	resultados
obesidade	quadrado	média	
objetivo	realizada	mediana	
óssea	realizado	médias	
pré	significância	meninas	
saúde	teste	meses	
terapia	utilizado	nascimento	
termo	utilizando	respectivamente	
		tabela	
		valores	

A terceira tabela contém o cópys do Jornal de Pediatria, que foi alinhado por Robert Coulthard na ocasião de elaboração de sua dissertação de mestrado (Coulthard, 2005). No entanto, o cópys estava alinhado frase a frase, ou parágrafo a parágrafo, por artigo. Foi preciso, então, reajustar essa situação a nossa realidade. Primeiramente, excluímos do cópys alinhados aqueles artigos que não eram artigos originais, ou seja, as cartas ao editor, as revisões e os estudos de caso. Depois, repetimos o mesmo passo que foi necessário com o cópys em português: separamos o texto dos artigos em quatro arquivos, um para cada seção dos artigos. Essa separação foi feita manualmente para os 120 artigos que compõem o cópys exclusivo de artigos originais. Como resultado, obtivemos 4 planilhas do excel, cada uma contendo apenas os trechos correspondentes a mesma seção: *Introduções*, *Métodos*, *Resultados* e *Discussão*. Explicitando ainda mais, cada planilha contém 120 *Introduções*, 120 seções *Métodos*, etc. Os trechos estão alinhados em duas colunas, uma para o português e outra para o inglês. Outro processo manual que fizemos foi separar os trechos que estavam alinhados por parágrafo para um alinhamento em frases dos segmentos maiores. A seguir, algumas células para demonstrar o formato do cópys alinhado:

Quadro 43 – Exemplo do formato do corpúsculo alinhado

Apesar dos avanços nas taxas de aleitamento materno observados na última década, a situação do aleitamento materno no Brasil ainda está longe da preconizada pela Organização Mundial da Saúde: amamentação exclusiva até o sexto mês e aleitamento materno com alimentos complementares até dois anos ou mais de idade (1).	Despite the growing tendency of breastfeeding practice in Brazil, we are far from meeting the WHO recommendations, which determine exclusive breastfeeding until the sixth month of life and continued breastfeeding with complementary foods up to the second year of life or later (1).
Estudo populacional realizado em 1995, no município onde este estudo foi desenvolvido, detectou um padrão insatisfatório de aleitamento materno; aos trinta dias, somente 29% das crianças eram exclusivamente amamentadas, e aos quatro meses, apenas 4,6% (2).	A population-based study carried out in 1995 revealed an unsatisfactory breastfeeding pattern: at 30 days of age, only 29% of the infants were exclusively breast-fed, and at four months of age this was equivalent to 4.6% (2).
A duração do aleitamento materno pode ser favorecida ou restringida por fatores biológicos, culturais, relativos à assistência à saúde e socioeconômicos (3).	The length of breastfeeding may be favored or restricted by biological, cultural, healthcare-related and socioeconomic factors (3).
Os profissionais de saúde, por meio de suas atitudes e práticas, podem influenciar positiva ou negativamente o início da amamentação e sua duração.	Health professionals, by means of their attitudes and practices, may positively or negatively influence the beginning of breastfeeding and its length.

A seguir, apresentamos um esquema com a visão do banco de dados, com os campos de cada tabela:

**Figura 46** – Esquema com a visão do banco de dados

Nos dois primeiros quadros à esquerda, a relação que existe é a de que 1 grupo composto por 2 palavras (há espaço para 4, mas utilizamos apenas 2) pode ter N outras palavras (da tabela tbwordsinta) associadas, assim como uma seção poderá conter várias frases alinhadas, como mostram os dois quadros da direita.

2. A composição do protótipo

Para este protótipo, incluímos na tabela do córpis alinhado somente as seções de *Introduções* e *Métodos*. Também trabalhamos apenas com os grupos de palavras associadas para uma palavra da seção *Introdução: objetivo* e uma para a seção *Métodos: amostras*. Apenas para ter uma ideia da quantidade de dados que foram gerados, somente a palavra *objetivos* associa-se com 35 outras palavras e cada uma dessas 35 se associa a outros grupos de palavras. Por exemplo: *objetivo* associa-se com *avaliar* e essa associação está associada ao grupo de palavras da coluna 3 do quadro acima. Devido a essa enorme quantidade de dados gerada, optamos por não incluir outros grupos de palavras neste protótipo. Para ter uma ideia de quantas associações resultam da investigação de uma palavra, há uma lista com todos os *congrams* (de 2, 3, 4 e 5 palavras para a palavra-nodo *objetivo*) no Anexo IV.

3. Funcionamento do sistema

Para o usuário, o sistema funciona da seguinte forma: primeiramente, o usuário cola no campo apropriado um trecho do texto que deseja pesquisar, preferencialmente uma frase. A seguir, o software analisa se esse novo trecho contém palavras que foram cadastradas no banco de dados de associações (*congrams*). Em caso afirmativo, essas palavras são ressaltadas em amarelo. A seguir, ao pressionar o botão *Processa*, o software busca, no banco de dados, formado pelo córpis paralelo, todos os trechos do córpis que contém essa mesma associação em português. Como o córpis está alinhado, o software mostra, além do trecho em português, a sua tradução, a qual pensamos, servirá de inspiração ou de ponto de partida para a tradução que o consultante precisa fazer. Além disso, o consultante pode saber em que seção do artigo a associação contida na frase que precisa traduzir foi encontrada no córpis alinhado.

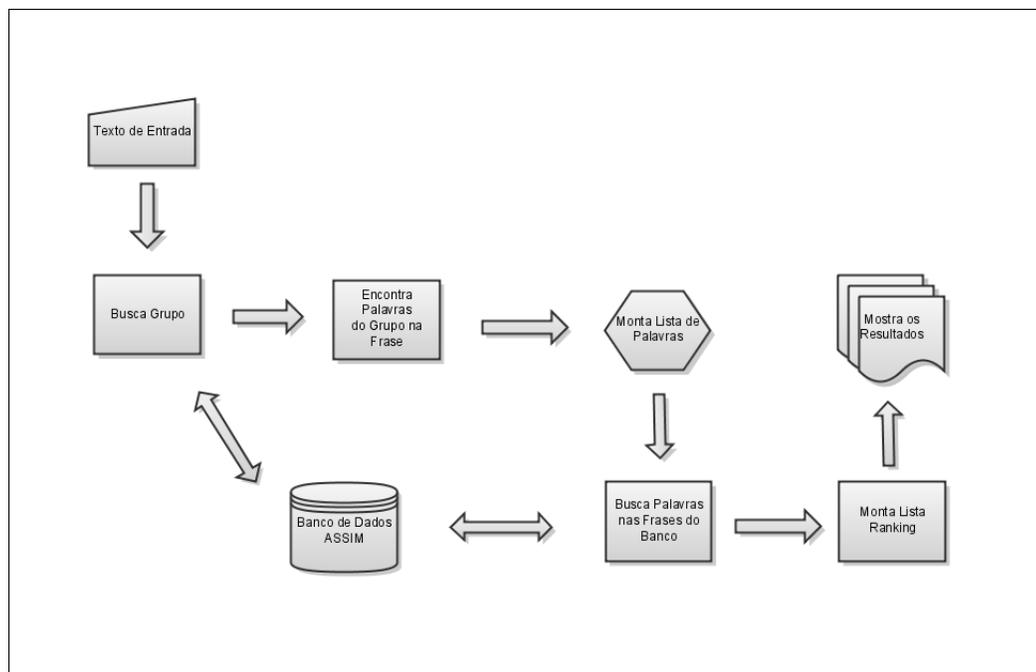


Figura 47 – Visão geral do funcionamento do sistema

A Figura a seguir mostra os principais campos da interface da ferramenta. O campo *Entrada* é onde é inserido o texto “novo”, o texto a ser traduzido; o campo PT-BR é onde é exibido o texto que o sistema encontra no cópús que contém os mesmos agrupamentos de palavras que o texto, inserido no campo *Entrada*; o campo EN mostra como está traduzido no cópús esse trecho identificado como semelhante ao que deve ser traduzido.

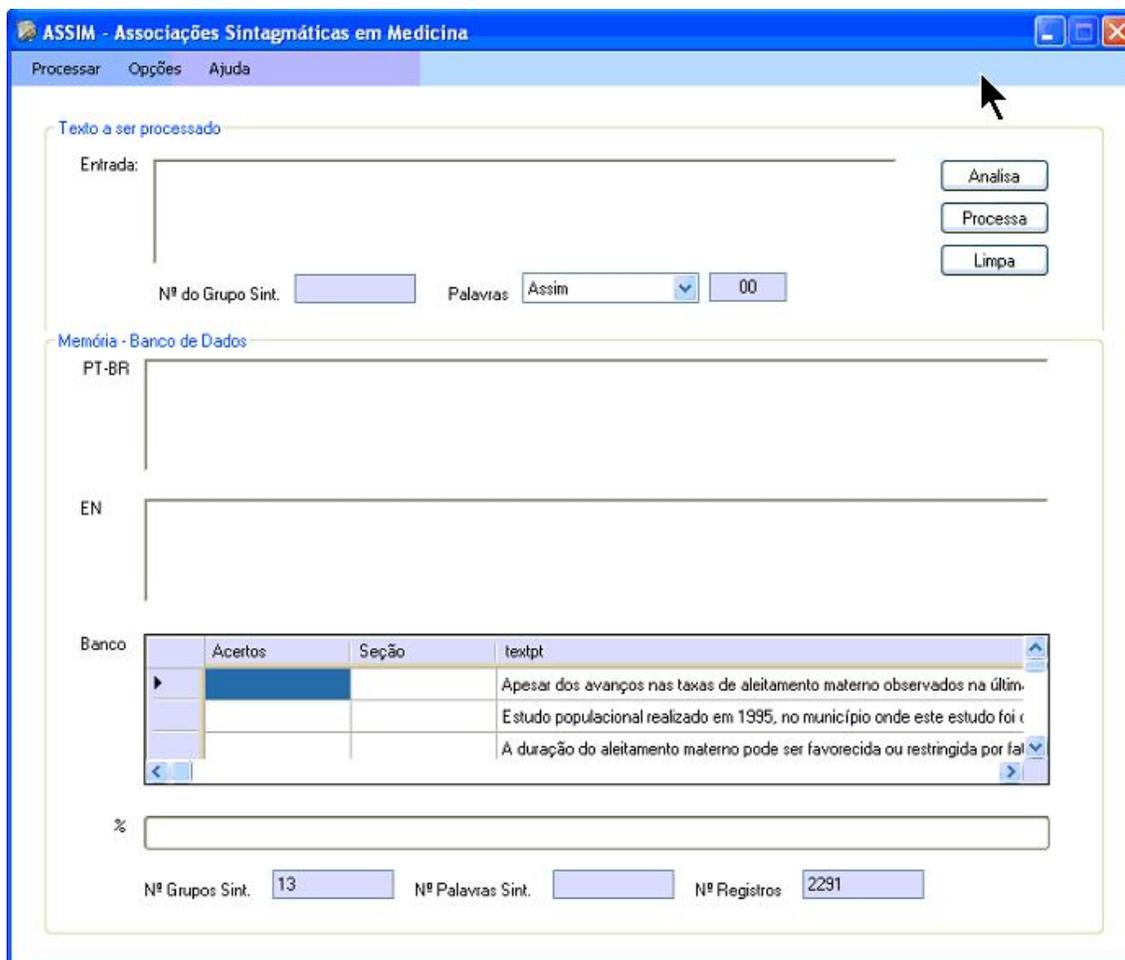


Figura 48: Tela do sistema ASSIM

A sequência de processamento configura-se da seguinte forma. O usuário insere o texto a ser traduzido no campo *Entrada* e clica no botão *Analisa*. A ferramenta, então, procura no banco de dados quais associações de palavras o TE (Texto de Entrada) contém que são cadastradas no banco como associações sintagmáticas relevantes. Tais palavras são sublinhadas, como mostra a Figura 49. No caso do exemplo a seguir, a associação é *amostras/antes*, associada ao grupo que contém as palavras *foram* e *sangue* como associadas.

Processar Opções Ajuda

Texto a ser processado

Entrada: As amostras foram coletadas antes do início da antibioticoterapia, juntamente com a coleta de sangue para exames de rotina.

Nº do Grupo Sint. 7 Palavras 4

Memória - Banco de Dados

PT-BR

EN

Banco

Acertos	Seção	textpt
		Apesar dos avanços nas taxas de aleitamento materno observados na últim.
		Estudo populacional realizado em 1995, no município onde este estudo foi c
		A duração do aleitamento materno pode ser favorecida ou restringida por fal

%

Nº Grupos Sint. 13 Nº Palavras Sint. Nº Registros 2291

Figura 49 – Tela do sistema ASSIM com uma frase a ser pesquisada, no momento em que o usuário clica em Analisa

A seguir, clica-se no botão *Processa*, e a ferramenta busca, no corpús alinhado, qual trecho de texto contém essas mesmas associações, como mostra a Figura 50.

