

UNIVERSIDADE FEDERAL DE MINAS GERAIS  
FACULDADE DE LETRAS

CARACTERIZAÇÃO DE REGISTROS ORIENTADA PARA A  
PRODUÇÃO TEXTUAL NO AMBIENTE MULTILÍNGUE:  
UM ESTUDO BASEADO EM *CORPORA* COMPARÁVEIS

Kelen Cristina Sant'Anna de Lima

BELO HORIZONTE

2013

KELEN CRISTINA SANT'ANNA DE LIMA

CARACTERIZAÇÃO DE REGISTROS ORIENTADA PARA A  
PRODUÇÃO TEXTUAL NO AMBIENTE MULTILÍNGUE:  
UM ESTUDO BASEADO EM *CORPORA* COMPARÁVEIS

Tese de doutorado apresentada ao Programa de Pós-Graduação em Estudos Linguísticos da Faculdade de Letras da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Doutora em Linguística Aplicada.

Área de concentração: Linguística Aplicada

Linha de Pesquisa: Estudos da Tradução

Orientadora: Profa. Dra. Adriana Silvina Pagano

Belo Horizonte  
Faculdade de Letras da UFMG

2013

Lima, Kelen Cristina Sant'Anna de

Caracterização de registros orientada para a produção textual no ambiente multilíngue: um estudo baseado em corpora comparáveis. [manuscrito] / Kelen Cristina Sant'Anna de Lima. - 2013.

251 f. il.

Orientadora: Adriana Silvina Pagano

Tese apresentada ao Programa de Pós-Graduação em Estudos Linguísticos da Faculdade de Letras da Universidade Federal de Minas Gerais, Área de concentração: Linguística Aplicada, Linha de Pesquisa: Estudos da Tradução para obtenção do título de Doutor em Linguística Aplicada.

1. Linguística de Corpus. 2. Corpus Comparável. 3. Linguística Sistêmico-Funcional. 4. Registro. 5. Produção Textual Multilíngue. I. Pagano, Adriana Silvina. II. Universidade Federal de Minas Gerais. Faculdade de Letras. III. Título.

CDD: 808:418



UNIVERSIDADE FEDERAL DE MINAS GERAIS  
PROGRAMA DE PÓS-GRADUAÇÃO EM ESTUDOS LINGÜÍSTICOS



## FOLHA DE APROVAÇÃO

### CARACTERIZAÇÃO DE REGISTROS ORIENTADA PARA A PRODUÇÃO TEXTUAL NO AMBIENTE MULTILÍNGUE: UM ESTUDO BASEADO EM CORPORA COMPARÁVEIS

#### KELEN CRISTINA SANT'ANNA DE LIMA

Tese submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em ESTUDOS LINGÜÍSTICOS, como requisito para obtenção do grau de Doutor em ESTUDOS LINGÜÍSTICOS, área de concentração LINGÜÍSTICA APLICADA, linha de pesquisa Linha H - Estudos da Tradução.

Aprovada em 27 de agosto de 2013, pela banca constituída pelos membros:

  
Prof(a). Adriana Silvina Pagano - Orientador  
UFMG

  
Prof(a). Pedro Henrique Lima Praxedes Filho  
UECE

  
Prof(a). Giacomo Patrocínio Figueredo  
UFOP

  
Prof(a). André Luiz Elias de Souza  
Concordia University

  
Prof(a). Rui Rothe-Neves  
UFMG

Belo Horizonte, 27 de agosto de 2013.

À minha família,  
amigos(as)

E ao grande desafio:  
o ato de pesquisar.

## AGRADECIMENTOS

É com grande satisfação e orgulho que encerro esta pesquisa que teve um caráter interinstitucional em função da sua inserção no âmbito de um projeto de cooperação entre o Laboratório Experimental de Tradução da Faculdade de Letras da UFMG (LETRA/FALE/UFMG) e o Núcleo de Ações e Pesquisa em Apoio Diagnóstico da Faculdade de Medicina da UFMG (NUPAD/FM/UFMG). Foi essencial para a finalização deste trabalho a participação direta ou indireta das pessoas que citarei abaixo, que sempre buscaram entender as dificuldades enfrentadas por uma doutoranda. A essas pessoas gostaria de deixar os meus sinceros agradecimentos.

Agradeço primeiramente a Deus, por sempre abençoar a minha vida e tornar possível mais um objetivo.

À Adriana Pagano agradeço imensamente a orientação, o carinho e a atenção durante tantos anos. Com a sua indicação dos caminhos a serem seguidos, conquistei o meu amadurecimento acadêmico e o entendimento de o que é ser uma pesquisadora.

Aos professores Giacomo Figueredo e Pedro Henrique Praxedes Filho, pela leitura detida e detalhada do texto de qualificação e por contribuírem para o aprimoramento desta pesquisa.

Ao André L. Souza, pela ajuda com o R, além das valiosas contribuições para a pesquisa, e à Aline Costa Braga, da UFOP, pelas pertinentes análises dos textos e companheirismo.

À minha filha, Isabella Victória e à minha mãe, Tânia, que apoiaram a minha decisão desde o início, mesmo sabendo que eu teria pouco tempo para me dedicar a elas durante essa fase. Aos demais membros da minha família, em especial, Alexandre, Cassius, Sandra, tia Jane e Kênia, pelo carinho e atenção incondicionais.