Processar Opções Ajuda

Texto a ser processado

Entrada: As amostras foram coletadas antes do início da antibioticoterapia, juntamente com a coleta de sangue para exames de rotina.

Analisa
Processa
Limpa

Nº do Grupo Sint. 7 Palavras 4

Memória - Banco de Dados

PT-BR Foram obtidas amostras de sangue antes da administração da vacina para quantificação no sangue periférico, considerada basal, da carga viral do HIV e para a avaliação dos níveis de linfócitos T CD4+, através de sua contagem absoluta e porcentagem.

EN Blood samples were obtained before the administration of the vaccine for quantification of the HIV viral load in peripheral blood (baseline value), and for the evaluation of CD4+ T lymphocyte levels, through absolute count and percentage values.

Banco

Acertos	Seção	textpt
9	Métodos	Foram colhidas amostras de sangue da criança aos 6 meses (antes da aplic
6	Métodos	As amostras de LHP foram coletadas dos recipientes enviados à UTI neona
5	Métodos	Foram obtidas amostras de sangue antes da administração da vacina para c

%

Nº Grupos Sint. 13 Nº Palavras Sint. 5 Nº Registros 2291

Figura 50: Tela do sistema ASSIM apresentando um segmento de exemplo ao consultente, no campo EN

Outras frases do corpúis alinhado que apresentaram as mesmas palavras são apresentadas a seguir:

Memória - Banco de Dados

PT-BR As amostras de LHP foram coletadas dos recipientes enviados à UTI neonatal imediatamente antes da oferta ao RN; o leite havia sido, portanto, coletado, pasteurizado e congelado no banco de leite e, posteriormente, enviado ao lactário, sendo descongelado, aquecido, distribuído e fornecido à UTI neonatal. As amostras de leite que apresentavam floculação a olho nu foram descartadas. A homogeneização do material antes da coleta foi realizada com movimentos manuais, segundo o recomendado por Wang et al. (4).

EN The PHM samples were collected from the receptacles sent to the NICU immediately before being offered to the newborn infant; the milk, therefore, had been collected, pasteurized and frozen at the human milk bank and, at a later date, sent to the lactation center, being defrosted, heated, distributed and supplied to the NICU. Milk samples that presented coagulation to the naked eye were discarded. Before collection, homogenization of the material was performed with manual movements as recommended by Wang et al. (4).

Banco

Acertos	Seção	textpt
9	Métodos	Foram colhidas amostras de sangue da criança aos 6 meses (antes da aplic
6	Métodos	As amostras de LHP foram coletadas dos recipientes enviados à UTI neona
5	Métodos	Foram obtidas amostras de sangue antes da administração da vacina para c

Figura 51 – Tela do sistema ASSIM apresentando uma segunda opção com as mesmas associações

Memória - Banco de Dados

PT-BR Este estudo foi aprovado pelo Comitê de Ética do Hospital de Clínicas de Porto Alegre e um consentimento informado foi obtido dos pais ou responsáveis pelos pacientes. As amostras foram coletadas antes do início da antibioticoterapia, juntamente com a coleta de sangue para exames de rotina.

EN The study protocol was approved by the Ethics Committee at the Hospital de Clínicas de Porto Alegre and informed consent was obtained from patients' parents or guardians. The samples were collected before starting antibiotic therapy, simultaneously with blood for routine exams.

Banco

Acertos	Seção	textpt
5	Métodos	Foram obtidas amostras de sangue antes da administração da vacina para c
4	Métodos	Este estudo foi aprovado pelo Comitê de Ética do Hospital de Clínicas de P
3	Métodos	Antes do preenchimento do questionário padrão ISAAC (3), o termo de cons

Figura 52 – Tela do sistema ASSIM apresentando uma terceira opção com as mesmas associações

Memória - Banco de Dados

PT-BR Antes do preenchimento do questionário padrão ISAAC (3), o termo de consentimento livre e esclarecido era lido e explicado, até que não houvessem dúvidas por parte do responsável legal. Foram incluídas 742 (67,8%) crianças, que entregaram as amostras fecais para realização do exame parasitológico no laboratório de patologia clínica do Hospital Universitário Alcides Carneiro (HUAC) da Universidade Federal da Paraíba (Campina Grande, PB).

EN Before the ISAAC (3) standardized questionnaire was filled in the Free and Informed Consent form was read and explained until there were no further doubts on the part of the guardian. Seven hundred and forty-two (67.8%) children were included who had submitted fecal samples for the performance of the parasitological exam at the Clinical pathology Laboratory of the Hospital Universitário Alcides Carneiro (HUAC) of the Universidade Federal da Paraíba (Campina Grande, PB).

Banco

Acertos	Seção	textpt
4	Métodos	Este estudo foi aprovado pelo Comitê de Ética do Hospital de Clínicas de P
3	Métodos	Antes do preenchimento do questionário padrão ISAAC (3), o termo de cons

Figura 53 – Tela do sistema ASSIM apresentando uma quarta opção com as mesmas associações

O campo Banco apresenta todos os resultados de sentenças do córpus que contém as associações que estão presentes no texto de entrada. A ordem de apresentação é das frases que têm mais palavras que correspondem às associações. Salientamos que como o córpus não estava alinhado frase a frase em alguns casos, como no resultado apresentado na Figura 53, a ferramenta acaba apresentando exemplos em que as associações não estão nas mesmas frases. O alinhamento do córpus frase a frase, nos casos que ainda esteja alinhado parágrafo por parágrafo, deverá ser feito posteriormente para o melhor funcionamento do sistema.

4. Vantagens e limitações

A ferramenta que delineamos tem o propósito de demonstrar como a abordagem via associações sintagmáticas frequentes poderia ser vantajosa em relação aos sistemas de memória de tradução e, além disso e talvez principalmente, em relação a outros materiais de consulta para tradutores. Quando pensamos em quais são as vantagens ou benefícios da ASSIM, pensamos nessas duas vias: um ganho no número de resultados de opções de tradução e, além disso, no provável aprendizado que ela pode proporcionar.

Com relação a essa segunda vantagem, pensamos que, como o sistema mostra para o tradutor quais são as palavras que estão associadas, ele poderá ter acesso a um "conhecimento" que só adquiriria após traduzir vários desses textos médicos. A possibilidade de consultar as diferentes traduções para uma determinada associação também permite a reflexão e o aprendizado.

Com vistas a demonstrar os ganhos no número de resultados ou sugestões de tradução que a abordagem de associações poderia ocasionar, fizemos uma comparação dos resultados obtidos com a ferramenta ASSIM e o sistema de memória de tradução SDL Trados⁶⁰. Primeiramente, elaboramos uma memória de tradução com os textos do *cópus alinhado*, o mesmo que compõe o banco de dados da nossa ferramenta. Selecionamos duas frases de um artigo recente do *Jornal de Pediatria*⁶¹ as submetendo aos dois sistemas: ASSIM e Trados. Como essa é uma simulação e a ferramenta é um protótipo inicial, selecionamos frases que continham palavras que estavam inseridas no banco de dados. Selecionamos uma frase da *Introdução*, contendo a palavra *objetivo* e uma frase de *Métodos*, contendo a palavra *amostras*. As frases estão apresentadas abaixo:

O **objetivo** deste trabalho foi documentar o perfil de sorotipos e a sensibilidade aos antimicrobianos das cepas de pneumococo obtidas em espécimes clínicos de crianças de até 5 anos de idade atendidas com doença invasiva no HC-UFU e avaliar as implicações para a formulação das vacinas antipneumocócicas.

Os espécimes obtidos assepticamente foram devidamente processados e semeados em frascos para hemocultura (**amostras** de sangue) ou em placas de ágar chocolate e ágar sangue (demais tipos de amostra) o mais rapidamente possível após a coleta e imediatamente à chegada no laboratório.

⁶⁰ O SDL Trados é o software de auxílio a tradução mais utilizado por agências de tradução e tradutores freelancer, principalmente pelo recurso de Memória de Tradução. <http://www.trados.com/en/>.

⁶¹ *Prevalência de sorotipos e resistência antimicrobiana de cepas invasivas do pneumococo em crianças: análise de 9 anos*. <http://www.jpmed.com.br/conteudo/09-85-06-495/port.asp?cod=2027>

A seguir, os resultados que o Trados apresenta para a primeira frase. Ele encontra uma frase, com 59% de semelhança com a frase que selecionamos.

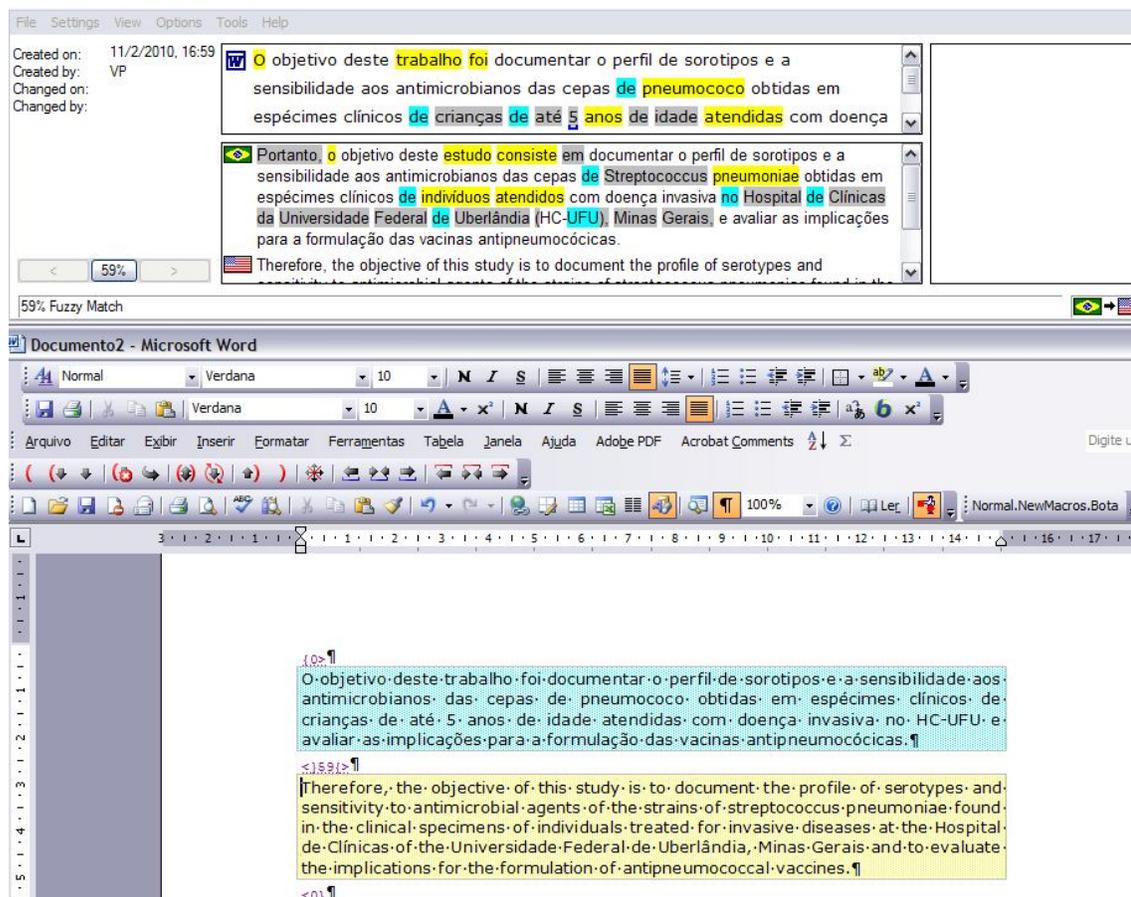


Figura 54 – Segmento encontrado pelo Trados na memória

FALTOU O ASSIM (20 ocorrências)

Já para a segunda frase, o Trados não encontra nenhuma correspondência na memória, que é formada com as frases do corpus alinhado, como mostra a figura abaixo.

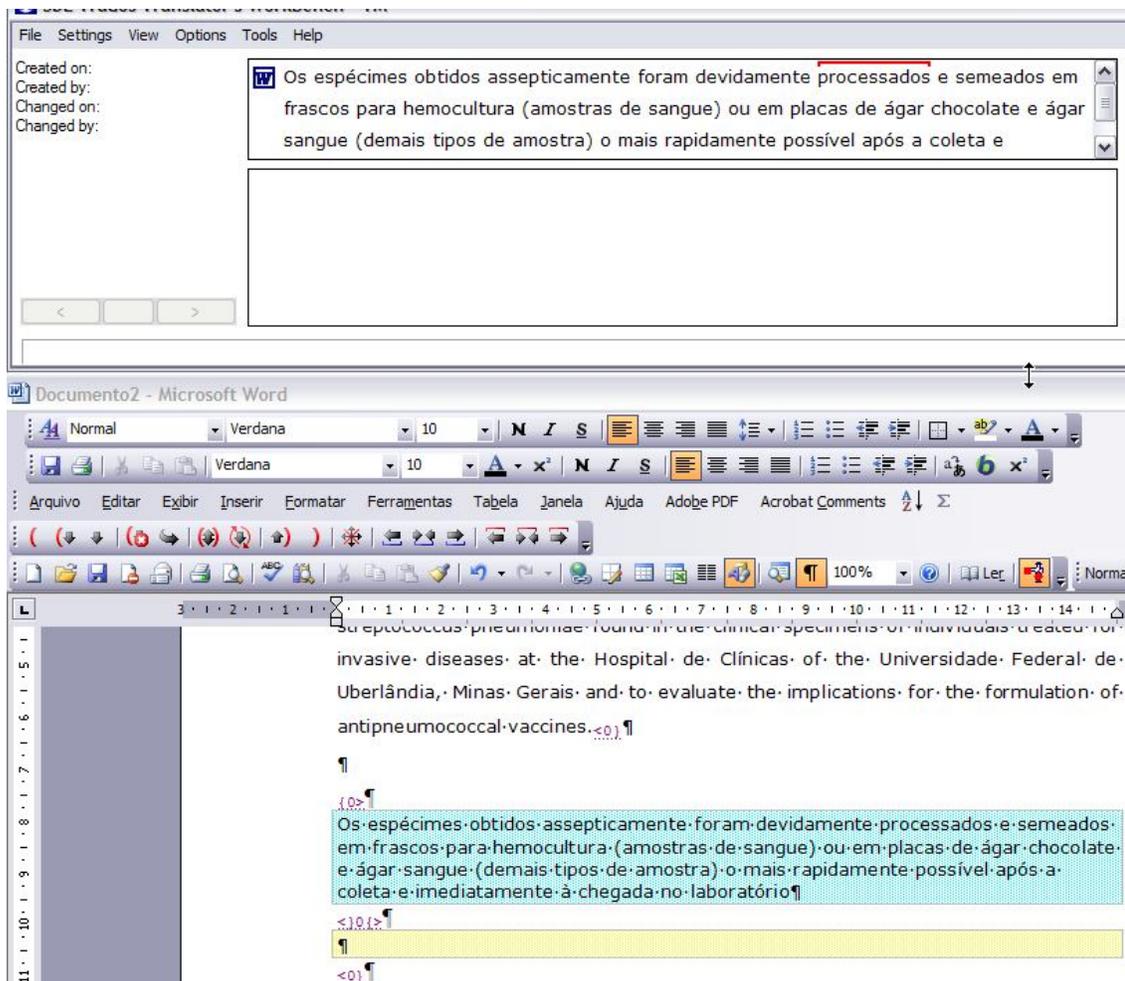


Figura 55 – Segmento encontrado pelo Trados na memória

Processamos agora as mesmas frases no sistema ASSIM. Para a primeira frase, que continha a palavra *objetivo*, a ferramenta apresentou 18 frases de exemplo, selecionamos algumas, mostradas nas figuras abaixo.

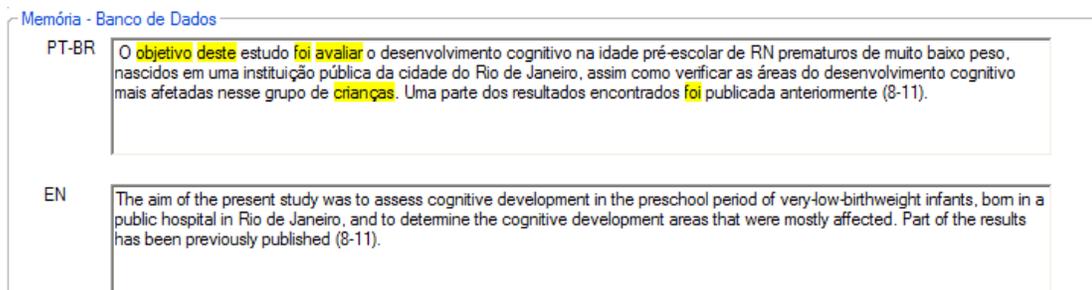


Figura 56 – Segmento encontrado pela ASSIM no banco

Memória - Banco de Dados

PT-BR O objetivo deste estudo foi avaliar a extensão do uso de medicamentos não apropriados para crianças em prescrições de uma unidade de tratamento intensivo pediátrica (UTIP) terciária, de acordo com os padrões estabelecidos pela FDA.

EN The objective of this study is to evaluate the extent of the use of drugs which are "not appropriate for children" in the prescriptions at a tertiary Pediatric Intensive Care Unit (PICU) according to standards established by the FDA.

Figura 57 – Segmento encontrado pela ASSIM no banco

Ainda a ferramenta indica que as frases foram encontradas, em sua maioria, na seção Introdução. Apenas um exemplo ocorreu na seção Métodos (ver Figura 59).

Memória - Banco de Dados

PT-BR Foi necessário desenvolver um questionário de coleta de dados, seguindo um roteiro preestabelecido composto de questões abertas e fechadas, uma vez que inexistem questionários padronizados para nossa população que permitam avaliar os aspectos específicos da ST que eram objetivo deste trabalho: a compreensão a respeito da síndrome, o impacto no momento do diagnóstico, o impacto sobre a vida atual e as expectativas de futuro.

EN A questionnaire had to be devised for data collection, following pre-established guidelines. The questionnaire included closed and open-ended questions, as no standardized questionnaires are available for this population that allow assessing specific aspects of TS, which were the aim of this study: improved knowledge about the syndrome, about the reaction at the time of diagnosis, about the impact on current life, and about future expectations.

Banco

	Acertos	Seção	textpt
	4	Introdução	O objetivo deste estudo consiste em avaliar aspectos clínicos e histológicos
	4	Introdução	O objetivo deste artigo é avaliar as condições nutricionais de um grupo de p
▶	4	Métodos	Foi necessário desenvolver um questionário de coleta de dados, seguindo u

Figura 58 – Indicação de que o exemplo pertence à seção Métodos

Observemos agora os resultados para a segunda frase selecionada. A ferramenta ASSIM encontra 8 frases, todas elas na seção Métodos. Lembramos que o sistema de memória de tradução não encontrou nem uma frase na memória formada pelo mesmo banco de dados que o que consta na ASSIM.

4.1. Limitações

No momento, identificamos como possíveis pontos de críticas o fato de a ferramenta estar sugerindo ao tradutor frases traduzidas, e não originalmente escritas em inglês. Quanto a isso, pensamos que as traduções se encontram publicadas por uma entidade respeitada, a Sociedade Brasileira de Pediatria, e em uma revista também reconhecida e indexada em

várias bases de dados. Além disso, propõe-se apenas que as traduções apresentadas sirvam de exemplo, não eximindo o tradutor de confirmar suas escolhas em outras fontes. Não podemos esquecer, também, que as memórias de tradução também funcionam com base em traduções anteriores, e que sempre é possível alterar a tradução que se está fazendo, usando as traduções prévias apenas como exemplos ou possibilidades de tradução. Também não investigamos a usabilidade do sistema e nem foi possível testá-lo em uma situação real, o que também não viria ao caso, pois estamos tratando de um projeto inicial para o qual elaboramos apenas um protótipo.

A seguir, tecemos as considerações a respeito desta segunda etapa desta tese.

CAPÍTULO IX

CONSIDERAÇÕES FINAIS – PARTE II

Nesta segunda etapa do trabalho, inicialmente, havia uma proposta e uma curiosidade muito grande de nos aproximarmos da área de Linguística Computacional e do Processamento da Linguagem Natural. Essa proposta, no entanto, acabou se modificando à medida que fomos entendendo e aceitando as limitações e os pontos fortes de ambas as áreas envolvidas. No que cabe a nós, enquanto linguistas, a maior dificuldade nessa tentativa de aproximação foi o desconhecimento das linguagens de programação e até mesmo do raciocínio abstrato que o desenvolvimento de uma ferramenta requer. Isso, de forma alguma nos desmotivou a prosseguir com o desenvolvimento do sistema pretendido, apenas entendemos melhor qual seria o nosso papel dentro desse contexto. Optamos por não nos tornarmos linguistas computacionais, mas por fazer o papel de um linguista, que levanta dados linguísticos a serem aplicados em um recurso computacional. Isso acaba sendo um pouco diferente do que desenvolver soluções em Linguística Computacional.

Essa aproximação entre linguistas e informatas vem sendo conseguida com maior ou menor êxito dependendo da “boa vontade” de ambos os lados. Maia (2008) considera que a cooperação não deve ser complicada se o linguista tiver uma teoria para testar e se o engenheiro computacional puder pensar em alguma maneira de testá-la, e conclui: “o que é necessário é boa vontade e tentativas sérias de ambos os lados de entender o ponto de vista um do outro” (p.21). Diana Santos (2009), em comunicação na Escola de Verão da Linguatca, afirma também que pode haver aqueles que possam atuar nos dois lados da disciplina igualmente, citando como exemplo como ela própria, cuja formação é em Computação. Particularmente, nos parece mais fácil o linguista computacional aprender

Linguística do que o contrário, mas de qualquer maneira, não nos parece que esse deva ser ponto nevrálgico da disciplina, alguns serão linguistas computacionais, outros serão linguistas e informatas trabalhando em colaboração. Por isso, ao fim e ao cabo, a nossa contribuição à ferramenta aconteceu no levantamento dos dados e na formulação do seu funcionamento e dos resultados que queríamos que o sistema apresentasse ao usuário.

A breve passagem que fizemos sobre tecnologias e tradução mostra que a área segue em desenvolvimento, tanto nos níveis acadêmicos como nos comerciais. A empresa que faz o SDL Trados, por exemplo, lançou em 2009 o SDL Trados Studio 2009 com um sistema patentado denominado AutoSuggest, que propõe *matches* de segmentos menores que frases conforme o tradutor digita, o que corrobora a ideia de que correspondências de segmentos menores são profícuas. Nosso protótipo também mostrou como os *matches* de um sistema de auxílio ao tradutor podem ser ampliados usando-se como base as associações sintagmáticas. A diferença é que não se trata de qualquer associação, mas sim daquelas que se mostram frequentes e significativas dentro de uma área especializada previamente examinada.

Constatamos também que a preparação dos dados a serem alimentados em um sistema computacional também requer bastante esforço e que um banco de dados pode requerer uma preparação trabalhosa dos dados linguísticos, por mais “brutos” que sejam. O processo de construção das listas de *concgrams* pode ser demorado dependendo do tamanho do cópulus e a inserção dos *concgrams* no banco de dados da ferramenta requer uma preparação inicial, em que temos que agrupar em uma célula todas as palavras que se relacionam a cada *concgram* de duas palavras. E dadas todas as palavras para as quais fizemos o levantamento, o volume de dados é grande; não é algo complicado de se fazer, até pode-se pensar em alguma automatização, mas exige algum tempo de preparação.

Voltando à questão de pesquisa que norteou essa segunda parte de nossa tese, podemos respondê-la que **sim**, é possível aproveitar as associações identificadas na Parte I em uma ferramenta computacional. Mesmo com dados controlados, consideramos que foi possível obter da ferramenta um desempenho satisfatório, que nos deu boas perspectivas para o prosseguimento do desenvolvimento. Optamos, neste momento, por observar o comportamento do sistema aos poucos, começando com um banco de dados menor, com as associações de duas palavras-nodo somente: *objetivos* e *amostras*. No entanto, daremos continuidade à inserção dos dados, monitorando o comportamento do sistema a cada nova inserção.

Mostramos com alguns exemplos como a ferramenta ASSIM pode ampliar a gama de resultados apresentados ao tradutor na ocasião de uma nova tradução, em comparação com o Trados. Isso corrobora as abordagens que utilizam critérios de segmentos menores de correspondência entre texto fonte e texto alvo, como o Similis, que apresentamos anteriormente. Ressaltamos novamente que a ASSIM, no entanto, não faz apenas essa comparação de segmentos entre o novo texto a ser traduzido e o que está armazenado. A ferramenta primeiro identifica quais palavras do texto a ser traduzido são associações frequentes, para depois buscar os trechos do *cópus* alinhado que contêm as mesmas associações. Isso, para nós, é uma maneira de evidenciar ao tradutor aprendiz que aquelas associações são frequentes e constituem o texto especializado da Pediatria, assim lhe dando acesso a um "conhecimento" que ele só poderia adquirir após a realização de várias traduções sobre um mesmo tema ou domínio.

Pretendemos dar prosseguimento ao desenvolvimento da ferramenta, inserindo todos os dados levantados e sempre pensando em formas de aprimorar tanto seus resultados como sua usabilidade. No momento, o sistema é um aplicativo que precisa ser instalado no computador do usuário, mas, futuramente, pensamos em disponibilizá-lo *online* para acesso público. Nossa ideia é, futuramente, poder fazer alguns testes com os alunos do curso de Tradução da UFRGS em colaboração com disciplinas de tradução para o inglês.

Até o momento de escritura desta tese fizemos o que estava ao nosso alcance, em termos de tempo, principalmente, mas vislumbramos um longo futuro para a ASSIM, principalmente porque acreditamos que ela pode estabelecer algumas bases, quem sabe, para aplicações mais robustas, com um processamento mais sofisticado. O que temos até o momento é um protótipo, mas acreditamos que foi possível mostrar o seu potencial de aplicação, e os benefícios e vantagens da abordagem aqui apresentada.