Ao Diretor Geral do NUPAD/FM/UFMG, Nelio, meu grande amigo e companheiro de longos anos, que me apoiou desde a graduação até hoje. Obrigada pela serenidade, compreensão, carinho e tanto trabalho conjunto; enfim, por entender minhas ausências e possíveis falhas. Sem a sua ajuda, o caminho trilhado, com certeza, teria sido infinitamente mais árduo.

Ao meu grande e melhor amigo Igor. Sem sua enorme ajuda, dedicação e apoio, a realização deste trabalho não seria possível. Obrigada pela presença em simplesmente todos os momentos. Ao Leonardo, pelo apoio nos trabalhos que desenvolvemos juntos na reta final de nossas pesquisas e pelas palavras carinhosas durante as longas madrugadas. Aos demais colegas do LETRA (Laboratório Experimental de Tradução) que me acolheram com enorme carinho, em especial: Tânia, Roberta, Malu, Renata, Silvana, Daniel, Leticia, Camila, Aline, Júlio, Roberto Carlos e Kícila.

A todos do NUPAD/FM/UFMG, especialmente a Thaís, pelas longas conversas de encorajamento, e ao Fernando, pela eficiência e perspicácia para lidar com os meus dados e entender exatamente cada detalhe da pesquisa.

À UNIPAC Contagem, principalmente Kátia Tofoli, Solange, Edneia e Marcelo Nery.

Aos novos colegas da Faculdade de Saúde e Ecologia Humana (FASEH), por entenderem que o primeiro semestre de 2013 seria ocupado pela finalização desta tese.

*É melhor tentar e falhar  
que preocupar-se e ver a vida passar;  
é melhor tentar, ainda que em vão,  
que sentar-se fazendo nada até o final.  
Eu prefiro na chuva caminhar  
a, em dias tristes, em casa me esconder.  
Prefiro ser feliz, embora louco,  
a em conformidade viver ...*

Martin Luther King

## RESUMO

Visando contribuir para a elaboração de um modelo que explique a produção e padrão de variabilidade de textos no ambiente multilíngue (FIGUEREDO 2011), esta tese se ampara na interface entre a Linguística de Corpus (MCENERY; XIAO, 2007; GRANGER, 2003) – mais especificamente as análises semiautomáticas de *corpora* comparáveis – e os Estudos da Tradução, com o aporte da Linguística Sistêmico-Funcional (HALLIDAY; MATTHIESSEN, 2004; EGGINS, 2004; MATTHIESSEN; TERUYA; WU, 2008; FIGUEREDO, 2011). Apresenta resultados de um estudo de padrões de uso linguístico (SINCLAIR, 1991; BERBER SARDINHA, 2004) em textos sobre triagem neonatal para anemia falciforme compilados em *corpora* comparáveis classificados de acordo com três tipos de textos específicos, considerando-se os rótulos a eles atribuídos pela comunidade de usuários da língua, a saber: (i) artigos acadêmicos (interação especialista – especialista); (ii) manuais técnicos (interação especialista – técnico da área da saúde); e (iii) cartilhas e *folders* de divulgação (interação especialista – leigo). Para a seleção da amostra de textos, foi adotada a metodologia descrita em Biber (1990), adaptada por Neumann (2005). Inicialmente, os textos foram anotados e manipulados através da utilização do ambiente de programação R, com o qual foram extraídos padrões de co-ocorrência entre itens lexicais e itens gramaticais específicos. Após a tabulação desses itens, foram identificados os padrões de uso linguístico que geraram indícios para a classificação de cada um dos tipos de texto de acordo com o seu processo sociosemiótico. Na sequência, os *corpora* foram anotados em classes de palavras por meio do etiquetador morfossintático Treetagger®. Os resultados obtidos com a aplicação do teste de associação qui-quadrado, do teste exato de Fisher e do teste Z possibilitaram identificar classes de palavras que se diferenciam significativamente e podem ser mais bem investigados para a caracterização do registro (MATTHIESSEN, 1993) ao qual estão vinculados. Excertos de mil palavras de cada um dos tipos de texto em inglês e em português foram classificados de acordo com a tipologia da língua no contexto de cultura (cf. MATTHIESSEN; TERUYA; WU, 2008). Os excertos foram anotados por meio do programa UAM CorpusTool®, com base nas metafunções ideacional, interpessoal e textual (HALLIDAY; MATTHIESSEN, 2004) e analisados de forma semiautomática. Os dados extraídos possibilitaram investigar frequências relativas dos sistemas de TRANSITIVIDADE, MODO, e TEMA e MENSAGEM, com o objetivo de propor uma descrição sistêmico-funcional dos tipos de textos (FIGUEREDO, 2011). Os resultados apontam diferenças, em função do registro, em relação à variação lexical, densidade lexical, frequência de ocorrência de itens lexicais e itens gramaticais, além do mapeamento da forma como esses itens estão distribuídos em classes de palavras. À luz da Linguística Sistêmico-Funcional, as diferenças e semelhanças entre os textos responderam ao impacto das variáveis do contexto (*i.e.*, campo, sintonia e modo) que se realizam no nível da léxico-gramática (EGGINS, 2004). Foi possível localizar os rótulos na tipologia dos textos com base nos parâmetros do contexto (MATTHIESSEN; TERUYA; WU, 2008), sendo classificados enquanto pertencentes aos processos sociosemióticos EXPLORAR (artigo acadêmico) e HABILITAR (manual técnico e cartilha e *folder* e divulgação), o que permite compreender a produção textual dentro do ambiente multilíngue. A descrição do perfil metafuncional dos textos revelou que, ideacionalmente, os processos mais significativos para a construção do mundo das experiências em todos os tipos de textos em português e em inglês foram os processos relacionais e materiais. Interpessoalmente, concernente à relação entre autor/leitor, o Modo que se destacou para todos os tipos de textos em português e em inglês foi o Indicativo Declarativo (Modo esse que apresenta como função semântica a oferta de informação), sendo encontrado o Modo Imperativo, cuja função semântica é a demanda de bens e serviços, apenas na cartilha e *folder* de divulgação em inglês. Textualmente, todos os tipos de texto em português e em inglês estão organizados, em função da semântica, a partir de mensagens iniciais, mensagens de continuidade e descontinuidade (mudança e desvio) e, em função da gramática, em tipos de Temas que se diferem para os textos em português e em inglês. As mensagens iniciais são responsáveis por separar as informações dos textos; as de continuidade acrescentam informações às mensagens iniciais; por fim, as de descontinuidade:mudança direcionam o fluxo de informações nos textos, considerando-se algum evento em particular e as mensagens de descontinuidade:desvio focalizam o fluxo de informações dos textos, levando-se em consideração algum participante em particular. No que toca aos tipos de Tema, os textuais e o *default* foram os mais frequentes em todos os tipos de texto em português, ao passo que os Tema simples e múltiplo (textual e tópico) foram os mais frequentes em todos os tipos de texto em inglês; o tema ângulo-fonte foi encontrado apenas no artigo acadêmico em português; e o Tema múltiplo (interpessoal e tópico) ocorreu apenas na cartilha e *folder* de divulgação em inglês. Com base no perfil metafuncional prototípico identificado, produziu-se um *template* que permitiu a elaboração de parte de um texto do tipo HABILITAR a partir das escolhas metafuncionais com base na gramática sistêmico-funcional hallidayana. A pesquisa é parte de um projeto conjunto desenvolvido entre o Laboratório Experimental de Tradução da Faculdade de Letras da UFMG (LETRA) e o Núcleo de Ações e Pesquisa em Apoio Diagnóstico da Faculdade de Medicina da UFMG (NUPAD/FM/UFMG).