CAPÍTULO X

CONCLUSÕES

Nesta etapa do trabalho, é útil retomar nossa hipótese e as questões de pesquisa. O trabalho se dividiu em duas partes, guiadas por duas questões de pesquisa. Nossa hipótese era de que:

Existem, nos textos especializados, associações no nível sintagmático que:

- e) são combinações mais ou menos fixas e frequentes de palavras;
- f) não são necessariamente contínuas;
- g) estão associadas no nível sintagmático por uma atração semântica e/ou pragmática e/ou gramatical;
- h) não necessariamente contêm um termo em sua composição, sendo termo entendido aqui como uma expressão lexical associada a um conceito.

Essas associações se configuram como Unidades de Tradução Especializadas porque são recorrentes, fazem parte dos “modos de dizer” do gênero e do domínio em questão e apresentam riscos e vantagens ao tradutor.

As duas questões de pesquisa eram:

- 1) É possível identificar e descrever, em larga escala e com base em corpus, associações sintagmáticas com o perfil acima?
- 2) Uma vez identificadas tais associações, é possível aproveitá-las em uma ferramenta computacional que as armazene e as mostre para o tradutor na situação de uma tradução nova, trazendo também exemplos de como já foram traduzidas?

A hipótese foi confirmada por meio de estudos empíricos que comprovaram e demonstraram o perfil das associações que foram nossas unidades de análise. Mostramos, nas análises das linhas de concordância dos resultados do *ConcGram*, nas Etapas de 1 a 4 do Capítulo V, que essas associações estão presentes nos textos especializados e se conformam como delineamos na hipótese.

As respostas para as questões de pesquisa, em síntese, são as seguintes:

1) Sim, foi possível identificar e descrever tais associações em larga escala, principalmente utilizando um software lançado recentemente e denominado *ConcGram*. Percebemos que a identificação de associações descontínuas também pode ser feita usando as ferramentas do *WordSmith Tools*, no entanto é um processo que demanda mais esforço manual, que praticamente inviabiliza a aplicação em larga escala, sem contar que está mais sujeito às subjetividades do pesquisador.

2) Sim, após a identificação das associações, foi possível inseri-las em um banco de dados que é utilizado pela ferramenta *ASSIM*. Essa ferramenta faz uso de um *cópus* alinhado para mostrar os exemplos de como as associações foram já traduzidas.

Cumprida essa trajetória de pesquisa, esperamos ter chamado a atenção para as associações sintagmáticas com o perfil delineado. Afinal, elas, muitas vezes, pouco são notadas na prática – e nas teorizações – pois parecem ficar em um ponto-cego entre as associações especializadas – as que contêm termos – e as de feição mais genérica, bastante fixas e muito contíguas. O trabalho realizado permitiu a sua percepção não só pela não-continuidade, mas também pela evidenciação de dois blocos de forças que as conformam: a força semântica e a pragmática, além de uma força gramatical que, em nosso ponto de vista, estará também condicionada a uma das outras duas.

Pensando nas constatações que fizemos voltadas ao âmbito da tradução e das atividades do tradutor, podemos dizer que o critério de especialidade atribuído às palavras de caráter mais "especializado" é acompanhado também de outras considerações, como quanto ao que é risco e vantagem em uma situação de tradução. Entendemos que esses fatores são oriundos, entre outros, dessa reconstrução do texto na língua de chegada, em que relações sintagmáticas precisam ser reconstruídas. Sendo assim, saber traduzir um texto técnico depende também de saber reformular tais associações e, por esse motivo, há a necessidade de se ajustar o foco também para a construção das frases e seus elementos.

Do nosso ponto de vista, no momento em que uma sequência de caracteres como *com média de x meses*, que não apresenta nenhuma palavra que seja considerada termo técnico, se mostra frequente em um texto especializado de Medicina, ela deixa de ser apenas uma recorrência do texto especializado, mas passa a configurar-se como pertencente ao âmbito da Medicina, pela sua relação com o restante do texto palavras. Do ponto de vista do tradutor, a expressão adquire relevância porque faz parte deste texto e ele precisa saber como dizê-la, com quais palavras, em outra língua. Se buscarmos por essa mesma expressão em um *córpus* de artigos de Informática ou Química, como o fizemos, não encontramos nenhuma ocorrência da expressão. Assim, o que, a princípio, parece ser uma expressão do âmbito da ciência em geral já não se comprova em outras áreas técnico-científicas. Poderíamos dizer que a palavra *média* já avança no espectro que vai de palavras a termos e se estabelece numa posição que se aproxima mais do caráter de termo, dado o contexto em que se realiza, isto é, sua relação com as outras palavras em seu entorno.

Esses aspectos que levantamos são apenas para demonstrar como, para os tradutores, é relevante que se incluam, em obras de referência, informações sobre o uso dos termos e palavras, colocações, variantes, frases de contexto. Essa seria uma proposta de uma *Terminologia Tradutiva*, como sugere Teixeira (2008), como já citamos anteriormente. Ressaltamos que é preciso pensar em alterações no formato das obras que venham a conter esses tipos de informação, assim como em outros formatos além do dos dicionários, tanto pela natureza variável desses itens como pela imensa quantidade de informações que eles geram em torno de si. Vimos, neste estudo, como as mesmas associações de palavras podem estar associadas a outras tantas (como classificadores, numerais) e como elas podem ser flexíveis tanto quanto à posição na frase como no número de palavras entre um item e o que se associa a ele.

Demonstramos como as fraseologias do texto especializado podem não ser tão fixas e como os chamados “modos de dizer” não são nem um pouco restritos ou fixos. Essa constatação evidencia, portanto, como é preciso repensar o modo de fazer obras de referência e de disponibilizar a gama de informações que podemos detectar em *córpus* atualmente. A ferramenta ASSIM, portanto, destaca-se também como uma contribuição para o paradigma de desenvolvimento de novas maneiras de desenvolver produtos de auxílio e referência para tradutores.

Nesta pesquisa, mais uma vez, confirmamos como o trabalho com um *córpus* que envolve considerações qualitativas, e não apenas estatísticas, é difícil, e como é árdua a tarefa

de entender e sistematizar tantos dados e que nos são apresentados com tantas variáveis. Há que se ter um tanto de resignação para entender dados que escapam aos desejos iniciais. Justamente por querermos incluir o máximo possível dos resultados apresentados pelas análises dos textos com o *ConcGram*, tivemos que encontrar uma forma de descrever os resultados que fosse sistemática, mas que também fosse abrangente e não excludente. Também, é por esse motivo que não estava em nossos objetivos delimitar um perfil sintático para as associações, pois com certeza teríamos que deixar de fora muitos dos dados encontrados.

Quando iniciamos esta tese com o objetivo de desenvolver algo aplicado, prático, usando recursos computacionais, nos parecia que a ideia de ter ao final uma ferramenta que fizesse o que imaginávamos era algo muito distante, talvez, a nossos olhos leigos, impossível. Conversamos com pessoas da área de Informática que nos apoiaram e disseram que tendo os dados seria possível, sim, desenvolver um protótipo, o que nos deu fôlego para investir no levantamento dos dados. Percebemos que conseguimos ir até além do protótipo, pois na configuração atual, tudo está preparado para que mais dados possam ser inseridos no banco.

Esperamos poder ter demonstrado que a relação entre Tecnologia e Tradução e que a relação de ambas com o estudo da linguagem especializada via *córpus* e texto não é uma “febre” passageira; é algo fundamental. Portanto, devemos estar cada vez mais alertas quanto às possibilidades que a associação entre essas áreas oferece, tanto no que tange ao levantamento de dados que possam ser usados no auxílio a tradutores como no estudo da tradução e dos textos especializados.

Destacamos ainda o ineditismo de estudar *concgrams* o que, até onde pudemos constatar, só foi feito pela equipe que desenvolveu o software e não se tem notícias ainda de outros trabalhos publicados no Brasil que o utilizem. Esperamos, em breve, publicar um artigo com os resultados alcançados para dar mais visibilidade à noção de *concgrams*, que ressaltam as discontinuidades da associações e a variação posicional, e ao próprio software *ConcGram*.

Por fim, resta-nos afirmar que esperamos que esse percurso de pesquisa possa render novas inquietações, novos desenvolvimentos e que sirva de encorajamento para novas pesquisas. Com certeza, esta nossa aventura de tese, nos seus altos e baixos, propiciou muitas preocupações e angústias, mas também mostrou que muitas heterogeneidades da linguagem podem ser apreendidas e compreendidas.

Além disso, a investigação possibilitou encontros agradabilíssimos com pesquisas brilhantes sobre o fenômeno da coocorrência de determinadas palavras e estudos com *cópus*. Com seus autores, dividimos certezas e dúvidas e pudemos criar novas perguntas em espaços em que se julgava não haveria mais nada a perguntar. Assim, esperamos ter contribuído um pouco com a reflexão sobre o caráter especializado dos textos, usando para isso principalmente o *cópus* de Pediatria/Medicina, e sobre os tipos de abordagens, teóricas e práticas, que podem vir realmente a ser úteis tanto para os tradutores quanto para os estudiosos dos fenômenos que envolvem os usos da linguagem científica em diferentes gêneros textuais e comunicativos.

REFERÊNCIAS

- ACOSTA, O. **Um Estudo sobre a Geração e Manipulação de Grafos Representativos de Textos**. Projeto de graduação, UCPEL, 2006. Disponível em: <<http://pg.ucpel.tche.br/doku.php?id=pg:2006.1>>. Acesso em: 20 fev. 2010.
- ALVES, F. **Zwischen Schweigen und Sprechen: Wie bildet sich eine transkulturelle Brücke? Eine psycholinguistisch orientierte Untersuchung von Übersetzungsvorgängen zwischen portugiesischen und brasilianischen Übersetzern**. Hamburg: Dr. Kovac. 1995.
- ALVES, F. **O processo de tradução: delimitação de um objeto de estudo**. Cadernos de Tradução, Florianópolis, v. 10, p. 9-22, 2002.
- ALVES, F.; MAGALHÃES, C. e PAGANO, A. **Traduzir com autonomia: estratégias para o tradutor em formação**. São Paulo: Contexto, 2000
- ALVES, F.; PAGANO, A. S.; MAGALHÃES, C. M.. **Autonomy in translation: approaching translators education through awareness of discourse processing**. Cadernos de Tradução, Florianópolis, v. 10, p. 167-213, 2002.
- AVILA, C. e LOH, S. **SISRECAC: Sistema de recomendação de artigos científicos**. Disponível em <<http://atlas.ucpel.tche.br/~loh/pdf>>. Acesso em: 28 jun. 2008.
- AZEREDO, S. **Expressões anunciadoras de paráfrase em manuais acadêmicos de química**. Dissertação de mestrado, Universidade Federal do Rio Grande do Sul, 2007.
- BAKHTIN, M. M. **Estética da criação verbal**. 2 ed. São Paulo: Martins Fontes, 1997.
- BATALHA, M. C., PONTES JR, G. R. **A tradução como prática da alteridade**. Cadernos de tradução, Santa Catarina, v. XIII, p. 27-44, 2005.
- BELL, R.T. **Translation and Translating**. Londres, Longman, 1991.
- BENEDUZZI, R. **Colocações substantivo+adjetivo: propostas para sua identificação e tratamento lexicográfico em dicionários ativos português-espanhol**. Dissertação de mestrado em Letras, Universidade Federal do Rio Grande do Sul, 2008.

- BERBER SARDINHA, T. **Padrões Lexicais e Colocações do Português**. LAEL, PUC/SP, 1999. Disponível em: <http://www2.lael.pucsp.br/~tony/1999padroes_inpla.pdf>. Acesso em 15 de novembro de 2009.
- BERBER SARDINHA, T. **Linguística de Corpus**. Barueri, São Paulo: Manole, 2004.
- BERRY-ROGGHE, G.L.M. **The Computation of Collocations and Their Relevance in Lexical Studies**. in *The Computer and Literary Studies*. Eds. A.J. Aitken, R.W. Bailey and N. Hamilton-Smith. Edinburgh: Edinburgh University Press, p 103-112, 1973.
- BEVILACQUA, C. R. **Unidades fraseológicas especializadas: elementos para seu reconhecimento em corpus textuais**. In: XII Intercâmbio em Linguística Aplicada, São Paulo. Intercâmbio. São Paulo : Lael - USP, v. 12. p. 215-223, 2003.
- BEVILACQUA, C. R. **Unidades Fraseológicas Especializadas Eventivas: Descripción Reglas De Formación En El Ámbito De La Energía Solar**. Tese de Doutorado, Barcelona: Institut Universitari de Linguística Aplicada, 2004.
- BEVILACQUA, C. R. **Fraseologia: perspectiva da língua comum e da língua especializada**. Revista *Língua & literatura*. Universidade Integrada do Alto Uruguai e das Missões, Departamento de Linguística, Letras e Artes. Frederico Westphalen, v. 6 e 7, n. 10/11, p. 73-86, 2005.
- BIBER, D. **Variations across speech and writing**, Cambridge University Press, Cambridge, 1988.
- BMJ. **British Medical Journal**. Disponível em <www.bmj.com>. Acesso em: 2 jun. 2008.
- BOWKER, L. **Computer Aided Translation Technology: A Practical Introduction**, University of Ottawa Press, Ottawa, Canada, 2002.
- BROWN, R.D. **Example-Based Machine Translation in the Pangloss System**. In *Proceedings of the 16th International Conference on Computational Linguistics (COLING-96)*, p. 169-174. Copenhagen, Denmark, August 5-9, 1996.
- BURKE, R. **Hybrid recommender systems: Survey and experiments**. *User Modeling and User Adapted Interaction*, v.12, n. 6, p. 331-370, nov. 2002. Disponível em: <<http://josquin.cti.depaul.edu/~rburke/pubs/burke-umuai02.pdf>>. Acesso em: 28 dez. 2006.
- CABRÉ, M. T. **La terminología: representación y comunicación**. Barcelona: IULA, 1999.
- CABRÉ, T. 2004. **La terminología en la traducción especializada**. In Gonzalo García, C. & García Yebra, V. (eds.). *Manual de documentación y terminología para la traducción especializada*. Madrid: Arcos/Libros S. L. 89-125.
- CARL, M. e HANSEN, S. **Linking Translation Memories with Example-Based Machine Translation**, Machine Translation Summit VII, Singapore, pp. 617-624, 1999.

- CHENG, W., GREAVES, C. and Warren M. **From n-gram to skipgram to conogram.** *International Journal of Corpus Linguistics* 11/4: 411-433, 2006.
- CHENG, W., GREAVES, C., SINCLAIR, J. and WARREN, M. Uncovering the extent of the phraseological tendency: towards a systematic analysis of congrams. *Applied Linguistics*, 2009.
- COULTHARD, R. **Paediatric Corpora and Translation: The JPED parallel corpus and the Pediatrics comparable corpus.** Dissertação de mestrado, Universidade Federal de Santa Catarina, 2005.
- DANCETTE, J. **Parcours de la traduction, essai.** Presses universitaires de Lille, 1995.
- DANIELSSON, Pernilla. **Automatic extraction of meaningful units from corpora.** *International Journal of Corpus Linguistics* 8, 1. 2003
- DELISLE, J. **Translation: An Interpretive Approach,** The University of Ottawa Press, 1988.
- DORR, B.J.; JORDAN, P. W.; BENOIT, J.W. **A Survey of Current Paradigms in Machine Translation.** In M. Zelkowitz (ed), *Advances in Computers*, Vol 49, pp. 1-68. Academic Press, London, 2000.
- ECO, U. **Quase a mesma coisa.** Rio de Janeiro: Record, 2007.
- ESTOPÀ, R. **Extracción de terminología: elementos para la construcción de un SEACUSE (Sistema de Extracción Automática de Candidatos a Unidades de Significación Especializada).** Tese de doutorado. Barcelona: IULA, Universitat Pompeu Fabra, 2000.
- FELTRIM, V. D. **Suporte computacional à escrita científica em português.** Tese de Doutorado, 2002
- FILLMORE, C. J. **Innocence: A Second Idealization for Linguistics,** *Berkeley Linguistic Society* 5, 63-76, 1979.
- FINATTO, M. J. B.; POSSAMAI, V. **Ultrapassando termos e dicionários: a pesquisa terminológica como elemento de diálogos transdisciplinares.** In: X SIMPÓSIO IBEROAMERICANO DE TERMINOLOGIA, 10. 2006, Montevideo. 2006
- FIRTH, J. R. **Papers in Linguistics 1934–1951.** Oxford University Press. Oxford, 1957.
- FIRTH, J. R. **Selected Papers of J. R. Firth, 1952–1959.** F. R. Palmer (Ed.), Indiana, University Press. London, 1968.
- FRACASSI, V. **Memórias de tradução, como esses programas podem afetar o seu trabalho.** 2007. Disponível em:
<<http://www.abrates.com.br/abreartigo.asp?onde=Memoria%20de%20Traducao.abr>>. Acesso em 05 dez. 2007.

- FRAWLEY, W. **Translation. Literary, Linguistic & Philosophical Perspectives**, Newark, 1984.
- GARRETSON, G. **What your words know: the theory of lexical priming**. International journal of corpus linguistics, Vol. 12, N° 3, p. 445-452, 2007.
- GÉMAR, J.C. **La traduction juridique ou le double défi: droit et (ou) langue?** Proceedings. First International Conference on Specialized Translation, p. 22-28, 2000.
- GILE, D. **Modalidades y tipos de traducción**. Vasos comunicantes, 4, 19027, 1995.
- GLEDHILL, C. **Collocations in Science Writing**. Tübingen: Narr. 2000.
- GOUADEC, D. **Nature et traitement des entités phraséologiques**. Terminologie et phraséologie. Acteurs et aménageurs: Actes de la deuxième Université d'Automne en Terminologie. Paris: La Maison du Dictionnaire, p. 167-193, 1994.
- GREAVES, C. **ConcGram 1.0: a phraseological search engine**. Manual de software. 2009.
- GREAVES, C. e WARREN, M. **Beyond clusters: A new look at word associations**. IVACS 4, 4th International Conference: Applying Corpus Linguistics. University of Limerick, Ireland, 13–14, June, 2008.
- GUTHRIE, D. et al., **A Closer Look at Skip-gram Modelling** in Proceedings of Fifth international Conference on Language Resources and Evaluation (LREC), Genoa, Italy, 1222—1225, 2006.
- GUTT, E. **Translation and relevance: cognition and context**. Oxford: Blackwell, 1991.
- GUZMÁN, R. **Manual MT Post-editing: “if it's not broken, don't fix it!**. Translation Journal, Volume 11, No. 4, October, 2007. Disponível em <<http://accurapid.com/journal/42mt.htm>>. Acessado em 20 fev. 2010.
- HALLIDAY, M. A. K. **Categories of the theory of grammar**. In: Halliday: System and function in language, edited by G. R. Kress, 52-72, Oxford: OUP, 1961.
- HALLIDAY, M.A.K and R. Hasan. **Cohesion in English**. London/New York: Longman, 1976.
- HOFFMANN, Lothar. **Conceitos básicos da Lingüística das Linguagens Especializadas**. In: KRIEGER, Maria da Graça; ARAÚJO, Luzia (Orgs.). Cadernos de Tradução da UFRGS, n° 17, outubro/dezembro de 2004. Porto Alegre: Instituto de Letras UFRGS, p. 79-90, 2004. Tradução de Maria José Bocorny Finatto, com colaboração de Ulla Pede Muss
- HOEY, M. **Lexical priming: A new theory of words and language**. London/New York: Routledge, Volume Functions of Language, Issue 14,2, p.283-294, 2005.
- HOEY, M. **Corpus-driven approaches to grammar: the search for common ground**. Apresentação em powerpoint. Disponível em <<http://lexicalpriming.org/wp->

- content/uploads/2007/07/corpus-driven-approaches-to-grammar-hanover-2006.pdf>. Acessado em 25 jun. 2010. 2007.
- HUNSTON, S. e FRANCIS, G. **Pattern Grammar: A Corpus-driven Approach to the Lexical Grammar of English**. Amsterdam: John Benjamins, 2000.
- HURTADO ALBIR, A. **Traducción y Traductología. Introducción a la Traductología**. Madrid: Ediciones Cátedra, 2001.
- IJCME. **Uniform Requirements for Manuscripts Submitted to Biomedical Journals: Writing and Editing for Biomedical Publication**. Disponível em <<http://www.icmje.org>>, em 26 jun. 2008.
- KILIAN, C. K. **A retomada de unidades de significação especializada em textos em língua alemã e portuguesa sobre gestão de resíduos: uma contribuição para a tradução técnico-científica**. Tese de doutorado, Universidade Federal do Rio Grande do Sul. 2007.
- KIRALY, D.C. **Pathways to Translation**, The Kent State University Press. 1995.
- KÖNIGS, F.G. **Der Vorgang des Übersetzens: Theoretische Modelle und praktischer Vollzug. Zum Verhältnis von Theorie und Praxis in der Übersetzungswissenschaft**. Lebende Sprachen 31/1, p.5-12, 1986.
- KÖNIGS, F.G. **Was beim Übersetzen passiert. Theoretische Aspekte, empirische Befunde und praktische Konsequenzen**. Die Neueren Sprachen 86/2, p.162-185, 1987.
- KRIEGER, M. G.; FINATTO, M. J. **Introdução à Terminologia: teoria e prática**. São Paulo: Contexto, 2004.
- LEDERER, M. **La Traduction Aujourd'hui – Le Modèle Interprétatif**. Paris: Hachette, 1994.
- LEE-JANKE, H. **Teaching medical translation; an easy job?** Panacea, vol. VI, n. 20, p.81-84. Julho, 2005. Disponível em <www.medtrad.org/panacea.html>. Acessado em 15 jul. 2005.
- LEECH, G. **Semantics: The Study of Meaning**. Harmondsworth: Penguin, 1974.
- LI, Y. e WARREN, M. **“in ... of”: What are collocational frameworks and should we be teaching them?** 4th International Conference on Teaching English at Tertiary Level. Zhejiang, China 11–12, October, 2008.
- LOUW, B. **Irony in the text or insincerity in the writer: the diagnostic potential of semantic prosodies**. In M. Baker, G. Francis & E. Tognini-Bonelli. Eds. Text and technology - Essays in honor of John McH Sinclair. Philadelphia/Amsterdam: John Benjamins, 1993.
- MACIEL, A. M. B. **Para o reconhecimento da especificidade do termo jurídico**. Tese de doutorado. Universidade Federal do Rio Grande do Sul, 2001.

- MAIA, B., SILVA, R, BARREIRO, A. et al. **N-grams in search of theories**. In Barbara Lewandowska-Tomaszczyk (ed.) *Corpus Linguistics, Computer Tools, and Applications – State-of-the Art*. Peter Lang, pp. 71-84, 2008.
- MANNING, C. D., & SCHÜTZE, H. **The Foundations of statistical natural language processing**. Cambridge, Mass.: The MIT Press. Disponível em: <<http://nlp.stanford.edu/fsnlp/promo/colloc.pdf>>. Acessado em 24 fev. 2009. 1999
- MARTÍNEZ MELIS, N. e HURTADO ALBIR, A. **Assessment in translation studies: research needs**. *Meta*, 46(2): 272-287, 2001.
- NELSON, M. **Semantic associations in Business English: A corpus-based analysis**. Source: *International Journal of Corpus Linguistics*, Volume 11, Number 4, 2006 , pp. 513-514(2), 2006.
- PAGURA, R. J.. **A interpretação de conferências: interfaces com a tradução escrita e implicações para a formação de intérpretes e tradutores**. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*. 19, 209-236, 2003.
- PARTINGTON, A. 1998. *Patterns and Meanings*. Amsterdam: John Benjamins.
- PEKAR, V. e MITKOV, R. **New Generation Translation Memory: Content-Sensitive Matching**. In proceedings of the 40th Anniversary Congress of the Swiss Association of Translators, Terminologists and Interpreters, 29-30 September 2006, Bern (ASTTI, 2007), 2007.
- PLANAS, E. **Extending Translation Memories**. In EAMT Machine Translation Workshop, Ljubljana, Slovenia, May, 2000.
- PLANAS, E. 2005. **SIMILIS Second-Generation Translation Memory Software**. Disponível em <<http://www.aslib.co.uk/conferences/programme27.html>>. Acessado em 6 mai. 2009.
- POSSAMAI, Viviane. **Marcadores textuais do artigo científico em comparação português-inglês : um estudo sob a perspectiva da tradução**. Dissertação de mestrado, Universidade Federal do Rio Grande do Sul. Porto Alegre, 2004.
- PYM, A. **A Theory of Cross-Cultural Communication**. 2003. Disponível em <<http://www.tinet.cat/~apym/on-line/intercultures/cross-cultural.pdf>>. Acessado em 23 set. 2009.
- PYM, A. **Text and risk in translation**. 2004. Disponível em <www.tinet.org/~apym/on-line/risk_analysis.pdf>. Acessado em 15 jul. 2009.
- QUAH, C. K. **Translation and Technology**. Hampshire: Palgrava Macmillan. 2006.
- RESNICK, P. and VARIAN. H. **Recommender Systems (introduction to special section)**. *Communications of the ACM* 40(3): 56-58, 1997.