Palavras-chave: Linguística de Corpus, Corpus Comparável, Linguística Sistêmico-Funcional, Registro, Produção Textual Multilíngue.

## ABSTRACT

This thesis aims to contribute to the development of a model to account for text production and variability within a multilingual environment (FIGUEREDO, 2011). It builds on Systemic-Functional Linguistics (HALLIDAY; MATTHIESSEN, 2004; EGGINS, 2004; MATTHIESSEN; TERUYA; WU, 2008; FIGUEREDO, 2011) as a framework that supports an interface between Translation Studies and Corpus Linguistics (MCENERY; XIAO, 2007; GRANGER, 2008) oriented towards the semiautomatic analysis of comparable corpora. It reports a study of patterns of language use (SINCLAIR, 1991; BERBER SARDINHA, 2004) in a comparable corpus compiled with texts on newborn screening for sickle cell disease ascribed to three specific text types and thus labelled by language users: 1) research articles (i.e., specialist-specialist interaction), 2) technical guides (specialist-technician interaction), and 3) pamphlets and patient information leaflets (specialist-layperson interaction). Text sampling was carried out following Biber (1990) as adapted by Neumann (2005). The texts were automatically and manually annotated and queried using the software R to obtain co-occurrence patterns of specific lexical and grammatical items. After classifying and computing these items, the analysis targeted patterns of use that account for how each text type could be classified according to their socio-semiotic process. The corpora were subsequently POS-tagged using Treetagger. Chi-square tests, Fisher's exact tests, and Z tests were carried out to identify patterns of word classes that could be taken to differentiate subcorpora and could support further analyses aimed at characterizing the registers with which the texts in each subcorpus were associated. Excerpts of 1,000 words were selected to represent each text type in English and Portuguese (BIBER, 1990) and classified according to language typology in the context of culture (cf. MATTHIESSEN; TERUYA; WU, 2008). These texts were pasted to UAM CorpusTool® for annotation and semi-automatic analyses of choices within the ideational, interpersonal and textual metafunctions (HALLIDAY; MATTHIESSEN, 2004). Frequencies of lexical and grammatical items in each text were computed with a view to proposing a systemic-functional (SFL) description (FIGUEREDO, 2011) of the TRANSITIVITY, MOOD, THEME and MESSAGE systems. The results pointed to registerial differences for lexical variation, lexical density, occurrence frequency of lexical and grammatical items, and provided a word class-based mapping of how these items are distributed in the texts. In the light of Systemic-Functional Linguistics (SFL), between-text differences and similarities were underscored building on the impact of context variables (i.e., field, tenor, and mode) on the lexico-grammar (EGGINS, 2004). Context parameters were used to locate the labels in the text typology (MATTHIESSEN; TERUYA; WU, 2008) and classify them as pertaining to the socio-semiotic processes EXPLORING (research article) and ENABLING (technical guide and pamphlets and patient information leaflets). This classification shed light on text production within the multilingual environment. The SLF-based description of the metafunctional profile of the texts showed that, ideationally, material and relational processes were the main processes used to construe the real world in all text types in both languages. Interpersonally, i.e. regarding author-reader interaction, the declarative mood, with the semantic function of information supply, was predominant in all text types both in Portuguese and in English, and the imperative mood, with the semantic function of demands of goods and services, was found only in the pamphlets and patient information leaflets. Textually, all text types are organized, at the semantic level, on the basis of initial messages, and continuity and discontinuity messages (change and shift) that are similar in both English and Portuguese, and, at the grammatical level, on the basis of theme types that are different for English and Portuguese. Initial messages are those that sort out text information; continuity messages add information to the initial messages; and discontinuity:change messages guide text information flow on the basis of a participant in particular. Textual and default themes were the most frequent type of Theme in all text types in Portuguese, whereas simple and multiple (textual and topical) themes were the most frequent in all text types in English; the angle-source Theme was found only in the Portuguese research articles; and the multiple (interpersonal and topical) Theme occurred only in the English pamphlets and patient information leaflets. Building on the prototypical metafunctional profile identified based on the metafunctional profile, a *template* was developed to serve as a basis for the production of part of an ENABLING text type that is based on SFL-informed metafunctional choices. The results reported are the first within a joint project developed by the Laboratory for Experimentation in Translation (LETRA, Faculty of Arts, Federal University of Minas Gerais) and the Center for Newborn Screening and Genetics Diagnosis (NUPAD, School of Medicine, Federal University of Minas Gerais).

Keywords: Corpus Linguistics, Comparable Corpus, Systemic-Functional Linguistics, Register, Multilingual Text Production.

## LISTA DE FIGURAS

FIGURA 1. Contextualização no Campo Disciplinar.....	21
FIGURA 2. Diagrama ilustrando os tipos de <i>corpus</i> .....	30
FIGURA 3. Estratificação e realização.....	33
FIGURA 4. Perspectiva trinocular da pesquisa .....	34
FIGURA 5. Realização entre os estratos semântico, fonológico e léxico-gramatical.....	35
FIGURA 6. Representação gráfica do componente semântico relacionado à léxico-gramática e realizado pelo Sistema de TRANSITIVIDADE.....	37
FIGURA 7. Sistemas de TRANSITIVIDADE das línguas portuguesa (A) e inglesa (B) .....	38
FIGURA 8. Sistema Geral do MODO e Funções discursivas .....	40
FIGURA 9. Sistema de TEMA em português brasileiro .....	43
FIGURA 10. Sistema de TEMA em inglês .....	45
FIGURA 11. Sistema de MENSAGEM .....	47
FIGURA 12. Perfil metafuncional do texto de popularização da ciência.....	52
FIGURA 13. Ambiente multilíngue no contínuo da instanciação do sistema linguístico.....	55
FIGURA 14. Variáveis do contexto de situação e seus respectivos parâmetros.....	59
FIGURA 15. A realização da língua através das três variáveis do contexto de situação.....	60
FIGURA 16. Tipologia de textos.....	64
FIGURA 17. Exemplo de cabeçalho de um dos textos comparáveis para o tipo de texto manual técnico.....	70
FIGURA 18. Exemplo do <i>corpus</i> compilado .....	72
FIGURA 19. Leitura do <i>corpus</i> a partir do ambiente de programação R .....	72
FIGURA 20. Tela inicial do ambiente de programação R, mostrando o comando para importar documentos em formato .txt.....	73
FIGURA 21. Criação da lista de palavras.....	73
FIGURA 22. Lista de <i>types</i> e frequência de <i>tokens</i> .....	74
FIGURA 23. Relação <i>type/token</i> simples e padronizada.....	75
FIGURA 24. <i>Script</i> para localizar as concordâncias por meio dos nódulos ou itens lexicais .....	77
FIGURA 25. <i>Script</i> para criar lista de frequência das palavras .....	77
FIGURA 26. Lista de palavras mais frequentes em ordem decrescente .....	78
FIGURA 27. Localização do item lexical no <i>corpus</i> .....	79
FIGURA 28. <i>Script</i> para aplicação do teste $\chi^2$ .....	80

FIGURA 29. Resultado da aplicação do teste $\chi^2$ .....	80
FIGURA 30. <i>Corpora</i> etiquetados por meio do etiquetador automático Treetagger .....	84
FIGURA 31. Exemplo do Esquema do Sistema Gramatical de TRANSITIVIDADE no UAM CorpusTool .....	88
FIGURA 32. Exemplo do Esquema do Sistema Gramatical de MODO no UAM CorpusTool .....	88
FIGURA 33. Exemplo do Esquema do Sistema Gramatical de TEMA em inglês no UAM CorpusTool .....	89
FIGURA 34. Exemplo do Esquema do Sistema Gramatical de TEMA em português no UAM CorpusTool .....	90
FIGURA 35. Exemplo do Esquema do Sistema Gramatical de TEMA em português no UAM CorpusTool .....	90
FIGURA 36. Exemplo de tela de projeto do UAM CorpusTool .....	91
FIGURA 37. Exemplo de análise realizada no UAM CorpusTool por meio de esquema previamente criado .....	91
FIGURA 38. Exemplo da extração dos dados estatísticos dos tipos de textos de forma comparada UAM CorpusTool .....	92
FIGURA 39. Disposição topológica das categorias.....	93
FIGURA 40. Gráfico tridimensional e dendrograma a partir dos dados obtidos pelo programa UAM CorpusTool .....	95
FIGURA 41. Exemplo de nuvem de palavras .....	98
FIGURA 42. Tipologia de textos.....	103
FIGURA 43. Síntese das frequências relativas do sistema de TRANSITIVIDADE nos textos do tipo EXPLORAR e HABILITAR .....	105
FIGURA 44. Síntese das frequências relativas do sistema de MODO nos textos do tipo EXPLORAR e HABILITAR.....	105
FIGURA 45. Síntese das frequências relativas do sistema de MENSAGEM nos textos do tipo EXPLORAR e HABILITAR.....	106
FIGURA 46. Síntese das frequências relativas do sistema de TEMA em português nos textos do tipo EXPLORAR e HABILITAR .....	107
FIGURA 47. Síntese das frequências relativas do sistema de TEMA em inglês nos textos do tipo EXPLORAR e HABILITAR.....	108
FIGURA 48. <i>Template</i> (modelo) desenvolvido para a elaboração de cartilha em português .....	110
FIGURA 49. Nuvem de palavras elaborada com base no <i>corpus</i> artigo_português.....	115
FIGURA 50. Nuvem de palavras elaborada com base no <i>corpus</i> artigo_inglês.....	115
FIGURA 51. Nuvem de palavras elaborada com base no <i>corpus</i> manual_português .....	116
FIGURA 52. Nuvem de palavras elaborada com base no <i>corpus</i> manual_inglês .....	116
FIGURA 53. Nuvem de palavras elaborada com base no <i>corpus</i> cartilha_português .....	116
FIGURA 54. Nuvem de palavras elaborada com base no <i>corpus</i> cartilha_inglês .....	116