- RODRIGUES, Cássio. **A abordagem processual no estudo da tradução: uma meta-análise qualitativa**. Cadernos de Tradução, Florianópolis, v. 10, p. 167-213, 2002.
- SANTOS, D. "**Relação entre informática e linguística, ou, entre informáticos e linguistas**". Escola de Verão Belinda Maia (FLUP, Porto, Portugal, 29 de Junho - 3 de Julho 2009). Disponível em <http://www.linguateca.pt/superb/busca_publicacoes.pl?idi=1247750437>. Acessado em 20 fev. 2010.
- SARMENTO, S. **O uso dos verbos modais em manuais de aviação em inglês: um estudo baseado em corpus**. Tese de doutorado, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2009.
- SAUSSURE, Ferdinand de. **Curso de Linguística Geral**. São Paulo: Cultrix, 1993.
- SCHLEIERMACHER, F. **Ueber die verschiedenen Methoden des Uebersetzens/Sobre os diferentes métodos de tradução**. In: HEIDERMANN, W. (Org.) *Clássicos da teoria da tradução*. Florianópolis: UFSC, p. 26-87, 2001.
- SELESKOVITCH, D. **Interpreting for international conferences**. Tradução de L'interprete dans les conferences internationales: problèmes de langage et de communication. Washington: DC, Pen and Booth, 1978.
- SELESKOVITCH, D. e LEDERER, M. **A systematic approach to teaching of interpretation. Tradução de Pédagogie raisonnée de l'interprétation**. n/c: The Registry of Interpreters for the Deaf, 1995.
- SHAROFF, S, BABYCH, B e HARTLEY, A: **Using collocations from comparable corpora to find translation equivalents**. LREC-2006: Fifth International Conference on Language Resources and Evaluation. Proceedings, Genoa, Italy, 22-28 May 2006; pp.465-470.
- SINCLAIR, J. **The nature of the evidence**. In J. McH. Sinclair (ed.) *Looking Up: An Account of the COBUILD Project in Lexical Computing*. London: Collins. 150-159, 1987.
- SINCLAIR, J.M. **Corpus, Concordance, Collocation**. Oxford: Oxford University Press, 1991.
- SINCLAIR, J. **The search for units of meaning**, *Textus* 9/1: 75-106. (Reprinted in Sinclair 2004a: 24-48.), 1996.
- SINCLAIR, J. **The Lexical Item**. In *Contrastive Lexical Semantics*, Weigand, Edda (ed.), SWALES, John M. (1990) *Genre analysis. English in academic and research settings*. Cambridge: Cambridge University press, 1998.
- SINCLAIR, J. **Trust the text**. in Coulthard, M. (ed). London: Routledge, 2004.
- SINCLAIR, J. **Collocation reviewed** (manuscript), Tuscan Word Centre, Italy. 2007.

- SLOCUM, J. **A Survey of Machine Translation: Its History, Current Status, and Future Prospects.** Machine Translation Systems. Org. Jonathan Slocum. Cambridge, Cambridge University Press, pp.1-41, 1985.
- SPECIA, L & RINO, L. H. **Introdução aos Métodos e Paradigmas de Tradução Automática.** 2002. Disponível em <http://www2.dc.ufscar.br/~lucia/TechRep/NILCTR0204-SpeciaRino.pdf>>. Acessado em 7 de fev de 2008.
- STUBBS, M.. **Corpus evidence for norms of lexical collocation.** In G. Cook & B. Seidlhofer (Eds.), Principle and Practice in Applied Linguistics. Studies in Honour of H.G. Widdowson (pp. 245–256). Oxford: Oxford University Press, 1995.
- STUBBS, M. **The search for units of meaning: Sinclair on empirical semantics.** Applied Linguistics, 30, 1: 115-37, 2009.
- SWALES, J.M. **Genre analysis: English in academic and research settings.** Cambridge: Cambridge University Press, 1990.
- TAGNIN, S. **O jeito que a gente diz.** São Paulo: Disal, 2005.
- TAGNIN, S. O. Os corpora: **Instrumentos de auto-ajuda para o tradutor.** In Cadernos de Tradução, número especial sobre Corpora e Tradução, Florianópolis, SC: Universidade Federal de Santa Catarina, 191-219. 2002
- TEIXEIRA, E.D. **A lingüística de corpus a serviço do tradutor: proposta de um dicionário de culinária voltado para a produção textual.** Tese de doutorado. Faculdade de Filosofia, Letras e Ciências Humanas (FFLCH), São Paulo, 2008.
- TEUFFEL, Simone Teufel. **Argumentative Zoning: Information Extraction from Scientific Text.** Tese de doutorado, School of Cognitive Science, University of Edinburgh, UK, 1999.
- TOGNINI-BONELLI, E. **Corpus Linguistics at Work.** Amsterdam: Benjamins, 2001.
- VAN ROEY, J: **French-English Contrastive Lexicology: An Introduction.** Louvain-la - Neuve: Peeters. 1990.
- VENUTTI, L. **Translator's invisibility: a history of translation.** New York: Routledge, 1995.
- VIEIRA R., LIMA, V. **Lingüística Computacional: Princípios e Aplicações.** In: As Tecnologias da Informação e a questão social: anais. Carlos Eduardo Ferreira (Ed.) Fortaleza, SBC. (v.2). pgs 47-88, 2001.
- WIERZBICKA, A. **Primitivos semánticos y universales léxicos: teoría y algunos ejemplos,** in Pamies, A & Luque, J.d.D. (eds): Trabajos de lexicología y fraseología contrastivas. Granada: Método, 2000. pp.1-28, 2000.

- WILKS, Y. 2005. **REVEAL: the notion of anomalous texts in a very large corpus**. Tuscan Word Centre International Workshop: Dial a Corpus. Certosa di Pontignano, Tuscany, Italy, 31 June – 3 July, 2005.
- WILSS, W. **Kognition und Übersetzen: Zu Theorie und Praxis der menschlichen und der maschinellen Übersetzung**. Tübingen: Niemeyer, 1988.
- WHITE, Peter R. R. **Functional Grammar**. CELS MA Program. Manual for Distance Students. 2009.
- ZILIO, Leonardo. **Colocações especializadas e Komposita: um estudo contrastivo alemão-português na área de Cardiologia**. Dissertação de mestrado, Porto Alegre: UFRGS, 2009.

ANEXO I

Lista de frequências da seção Introdução

1	DE	3101	5,9668
2	#	2705	5,2048
3	A	1911	3,6771
4	E	1636	3,1479
5	O	1061	2,0415
6	EM	909	1,7491
7	DA	889	1,7106
8	DO	697	1,3411
9	QUE	666	1,2815
10	COM	604	1,1622
11	NA	479	0,9217
12	PARA	462	0,889
13	É	436	0,8389
14	NO	414	0,7966
15	OS	411	0,7908
16	UM	348	0,6696
17	SE	342	0,6581
18	COMO	337	0,6484
19	UMA	335	0,6446
20	AS	315	0,6061
21	DOS	306	0,5888
22	CRIANÇAS	294	0,5657
23	POR	292	0,5619
24	OU	251	0,483
25	DAS	238	0,4579
26	MAIS	221	0,4252
27	NÃO	218	0,4195
28	À	205	0,3945
29	SÃO	192	0,3694
30	ENTRE	171	0,329
31	AO	168	0,3233
32	PACIENTES	150	0,2886
33	SER	142	0,2732
34	ESTUDOS	141	0,2713
35	ESTUDO	140	0,2694
36	NOS	127	0,2444
37	SAÚDE	127	0,2444
38	FOI	124	0,2386
39	ANOS	122	0,2347
40	DOENÇA	122	0,2347
41	TRATAMENTO	117	0,2251
42	PODE	108	0,2078

43	SUA	106	0,204
44	IDADE	105	0,202
45	TEM	105	0,202
46	FATORES	103	0,1982
47	NAS	99	0,1905
48	PREVALÊNCIA	99	0,1905
49	VIDA	98	0,1886
50	RISCO	92	0,177
51	SOBRE	90	0,1732
52	MAIOR	87	0,1674
53	PELO	86	0,1655
54	DIAGNÓSTICO	85	0,1636
55	SENDO	84	0,1616
56	TÊM	84	0,1616
57	RECÉM	83	0,1597
58	DESENVOLVIMENTO	82	0,1578
59	DURANTE	81	0,1559
60	TAMBÉM	80	0,1539
61	USO	80	0,1539
62	AUMENTO	79	0,152
63	SIDO	79	0,152
64	PELA	74	0,1424
65	OBJETIVO	73	0,1405
66	CRIANÇA	72	0,1385
67	NASCIDOS	70	0,1347
68	AOS	68	0,1308
69	PODEM	68	0,1308
70	AINDA	67	0,1289
71	ALEITAMENTO	66	0,127
72	ATÉ	66	0,127
73	FORAM	66	0,127
74	MATERNO	66	0,127
75	BRASIL	65	0,1251
76	CRESCIMENTO	64	0,1231
77	QUANDO	64	0,1231
78	ADOLESCENTES	63	0,1212
79	ATRAVÉS	61	0,1174
80	RELAÇÃO	61	0,1174
81	INFECÇÃO	60	0,1154
82	AL	59	0,1135
83	ET	59	0,1135

Lista de frequências da seção Métodos

1	DE	5957
2	#	4568
3	E	2710
4	A	2459
5	O	1626
6	DO	1489
7	DA	1306
8	COM	1079
9	EM	1077
10	PARA	1076
11	FORAM	1056
12	FOI	1022
13	OS	1005
14	AS	745
15	NO	682
16	OU	677
17	SE	659
18	QUE	646
19	POR	562
20	UM	467
21	ESTUDO	425
22	DOS	423
23	NA	408
24	DAS	372
25	CRIANÇAS	351
26	NÃO	310
27	PELO	296
28	UMA	293
29	COMO	290
30	IDADE	289
31	PACIENTES	288
32	ENTRE	287
33	AO	284
34	À	266
35	TESTE	222
36	DADOS	209
37	VARIÁVEIS	208
38	PELA	207
39	PESO	197
40	ANÁLISE	190
41	PERÍODO	187
42	APÓS	186
43	ANOS	182
44	PESQUISA	179
45	ATRAVÉS	175
46	CADA	160
47	GRUPO	160

48	DURANTE	153
49	MESES	153
50	TODOS	139
51	REALIZADA	131
52	NOS	130
53	É	129
54	QUANDO	129
55	GRUPOS	127
56	AVALIAÇÃO	125
57	DOIS	122
58	NÍVEL	122
59	SÃO	122
60	DIAGNÓSTICO	118
61	RECÉM	118
62	CRIANÇA	116
63	HOSPITAL	115
64	TEMPO	113
65	USO	113
66	DIAS	109
67	ÉTICA	108
68	ACORDO	107
69	AMOSTRA	104
70	MAIS	104
71	NÚMERO	104
72	TRÊS	104
73	DIA	103
74	PRESENÇA	103
75	SENDO	102
76	MÉDIA	101
77	NASCIDOS	100
78	NAS	99
79	VALORES	99
80	ATÉ	98
81	CRITÉRIOS	96
82	DUAS	96
83	PROGRAMA	96

Lista de frequências da seção Resultados

1	#	11116
2	DE	4733
3	E	2733
4	A	2211
5	EM	1389
6	COM	1204
7	O	1166
8	DO	1025
9	OS	903
10	QUE	854
11	DA	841
12	TABELA	729
13	FOI	726
14	DOS	713
15	PACIENTES	708
16	NO	695
17	NÃO	640
18	ENTRE	608
19	NA	589
20	AS	588
21	PARA	580
22	SE	564
23	CRIANÇAS	542
24	P	525
25	DAS	504
26	FORAM	442
27	GRUPO	435
28	AO	413
29	UM	383
30	POR	368
31	IDADE	363
32	OU	353
33	ANOS	329
34	À	321
35	UMA	315
36	NOS	298
37	MESES	295
38	FIGURA	262
39	MAIS	250
40	GRUPOS	237
41	RECÉM	236
42	DOIS	234
43	MÉDIA	230
44	RELAÇÃO	230
45	PESO	225
46	MAIOR	211
47	NASCIDOS	208
48	APRESENTARAM	205
49	SEXO	190

50	DIFERENÇA	182
51	SENDO	179
52	CASOS	173
53	COMO	169
54	NAS	168
55	TRÊS	165
56	APÓS	159
57	HOUVE	159
58	ESTUDO	157
59	DIAS	153
60	SIGNIFICATIVA	145
61	QUANDO	142
62	AOS	141
63	VARIÁVEIS	141
64	VALORES	140
65	TEMPO	133
66	N	131
67	ESTATISTICAMENTE	128
68	QUANTO	127
69	PERÍODO	121
70	SEGUNDO	121
71	MEDIANA	120
72	NÚMERO	120
73	QUATRO	115
74	TOTAL	115
75	DISTRIBUIÇÃO	113
76	APENAS	109
77	USO	109
78	DIAGNÓSTICO	107
79	GRAU	105
80	CARACTERÍSTICAS	103
81	DIA	102
82	MÃES	102
83	ANÁLISE	96

Lista de frequências da seção Discussão

1	DE	7764
2	#	6930
3	A	4207
4	E	3283
5	QUE	2653
6	O	2437
7	EM	2241
8	DA	1862
9	COM	1736
10	DO	1665
11	OS	1266
12	PARA	1190
13	NO	1169
14	NA	1064
15	SE	1017
16	NÃO	987
17	DOS	904
18	AS	812
19	UMA	802
20	UM	795
21	POR	775
22	CRIANÇAS	772
23	ESTUDO	721
24	À	689
25	É	671
26	PACIENTES	668
27	AO	663
28	FOI	630
29	ENTRE	622
30	MAIS	604
31	COMO	590
32	DAS	563
33	SER	474
34	OU	463
35	MAIOR	426
36	NOS	418
37	PODE	325
38	SÃO	320
39	FORAM	316
40	ESTUDOS	295
41	IDADE	284
42	USO	278
43	AL	275
44	GRUPO	274
45	ANOS	273
46	AOS	273
47	ET	273
48	TAMBÉM	266
49	RELAÇÃO	245

50	RESULTADOS	245
51	QUANDO	239
52	DIAGNÓSTICO	238
53	RISCO	221
54	DOENÇA	220
55	OUTROS	220
56	NAS	218
57	PRESENTE	217
58	PREVALÊNCIA	205
59	RECÉM	204
60	PELO	198
61	PELA	194
62	NÍVEIS	187
63	APÓS	181
64	DADOS	177
65	MENOR	173
66	LITERATURA	172
67	PESO	172
68	VIDA	170
69	POPULAÇÃO	168
70	SENDO	168
71	NASCIDOS	166
72	CASOS	163
73	TRATAMENTO	163
74	NOSSO	162
75	FATORES	160
76	AValiação	156
77	SOBRE	154
78	AUTORES	152
79	SIDO	151
80	QUANTO	147
81	ESTE	146
82	JÁ	146
83	TEMPO	142

ANEXO II

Lista de palavras excluídas da análise com o ConcGram (*Stopwords*)

1. à
2. a
3. ao
4. aos
5. as
6. às
7. cá
8. com
9. da
10. das
11. de
12. do
13. dos
14. dum
15. duns
16. e
17. ele
18. eles
19. em
20. eu
21. lá
22. lhe
23. lhes
24. la
25. lo
26. los
27. las
28. me
29. meu
30. meus
31. mim
32. minha
33. minhas
34. muito
35. na
36. nas
37. no
38. nos
39. nós
40. nossa
41. nossas

42. nosso
43. nossos
44. nunca
45. o
46. ora
47. por
48. os
49. ou
50. para
51. que
52. seu
53. seu
54. seus
55. si
56. te
57. ti
58. tu
59. um
60. uma
61. umas
62. uns
63. você
64. vocês
65. vos
66. vós
67. vossa
68. vossas
69. vosso
70. vossos

ANEXO III

Listas de conegrams de 3 palavras na seção Introdução para a palavra *crianças*

CRIANÇAS

crianças	al	et	11	
crianças	anos	idade	10	
crianças	arterial	pressão	9	
crianças	não	uso	9	
crianças	anos	entre	8	
crianças	anos	menores	8	
crianças	abaixo	idade	5	
crianças	anos	dois	5	
crianças	aprovados	uso	5	
crianças	contra	uso	5	
crianças	estudo	et	5	
crianças	idade	meses	5	
crianças	idade	menores	5	
crianças	adolescentes	sobrepeso	4	
crianças	al	estudo	4	
crianças	alto	Xingu	4	
crianças	Além	disso	4	
crianças	ano	menores	4	
crianças	anos	mais	4	
crianças	anos	prevalência	4	
crianças	aprovados	não	4	
crianças	como	não	4	
crianças	corporal	composição	4	
crianças	desenvolvimento	países	4	
crianças	dieta	isenta	4	
crianças	dieta	leite	4	
crianças	dieta	vaca	4	
crianças	estado	nutricional	4	
crianças	idade	inferior	4	
crianças	isenta	leite	4	
crianças	isenta	vaca	4	
crianças	leite	vaca	4	
crianças	sido	tem	4	

Listas de conegrams de 3 palavras na seção Métodos para a palavra *estudo*

estudo foi realizado	38	
estudo foi pelo	35	
estudo aprovado foi	32	
estudo foi Vollss	32	
estudo aprovado pelo	28	
estudo realizado Vollss	28	
estudo Comitê pelo	27	
estudo Comitê foi	26	
estudo aprovado Comitê	23	
estudo transversal Vollss	19	
estudo crianças Foram	16	
estudo foi Pesquisa	16	
estudo foram pacientes	16	
estudo Este Vollss	15	
estudo aprovado Pesquisa	14	14
estudo Este foi	14	
estudo Comitê Pesquisa	13	
estudo foi transversal	13	
estudo se realizou	13	
estudo se Vollss	13	
estudo durante período	12	12
estudo foi foram	12	
estudo pelo Pesquisa	12	
estudo foram incluídos	11	
estudo foi pela	10	
estudo foram incluídas	10	
estudo comissão pela	9	
estudo coorte Vollss	9	
estudo crianças Vollss	9	
estudo etária faixa	9	
estudo foi não	9	
estudo foi população	9	
estudo foi pacientes	9	
estudo foram não	9	
estudo prospectivo Vollss	9	
estudo retrospectivo Vollss	9	
estudo se crianças	9	
estudo anos foram	8	
estudo aprovado pela	8	
estudo crianças incluídas	8	8
estudo entre foram	8	
estudo excluídos foram	8	
estudo foi prospectivo	8	
estudo foi protocolo	8	
estudo foram Vollss	8	

estudo realizado	transversal	8
estudo se	transversal	8
estudo tipo	Vollss	8
estudo amostra	foi	7
estudo aprovado	Comissão	7
estudo crianças	idade	7
estudo crianças	todas	7
estudo deste	foram	7
estudo foi	período	7
estudo foi	todos	7
estudo foram	inclusão	7
estudo foram	selecionados	7
estudo incluídos	pacientes	7
estudo Participaram	Vollss	7
estudo realizou	Vollss	7
estudo tipo	transversal	7
estudo caso	Vollss	6
estudo comissão	foi	6
estudo comissão	pesquisa	6
estudo controle	caso	6
estudo controle	Vollss	6
estudo crianças	foi	6
estudo crianças	População	6
estudo critérios	foram	6
estudo critérios	inclusão	6
estudo dados	foi	6
estudo deste	Vollss	6
estudo durante	foi	6
estudo Este	realizado	6
estudo foi	submetido	6
estudo foram	período	6
estudo observacional	Vollss	6
estudo pelo	protocolo	6
estudo piloto	realizado	6

Listas de conegrams de 3 palavras na seção Resultados para a palavra *pacientes*

pacientes	masculino	sexo	20	
pacientes	foi	grupo	15	
pacientes	se	Observou	15	
pacientes	distribuição	Tabela	14	
pacientes	dois	grupo	14	
pacientes	foi	não	13	
pacientes	dois	foi	12	
pacientes	Entre	Tabela	12	
pacientes	foi	três	12	
pacientes	anos	idade	11	
pacientes	B	grupo	11	
pacientes	encaminhados	foram	11	
pacientes	Entre	foi	11	
pacientes	etária	faixa	11	
pacientes	foi	maior	11	
pacientes	foram	incluídos	11	
pacientes	alogênico	TMO	10	
pacientes	distribuição	segundo	10	10
pacientes	foram	não	10	
pacientes	mediana	meses	10	
pacientes	segundo	Tabela	10	
pacientes	apenas	foi	9	
pacientes	casos	foi	9	
pacientes	Entre	não	9	
pacientes	Entre	p	9	
pacientes	foi	mais	9	
pacientes	foram	submetidos	9	
pacientes	foram	Vollss	9	
pacientes	grau	I	9	
pacientes	I	II	9	
pacientes	não	três	9	
pacientes	RM	TC	9	
pacientes	submetidos	TMO	9	
pacientes	anos	mediana	8	
pacientes	anos	meses	8	
pacientes	apresentaram	Dois	8	
pacientes	cada	grupo	8	
pacientes	Características	Tabela	8	
pacientes	dois	foram	8	
pacientes	EDA	grau	8	
pacientes	eram	masculino	8	
pacientes	eram	sexo	8	
pacientes	esofagite	grau	8	
pacientes	foi	normal	8	
pacientes	foi	observada	8	

pacientes	idade	meses	8	
pacientes	início	sintomas	8	
pacientes	se	CF	8	
pacientes	se	Tabela	8	
pacientes	alotômico	submetidos	7	
pacientes	anos	foi	7	
pacientes	apresentaram	grau	7	
pacientes	cicatrizrenal		7	
pacientes	clínicos	cirúrgicos	7	
pacientes	compareceram	grupo	7	
pacientes	diagnóstico	Figura	7	
pacientes	distribuição	etária	7	
pacientes	distribuição	faixa	7	
pacientes	Entre	foram	7	
pacientes	Entre	grupos	7	
pacientes	enurese	noturna	7	
pacientes	esofagite	I	7	
pacientes	etária	Tabela	7	
pacientes	faixa	Tabela	7	
pacientes	feminino	sexo	7	
pacientes	foi	média	7	
pacientes	foi	meses	7	
pacientes	foi	sendo	7	
pacientes	foram	sendo	7	
pacientes	foram	sexo	7	
pacientes	grupo	três	7	
pacientes	idade	sexo	7	
pacientes	mecânica	ventilação	7	
pacientes	muscular	relaxamento	7	
pacientes	média	meses	7	
pacientes	se	grau	7	
pacientes	acordo	Tabela	6	
pacientes	alérgicos	Tabela	6	
pacientes	anos	variou	6	
pacientes	apresentaram	durante	6	
pacientes	apresentaram	febre	6	
pacientes	autotômico	TMO	6	
pacientes	avaliação	foram	6	
pacientes	cada	número	6	
pacientes	Características	clínicas	6	
pacientes	CF	IV	6	
pacientes	cinco	foi	6	
pacientes	cinco	grupo	6	
pacientes	cinco	três	6	
pacientes	controles	Não	6	
pacientes	controles	Tabela	6	
pacientes	dez	foram	6	
pacientes	disfunção	miccional	6	
pacientes	distribuição	Figura	6	
pacientes	dois	entre	6	

pacientes	dois	grupos	6
pacientes	Down	síndrome	6
pacientes	EDA	I	6
pacientes	foi	foram	6
pacientes	foi	menor	6
pacientes	foi	possível	6
pacientes	foi	p	6
pacientes	foi	quatro	6
pacientes	foram	grupo	6
pacientes	foram	idade	6
pacientes	foram	masculino	6
pacientes	foram	observadas	6
pacientes	foram	período	6
pacientes	grau	normal	6
pacientes	grupo	tiveram	6
pacientes	I	normal	6
pacientes	IC	TMO	6
pacientes	idade	média	6
pacientes	livre	sobrevida	6
pacientes	mediana	variou	6
pacientes	ml	ng	6
pacientes	não	quatro	6
pacientes	se	I	6

Listas de congrams de 3 palavras na seção Discussão para a palavra *crianças*

crianças	al	et	37	
crianças	anos	idade	27	
crianças	anos	entre	16	
crianças	estudo	presente		16
crianças	entre	idade	14	
crianças	anos	até	13	
crianças	estudo	foi	13	
crianças	foi	não	13	
crianças	estudo	neste	12	
crianças	até	idade	11	
crianças	dieta	leite	11	
crianças	entre	mais	11	
crianças	foi	idade	11	
crianças	idade	meses	11	
crianças	leite	vaca	11	
crianças	anos	foi	10	
crianças	dois	idade	10	
crianças	foi	mais	10	
crianças	meses	vida	10	
crianças	não	se	10	
crianças	adolescentes	TVP		9
crianças	anos	dois	9	
crianças	anos	estudo	9	
crianças	dieta	isenta	9	
crianças	dieta	vaca	9	
crianças	entre	foi	9	
crianças	estudo	não	9	
crianças	foram	não	9	
crianças	isenta	leite	9	
crianças	isenta	vaca	9	
crianças	adolescentes	estudo	8	
crianças	adolescentes	foi	8	
crianças	anos	menores		8
crianças	anos	vida	8	
crianças	baixo	peso	8	
crianças	comparadas	quando	8	
crianças	cujas	mães	8	
crianças	DBP	sem	8	
crianças	dieta	exclusão		8
crianças	entre	foram	8	
crianças	entre	maior	8	
crianças	entre	não	8	
crianças	estado	nutricional		8
crianças	estudo	entre	8	
crianças	estudo	foram	8	
crianças	etária	faixa	8	

crianças	foi	grupo	8	
crianças	medicamentos	não	8	
crianças	ano	menores	7	
crianças	anos	acima	7	
crianças	anos	não	7	
crianças	arterial	pressão	7	
crianças	corticosteróide		mães	7
crianças	derivados	vaca	7	
crianças	dieta	ingestão	7	
crianças	entre	meses	7	
crianças	entre	pelo	7	
crianças	idade	mais	7	
crianças	mais	não	7	
crianças	mais	são	7	
crianças	pode	ser	7	
crianças	adolescentes	et	6	
crianças	adolescentes	obesidade	6	
crianças	al	anos	6	
crianças	al	idade	6	
crianças	anos	cinco	6	
crianças	anos	et	6	
crianças	anos	maiores	6	
crianças	apropriados	medicamentos	6	
crianças	apropriados	não	6	
crianças	assim	não	6	
crianças	baixa	estatura	6	
crianças	como	mães	6	
crianças	como	não	6	
crianças	como	tanto	6	
crianças	cujas	corticosteróide	6	
crianças	D	vitamina	6	
crianças	DBP	foram	6	
crianças	derivados	leite	6	
crianças	dessas	maior	6	
crianças	entre	mães	6	
crianças	estudo	idade	6	
crianças	estudo	meses	6	
crianças	estudo	realizado	6	
crianças	estudo	vida	6	
crianças	et	idade	6	
crianças	foram	mais	6	
crianças	jovens	mais	6	
crianças	maior	risco	6	
crianças	mais	menor	6	
crianças	mães	receberam	6	
crianças	obesidade	sobrepeso	6	
crianças	primeiro	vida	6	