FIGURA 55. Condensado dos gráficos sobre classe de palavras por língua e por rótulos (artigos acadêmicos, manuais técnicos e cartilhas e <i>folders</i> de divulgação).....	126
FIGURA 56. Estratificação da língua e os sistemas de TRANSITIVIDADE, MODO e TEMA relacionados às variáveis do contexto ( <i>i.e.</i> , campo, sintonia e modo) e às metafunções (ideacional, interpessoal e textual) .....	153
FIGURA 57. Representação gráfica tridimensional dos perfis metafuncionais dos tipos de textos EXPLORAR e HABILITAR .....	191
FIGURA 58. Dendrograma dos artigos científicos em português e em inglês .....	193
FIGURA 59. Dendrograma dos manuais técnicos em inglês e em português .....	194
FIGURA 60. Dendrograma das cartilhas e <i>folders</i> de divulgação em inglês e em português .....	195

## LISTA DE QUADROS

QUADRO 1.	Tipos, definição e exemplos dos processos que compõem o sistema de TRANSITIVIDADE....	39
QUADRO 2.	Exemplos de MODO dos tipos de texto.....	42
QUADRO 3.	Definição e exemplo de TEMA em português brasileiro dos tipos de texto .....	44
QUADRO 4.	Definição e exemplo de TEMA em inglês dos tipos de texto .....	46
QUADRO 5.	Compilação dos <i>corpora</i> comparáveis bilíngue por tipo de texto conforme processos sociossemióticos.....	70
QUADRO 6.	Etiquetas ( <i>tagset</i> ) para anotação dos textos em inglês e em português identificadas pelo programa Treetagger .....	82
QUADRO 7.	Agrupamento das etiquetas dos tagsets do inglês e do português .....	85
QUADRO 8.	Exemplo de tabela gerada a partir dos dados extraídos pelo UAM CorpusTool.....	93
QUADRO 9.	Classificação dos textos conforme a convergência do processo sociossemiótico EXPLORAR com as variáveis campo, sintonia e modo .....	139
QUADRO 10.	Classificação dos textos conforme a convergência do processo sociossemiótico HABILITAR com as variáveis campo, sintonia e modo .....	140
QUADRO 11.	Classificação dos textos conforme as variáveis campo, sintonia e modo na convergência do processo sociossemiótico HABILITAR.....	141
QUADRO 12.	Distribuição dos sistemas de TEMA e MENSAGEM no artigo acadêmico em português .....	161
QUADRO 13.	Distribuição dos sistemas de TEMA e MENSAGEM no artigo acadêmico em inglês .....	165
QUADRO 14.	Distribuição dos sistemas de TEMA e MENSAGEM no manual técnico em português .....	173
QUADRO 15.	Distribuição dos sistemas de TEMA e MENSAGEM no manual técnico em inglês .....	177
QUADRO 16.	Distribuição dos sistemas de TEMA e MENSAGEM na cartilha e <i>folder</i> de divulgação em português.....	185
QUADRO 17.	Distribuição dos sistemas de TEMA e MENSAGEM na cartilha e <i>folder</i> de divulgação em inglês .....	189

## LISTA DE TABELAS

TABELA 1.	Dados gerais dos <i>corpora</i> comparáveis.....	114
TABELA 2.	Frequência das classes de palavras no artigo acadêmico em inglês e em português .....	118
TABELA 3.	Frequência absoluta das classes de palavras no manual técnico em inglês e em português ....	120
TABELA 4.	Frequência absoluta das classes de palavras na cartilha e <i>folder</i> de divulgação em inglês e em português .....	121
TABELA 5.	Síntese dos dados empíricos dos <i>corpora</i> .....	122
TABELA 6.	Distribuições do teste Z para frequência das classes de palavras presentes nos artigos acadêmicos .....	127
TABELA 7.	Distribuições do teste Z para frequência das classes de palavras presentes nos manuais técnicos .....	128
TABELA 8.	Distribuições do teste Z para frequência das classes de palavras presentes nas cartilhas e <i>folders</i> de divulgação .....	129
TABELA 9.	Frequência de ocorrência dos nódulos doença falciforme e anemia falciforme para cada um dos <i>corpora</i> em português .....	130
TABELA 10.	Frequência de ocorrência dos nódulos “ <i>sickle cell disease</i> ” e “ <i>sickle cell anemia/anaemia</i> ” para cada um dos <i>corpora</i> em Inglês .....	132
TABELA 11.	Frequência de ocorrência dos nódulos “ <i>dor</i> ” (português) e “ <i>pain</i> ” (inglês) por tipos de texto	135
TABELA 12.	Frequência de ocorrência dos nódulos “ <i>pacientes</i> ” e “ <i>patients</i> ” em cada um dos tipos de texto .....	136
TABELA 13.	Frequência de ocorrência dos nódulos “ <i>você</i> ” e “ <i>you</i> ” em cada um dos tipos de texto .....	137
TABELA 14.	Frequência de ocorrência de substantivos em português por tipo de texto.....	143
TABELA 15.	Frequência de ocorrência de substantivos em inglês por tipo de texto.....	145
TABELA 16.	Frequência de ocorrência de pronomes pessoais em português .....	147
TABELA 17.	Frequência de ocorrência de pronomes pessoais em inglês.....	148
TABELA 18.	Frequência de ocorrência de conjunções em português .....	150
TABELA 19.	Frequência de ocorrência de conjunções em inglês.....	151
TABELA 20.	Frequência relativa das funções do sistema de TRANSITIVIDADE nos artigos acadêmicos em português e inglês.....	155