Anexo IV

Listas de congrams para a palavra *objetivo*

objetivo	estudo
objetivo	foi
objetivo	deste
objetivo	avaliar
objetivo	como
objetivo	este
objetivo	presente
objetivo	trabalho
objetivo	tem
objetivo	teve
objetivo	comparar
objetivo	verificar
objetivo	analisar
objetivo	determinar
objetivo	descrever
objetivo	fatores
objetivo	identificar
objetivo	investigar
objetivo	principal
objetivo	prevalência
objetivo	realizado
objetivo	crianças
objetivo	níveis
objetivo	perfil
objetivo	características
objetivo	infecção
objetivo	impacto
objetivo	mecânica
objetivo	não
objetivo	Nesse
objetivo	pelo
objetivo	reduzir
objetivo	tratamento
objetivo	uso
objetivo	após
objetivo	artigo
objetivo	adultos
objetivo	análise
objetivo	alimentar
objetivo	Alguns
objetivo	Assim
objetivo	cujo
objetivo	células
objetivo	contexto
objetivo	colostro
objetivo	consiste
objetivo	doença

objetivo	dessas
objetivo	desta
objetivo	efeitos
objetivo	estudar
objetivo	esta
objetivo	exame
objetivo	energético
objetivo	grupo
objetivo	incidência
objetivo	manutenção
objetivo	meio
objetivo	massa
objetivo	meses
objetivo	maior
objetivo	Neste
objetivo	obesidade
objetivo	pesquisa
objetivo	portanto
objetivo	possíveis
objetivo	presença
objetivo	plasmáticos
objetivo	prevenção
objetivo	pacientes
objetivo	Recém
objetivo	região
objetivo	redução
objetivo	relacionados
objetivo	Realizamos
objetivo	Saúde
objetivo	se
objetivo	seis
objetivo	terapêuticas
objetivo	transversal
objetivo	valor

objetivo	deste	estudo
objetivo	estudo	foi
objetivo	avaliar	estudo
objetivo	deste	foi
objetivo	avaliar	deste
objetivo	estudo	presente
objetivo	avaliar	foi
objetivo	como	estudo
objetivo	como	este
objetivo	este	estudo
objetivo	este	trabalho
objetivo	estudo	teve
objetivo	como	teve
objetivo	estudo	prevalência
objetivo	estudo	verificar
objetivo	foi	realizado
objetivo	foi	trabalho
objetivo	avaliar	presente
objetivo	como	tem
objetivo	comparar	foi
objetivo	comparar	trabalho
objetivo	descrever	foi
objetivo	este	tem
objetivo	estudo	investigar
objetivo	foi	presente
objetivo	analisar	estudo
objetivo	como	determinar
objetivo	como	presente
objetivo	como	prevalência
objetivo	como	trabalho
objetivo	comparar	deste
objetivo	comparar	Este
objetivo	crianças	estudo
objetivo	determinar	estudo
objetivo	determinar	prevalência
objetivo	este	foi
objetivo	este	realizado
objetivo	este	teve
objetivo	estudo	fatores
objetivo	estudo	identificar
objetivo	estudo	níveis
objetivo	estudo	pelo
objetivo	estudo	tem
objetivo	foi	uso
objetivo	presente	teve
objetivo	realizado	trabalho
objetivo	adultos	estudo
objetivo	analisar	Assim
objetivo	analisar	deste
objetivo	análise	características
objetivo	análise	deste
objetivo	análise	estudo
objetivo	análise	foi
objetivo	após	cujo
objetivo	avaliar	crianças
objetivo	avaliar	energético
objetivo	avaliar	níveis
objetivo	avaliar	tem

objetivo	avaliar	uso
objetivo	características	estudo
objetivo	como	colostro
objetivo	como	energético
objetivo	como	grupo
objetivo	comparar	estudo
objetivo	comparar	energético
objetivo	comparar	realizado
objetivo	consiste	deste
objetivo	consiste	estudo
objetivo	contexto	estudo
objetivo	contexto	Nesse
objetivo	crianças	foi
objetivo	crianças	uso
objetivo	descrever	características
objetivo	descrever	desta
objetivo	descrever	estudo
objetivo	descrever	presente
objetivo	descrever	perfil
objetivo	dessas	estudo
objetivo	deste	características
objetivo	deste	investigar
objetivo	deste	incidência
objetivo	deste	níveis
objetivo	deste	pelo
objetivo	deste	principal
objetivo	deste	prevalência
objetivo	deste	pacientes
objetivo	deste	trabalho
objetivo	deste	uso
objetivo	deste	verificar
objetivo	determinar	colostro
objetivo	esta	pesquisa
objetivo	esta	tem
objetivo	este	colostro
objetivo	este	energético
objetivo	este	investigar
objetivo	este	massa
objetivo	este	níveis
objetivo	este	Realizamos
objetivo	este	verificar
objetivo	estudo	incidência
objetivo	estudo	meio
objetivo	estudo	massa
objetivo	estudo	Nesse
objetivo	estudo	não
objetivo	estudo	obesidade
objetivo	estudo	plasmáticos
objetivo	estudo	prevenção
objetivo	estudo	pacientes
objetivo	estudo	perfil
objetivo	estudo	realizado
objetivo	estudo	se
objetivo	estudo	transversal
objetivo	estudo	uso
objetivo	fatores	identificar
objetivo	fatores	possíveis
objetivo	fatores	relacionados

objetivo	fatores	tem
objetivo	foi	características
objetivo	foi	desta
objetivo	foi	incidência
objetivo	foi	investigar
objetivo	foi	prevalência
objetivo	foi	verificar
objetivo	grupo	este
objetivo	grupo	teve
objetivo	identificar	presente
objetivo	incidência	verificar
objetivo	infecção	desta
objetivo	investigar	teve
objetivo	meses	seis
objetivo	níveis	colostro
objetivo	níveis	plasmáticos
objetivo	níveis	tem
objetivo	pesquisa	tem
objetivo	presente	tem
objetivo	prevalência	teve
objetivo	recém	este
objetivo	redução	verificar
objetivo	tem	colostro
objetivo	tem	energético
objetivo	tem	verificar
objetivo	trabalho	colostro
objetivo	trabalho	energético
objetivo	trabalho	tem
objetivo	tratamento	desta

objetivo	deste	estudo	foi
objetivo	avaliar	deste	estudo
objetivo	avaliar	estudo	foi
objetivo	avaliar	deste	foi
objetivo	avaliar	estudo	presente
objetivo	como	estudo	teve
objetivo	como	este	trabalho
objetivo	como	este	tem
objetivo	como	estudo	prevalência
objetivo	determinar	estudo	prevalência
objetivo	este	foi	realizado
objetivo	estudo	foi	presente
objetivo	estudo	presente	teve
objetivo	foi	realizado	trabalho
objetivo	analisar	deste	estudo
objetivo	análise	deste	características
objetivo	análise	deste	estudo
objetivo	análise	deste	foi
objetivo	análise	estudo	características
objetivo	análise	estudo	foi
objetivo	análise	foi	características
objetivo	avaliar	crianças	estudo
objetivo	avaliar	crianças	uso
objetivo	avaliar	deste	níveis
objetivo	avaliar	deste	uso
objetivo	avaliar	estudo	níveis
objetivo	avaliar	estudo	uso
objetivo	avaliar	foi	presente
objetivo	avaliar	foi	uso
objetivo	avaliar	tem	energético
objetivo	como	determinar	colostro
objetivo	como	determinar	estudo
objetivo	como	determinar	prevalência
objetivo	como	este	colostro
objetivo	como	este	estudo
objetivo	como	este	energético
objetivo	como	este	teve
objetivo	como	estudo	presente
objetivo	como	grupo	este
objetivo	como	grupo	teve
objetivo	como	presente	teve
objetivo	como	prevalência	teve
objetivo	como	tem	colostro
objetivo	como	tem	energético
objetivo	como	trabalho	colostro
objetivo	como	trabalho	energético
objetivo	como	trabalho	tem
objetivo	comparar	deste	foi
objetivo	comparar	deste	trabalho
objetivo	comparar	Este	energético
objetivo	comparar	Este	foi
objetivo	comparar	Este	realizado
objetivo	comparar	Este	trabalho
objetivo	comparar	estudo	foi
objetivo	comparar	foi	realizado
objetivo	comparar	foi	trabalho
objetivo	comparar	trabalho	energético
objetivo	consiste	deste	estudo

objetivo	contexto	estudo	Nesse
objetivo	crianças	estudo	foi
objetivo	crianças	estudo	uso
objetivo	crianças	foi	uso
objetivo	descrever	estudo	características
objetivo	descrever	estudo	foi
objetivo	descrever	foi	características
objetivo	descrever	foi	desta
objetivo	descrever	foi	presente
objetivo	deste	estudo	características
objetivo	deste	estudo	investigar
objetivo	deste	estudo	incidência
objetivo	deste	estudo	níveis
objetivo	deste	estudo	pelo
objetivo	deste	estudo	prevalência
objetivo	deste	estudo	pacientes
objetivo	deste	estudo	uso
objetivo	deste	estudo	verificar
objetivo	deste	foi	características
objetivo	deste	foi	incidência
objetivo	deste	foi	uso
objetivo	deste	foi	verificar
objetivo	deste	incidência	verificar
objetivo	esta	pesquisa	tem
objetivo	este	estudo	investigar
objetivo	este	estudo	massa
objetivo	este	estudo	tem
objetivo	este	estudo	teve
objetivo	este	estudo	verificar
objetivo	este	foi	trabalho
objetivo	este	investigar	teve
objetivo	este	níveis	colostro
objetivo	este	níveis	tem
objetivo	este	realizado	trabalho
objetivo	este	tem	colostro
objetivo	este	tem	energético
objetivo	este	trabalho	colostro
objetivo	este	trabalho	energético
objetivo	este	trabalho	tem
objetivo	estudo	foi	características
objetivo	estudo	foi	incidência
objetivo	estudo	foi	prevalência
objetivo	estudo	foi	realizado
objetivo	estudo	foi	uso
objetivo	estudo	foi	verificar
objetivo	estudo	identificar	presente
objetivo	estudo	incidência	verificar
objetivo	estudo	investigar	teve
objetivo	estudo	níveis	plasmáticos
objetivo	estudo	prevalência	teve
objetivo	foi	incidência	verificar
objetivo	grupo	teve	este
objetivo	níveis	tem	colostro
objetivo	trabalho	tem	colostro
objetivo	trabalho	tem	energético

objetivo	avaliar	deste	estudo	foi
objetivo	análise	deste	estudo	características
objetivo	análise	deste	estudo	foi
objetivo	análise	deste	foi	características
objetivo	análise	estudo	foi	características
objetivo	avaliar	crianças	estudo	uso
objetivo	avaliar	deste	estudo	níveis
objetivo	avaliar	deste	estudo	uso
objetivo	avaliar	deste	foi	uso
objetivo	avaliar	estudo	foi	presente
objetivo	avaliar	estudo	foi	uso
objetivo	como	determinar	estudo	prevalência
objetivo	como	este	tem	coloostro
objetivo	como	este	tem	energético
objetivo	como	este	trabalho	coloostro
objetivo	como	este	trabalho	energético
objetivo	como	este	trabalho	tem
objetivo	como	estudo	presente	teve
objetivo	como	estudo	prevalência	teve
objetivo	como	grupo	teve	este
objetivo	como	trabalho	tem	coloostro
objetivo	como	trabalho	tem	energético
objetivo	comparar	Este	foi	realizado
objetivo	comparar	Este	trabalho	energético
objetivo	crianças	estudo	foi	uso
objetivo	descrever	estudo	foi	características
objetivo	deste	estudo	foi	características
objetivo	deste	estudo	foi	incidência
objetivo	deste	estudo	foi	uso
objetivo	deste	estudo	foi	verificar
objetivo	deste	estudo	incidência	verificar
objetivo	deste	foi	incidência	verificar
objetivo	este	estudo	investigar	teve
objetivo	este	foi	realizado	trabalho
objetivo	este	níveis	tem	coloostro
objetivo	este	trabalho	tem	coloostro
objetivo	este	trabalho	tem	energético
objetivo	estudo	foi	incidência	verificar