TABELA 21.	Frequência relativa das funções do sistema de MODO e Funções Discursivas nos artigos acadêmicos em português e inglês .....	158
TABELA 22.	Frequência relativa das funções do sistema de MENSAGEM nos artigos acadêmicos em português .....	160
TABELA 23.	Frequência relativa das funções do sistema de TEMA nos artigos acadêmicos em português	163
TABELA 24.	Frequência relativa das funções do sistema de MENSAGEM no artigo acadêmico em inglês .....	164
TABELA 25.	Frequência relativa das funções do sistema de TEMA no artigo acadêmico em inglês .....	166
TABELA 26.	Frequência relativa das funções do sistema de TRANSITIVIDADE nos manuais em português e inglês .....	167
TABELA 27.	Frequência relativa das funções do sistema de MODO e Funções Discursivas nos manuais técnicos em português e inglês .....	170
TABELA 28.	Frequência relativa das funções do sistema de MENSAGEM no manual técnico em português .....	172
TABELA 29.	Frequência relativa das funções do sistema de TEMA no manual técnico em português .....	175
TABELA 30.	Frequência relativa das funções do sistema de MENSAGEM no manual técnico em inglês .....	176
TABELA 31.	Frequência relativa das funções do sistema de TEMA no manual técnico em inglês .....	178
TABELA 32.	Frequência relativa das funções do sistema de TEMA nos manuais em português e inglês ...	179
TABELA 33.	Frequência relativa das funções do sistema de MODO e Funções Discursivas nas cartilhas e <i>folders</i> de divulgação em português e inglês .....	182
TABELA 34.	Frequência relativa das funções do sistema de MENSAGEM na cartilha e <i>folder</i> de divulgação em português .....	184
TABELA 35.	Frequência relativa das funções do sistema de TEMA nas cartilhas e <i>folders</i> de divulgação em português .....	186
TABELA 36.	Frequência relativa das funções do sistema de MENSAGEM na cartilha e <i>folder</i> de divulgação em inglês .....	187
TABELA 37.	Frequência relativa das funções do sistema de TEMA na cartilha e <i>folder</i> de divulgação em inglês .....	190

## LISTA DE GRÁFICOS

GRÁFICO 1. Frequência de ocorrência das principais classes de palavras no artigo acadêmico em inglês e em português.....	123
GRÁFICO 2. Frequência de ocorrência das principais classes gramaticais no manual técnico em inglês e em português.....	124
GRÁFICO 3. Frequência de ocorrência das principais classes gramaticais para a cartilha e <i>folder</i> de divulgação em inglês e em português.....	124
GRÁFICO 4. Associação das frequências absolutas dos nódulos “doença falciforme” e “anemia falciforme” em português.....	132
GRÁFICO 5. Associação das frequências absolutas dos nódulos em inglês.....	134
GRÁFICO 6. Associação das frequências absoluta dos nódulos “dor” e “ <i>pain</i> ” por tipo de texto .....	135
GRÁFICO 7. Associação das frequências dos nódulos “pacientes” e “ <i>patients</i> ” por tipo de texto .....	137

## LISTA DE SIGLAS E ABREVIATURAS

CEHMOB-MG	Centro de Educação e Apoio para Hemoglobinopatias
DeCS	Descritores em Ciências da Saúde
DF	Doença Falciforme
ET	Estudos da Tradução
FALE	Faculdade de Letras
FM	Faculdade de Medicina
Hb A	Hemoglobina A
Hb C	Hemoglobina C
Hb D	Hemoglobina D
Hb S	Hemoglobina S
LC	Linguística de Corpus
LETRA	Laboratório Experimental de Tradução
LSF	Linguística Sistêmico-Funcional
MeSH	Medical Subject Headings
NUPAD	Núcleo de Ações e Pesquisa em Apoio Diagnóstico da Faculdade de Medicina da UFMG
OPAS/OMS	Organização Pan-Americana da Saúde/Organização Mundial da Saúde
Poslin	Programa de Pós-Graduação em Estudos Linguísticos da Faculdade de Letras da Universidade Federal de Minas Gerais
SBthal	Hemoglobinopatia SBthal
SC	Hemoglobinopatia SC
SCD	Sickle Cell Disease
SD	Hemoglobinopatia SD
SS	Anemia Falciforme
S-Talassemia	Hemoglobinopatia S-Talassemia
Thal	Thalassemia
UFMG	Universidade Federal de Minas Gerais
UFOP	Universidade Federal de Ouro Preto

## SUMÁRIO

INTRODUÇÃO.....	19
1 FUNDAMENTAÇÃO TEÓRICA.....	26
1.1 A Linguística de Corpus e Sua Interface com os Estudos da Tradução.....	26
1.2 A Teoria Sistêmico-Funcional.....	31
1.2.1 O sistema de TRANSITIVIDADE.....	36
1.2.2 O sistema de MODO.....	39
1.2.3 O sistema de TEMA e MENSAGEM.....	42
1.3 Perfil Metafuncional dos Tipos de Texto.....	48
1.4 Estudos Multilíngues.....	53
1.5 Estudos de Registro e Estudos de Corpora.....	56
1.6 A Tipologia de Textos Baseada nos Parâmetros do Contexto.....	63
2 METODOLOGIA.....	67
2.1 Metodologia de Coleta de Dados.....	67
2.1.1 Compilação dos <i>corpora</i> .....	68
2.1.2 Coleta de dados.....	70
2.1.2.1 Extração de dados através do ambiente de programação R.....	71
2.1.2.2 Anotação dos <i>corpora</i> em classes de palavras: etiquetador morfossintático Treetagger.....	81
2.1.2.3 Anotação e etiquetamento dos <i>corpora</i> por meio do UAM CorpusTool.....	86
2.2 Metodologia de Análise dos Dados.....	96
2.2.1 Procedimentos de análise a partir dos dados proveniente dos <i>corpora</i> .....	97
2.2.2 Procedimentos de análise sistêmico-funcional.....	100
2.2.2.1 Procedimento de análise de textos em contexto de uso.....	101
2.2.2.2 Procedimento de análise do perfil metafuncional dos textos.....	104
3 ANÁLISE DOS DADOS.....	113
3.1 Análise dos <i>Corpora</i> .....	113
3.1.1 Dimensão dos <i>corpora</i> .....	113
3.1.2 Anotação dos <i>corpora</i> em classes de palavras.....	118
3.1.3 Análise de nódulos extraídos dos <i>corpora</i> em função da língua e do tipo de texto.....	129
3.1.3.1 Análise dos nódulos “doença falciforme” e “anemia falciforme” “sickle cell disease” e “sickle cell anemia/anaemia”.....	130
3.1.3.1.1 Análise dos nódulos nos <i>corpora</i> em português.....	130
3.1.3.1.2 Análise dos nódulos nos <i>corpora</i> em inglês.....	132
3.1.3.2 Análise dos nódulos dor”, “pain”, “paciente”, “patient” e “you”.....	134
3.2 Caracterização dos Textos sob a Perspectiva da Linguística Sistêmico-Funcional.....	138
3.2.1 A variável CAMPO.....	142
3.2.2 A variável SINTONIA.....	147
3.2.3 A variável MODO.....	149

3.3	Descrição dos Perfis Metafuncionais dos Textos .....	152
3.3.1	Artigo acadêmico: análise ideacional, interpessoal e textual dos textos em português e inglês... 155	
3.3.1.1	<i>Análise ideacional dos artigos acadêmicos em português e em inglês</i> .....	155
3.3.1.2	<i>Análise Interpessoal dos artigos acadêmicos em português e em inglês</i> .....	158
3.3.1.3	<i>Análise textual dos artigos acadêmicos em português e em inglês</i> .....	159
3.3.2	Manuais técnicos: análise ideacional, interpessoal e textual dos textos em português e inglês ... 167	
3.3.2.1	<i>Análise ideacional dos manuais técnicos em português e em inglês</i> .....	167
3.3.2.2	<i>Análise Interpessoal dos manuais técnicos em português e em inglês</i> .....	170
3.3.2.3	<i>Análise Textual do manual técnico em português e em inglês</i> .....	171
3.3.3	Cartilhas e <i>folders</i> de divulgação: análise ideacional, interpessoal e textual dos textos em português e inglês .....	179
3.3.3.1	<i>Análise ideacional das cartilhas e folders de divulgação em português e em inglês</i> .....	179
3.3.3.2	<i>Análise interpessoal das cartilhas e folders de divulgação em português e em inglês</i> .....	181
3.3.3.3	<i>Análise textual das cartilha e folder de divulgação em português e em inglês</i> .....	183
3.3.4	Elaboração do texto do tipo HABILITAR a partir do template (modelo).....	196
4	DISCUSSÃO DOS RESULTADOS .....	202
4.1	Revisão dos Achados.....	202
4.2	Revisão dos Achados à Luz da Literatura.....	205
4.2.1	Interface entre as diferentes áreas do conhecimento .....	205
4.2.2	Análise de <i>corpora</i> comparáveis.....	206
4.2.3	Caracterização dos textos sob a perspectiva da Linguística Sistêmico-Funcional .....	210
4.2.4	Descrição do perfil metafuncional no ambiente multilíngue.....	212
5	CONCLUSÃO .....	215
	REFERÊNCIAS .....	221
	ANEXOS.....	226
	Anexo 1. Relação dos <i>Scripts</i> Utilizados na Pesquisa .....	227
	Anexo 2. Tabelas Construídas com Base nos Dados do UAM CorpusTool.....	236
	Anexo 3. Textos que Compuseram os <i>Corpora</i> da Descrição do Perfil Metafuncional .....	244

# INTRODUÇÃO

## INTRODUÇÃO

*In any given language processing task, there will be one or more domains to be modeled as part of our general taxonomy (...). It might be objected that this view leaves no room for scientific or metaphysical models — for example, that we do not allow for the possibility that science has advanced our understanding of the world. This objection would be misplaced: such models are construed in the ideation base as domain models within the overall meaning potential.*

M. A. K. HALLIDAY. & C. M. I. M. MATTHIESSEN,  
*Construing experience through meaning.*

**E**sta tese insere-se no escopo de pesquisas que visam contribuir para a elaboração de um modelo que possibilite a produção e tradução de textos no ambiente multilíngue, entendido como um mapa de “configurações distintas de produção de significados no contexto de cultura cuja instanciamento demanda a participação de um número variável de sistemas linguísticos” (FIGUEREDO, 2013)<sup>1</sup>. Essa perspectiva se insere na proposta de articulação de diferentes campos disciplinares e subáreas, propiciada pela teoria Sistêmico-Funcional, considerando-se o enfoque e o interesse de cada um deles através do estudo do contato linguístico (MATTHIESSEN; TERUYA; WU, 2008).

Nesta tese foi adotada uma abordagem de interface entre a Linguística de Corpus (MCENERY; XIAO, 2007; GRANGER, 2008) – no que toca às análises semiautomáticas de *corpora* comparáveis – e os Estudos da Tradução (MUNDAY, 2001), mediante o aporte da Linguística Sistêmico-Funcional (HALLIDAY; MATTHIESSEN, 2004; MATTHIESSEN; TERUYA; WU, 2008; FIGUEREDO, 2011). Mais especificamente, com base nas variáveis do contexto (MATTHIESSEN; TERUYA; WU, 2008; FIGUEREDO, 2011), no escopo da Linguística Sistêmico-Funcional, a tese apresenta um estudo de padrões de uso linguístico (SINCLAIR, 1991; BERBER SARDINHA, 2004) em *corpora* constituídos de três tipos de texto rotulados no contexto de cultura de acordo com distintas configurações explícitas de sintonia (*tenor*) entre interlocutores (HALLIDAY, 1964). Os textos que compuseram os *corpora* foram selecionados com base nos rótulos a eles atribuídos, a saber:

---

<sup>1</sup> Comunicação a ser apresentada em: FIGUEREDO, G. P. Developing tools to model meaning in multilingual environments. In: SIMPÓSIO INTERNACIONAL DE LINGUÍSTICA FUNCIONAL, 2, UFSCar, São Carlos, 14-16 ago. 2013. Comunicação Pessoal.

1. Artigos acadêmicos (interação especialista – especialista);
2. Manuais técnicos (interação especialista – técnico da área da saúde); e
3. Cartilhas e *folders* de divulgação (interação especialista – leigo).

Para a seleção das amostras de textos, adotou-se metodologia de Neumann (2005), a qual foi elaborada com base em Biber (1990). Os *corpora* analisados são comparáveis bilíngues, coletados de conjuntos de textos autênticos e não traduzidos da área da medicina (subárea triagem neonatal para anemia falciforme), em português e em inglês, provenientes do banco de dados do Núcleo de Ações em Pesquisa em Apoio Diagnóstico da Faculdade de Medicina da Universidade Federal de Minas Gerais (NUPAD/FM/UFMG)<sup>2</sup>.

Constituem os objetivos gerais desta pesquisa:

1. Desenvolver uma abordagem teórica e metodológica propiciada pela interface entre a Linguística de Corpus e os Estudos da Tradução, com subsídios da Linguística Sistêmico-Funcional, para estudar como textos que instanciam os tipos de textos – artigo acadêmico, manual técnico e cartilha e *folders* de divulgação – são produzidos em português e em inglês;
2. Elaborar modelos para a tradução e produção textual no ambiente multilíngue no âmbito de Projeto de Cooperação entre o LETRA/FALE/UFMG e o NUPAD/FM/UFMG;

No que diz respeito ao campo disciplinar, esta tese está inserida nos estudos puros da tradução (HOLMES, 1972) (*cf.* FIG. 1). Em seu trabalho, Holmes (1972) considera os Estudos da Tradução um campo independente dos estudos linguísticos e apresenta uma proposta programática para ele dividida em estudos aplicados e puros – estes últimos compreendendo os estudos descritivos e os estudos teóricos. Este estudo se baseia nos resultados provenientes dos *corpora*, com o aporte teórico da linguística sistêmico-funcional, no intuito de propor um modelo que sirva para explicar como os textos são produzidos em diferentes línguas e

---

<sup>2</sup> O Núcleo de Ações e Pesquisa em Apoio Diagnóstico (NUPAD), criado em 1993, é órgão complementar da Faculdade de Medicina da Universidade Federal de Minas Gerais e credenciado em 2001 pelo Ministério da Saúde como Serviço de Referência em Triagem Neonatal do Estado. O núcleo é reconhecido tanto no cenário nacional quanto no cenário internacional como centro de referência em estudos diagnósticos da anemia falciforme. Para mais informações, *cf.* o sítio eletrônico: <[www.nupad.medicina.ufmg.br](http://www.nupad.medicina.ufmg.br)>.

traduzidos em ambiente multilíngue. Portanto, conforme ilustra a FIG. 1, esta tese se localiza nos estudos puros, descritivos, orientados ao produto (HOLMES, 1972).

A pesquisa também enseja uma contribuição para os estudos aplicados, no que toca as ferramentas de auxílio à tradução, com aplicação da tecnologia da informação, no sentido de subsidiar, com base em análises semiautomáticas de *corpora* comparáveis, uma proposta de *pipeline* ou linha de produção textual que direcione a tradução e elaboração de documentos em diferentes tipos de textos. Com esse direcionamento, espera-se automatizar alguns processos, uniformizar/padronizar termos e agilizar a produção multilíngue de textos (cf. FIG. 1).

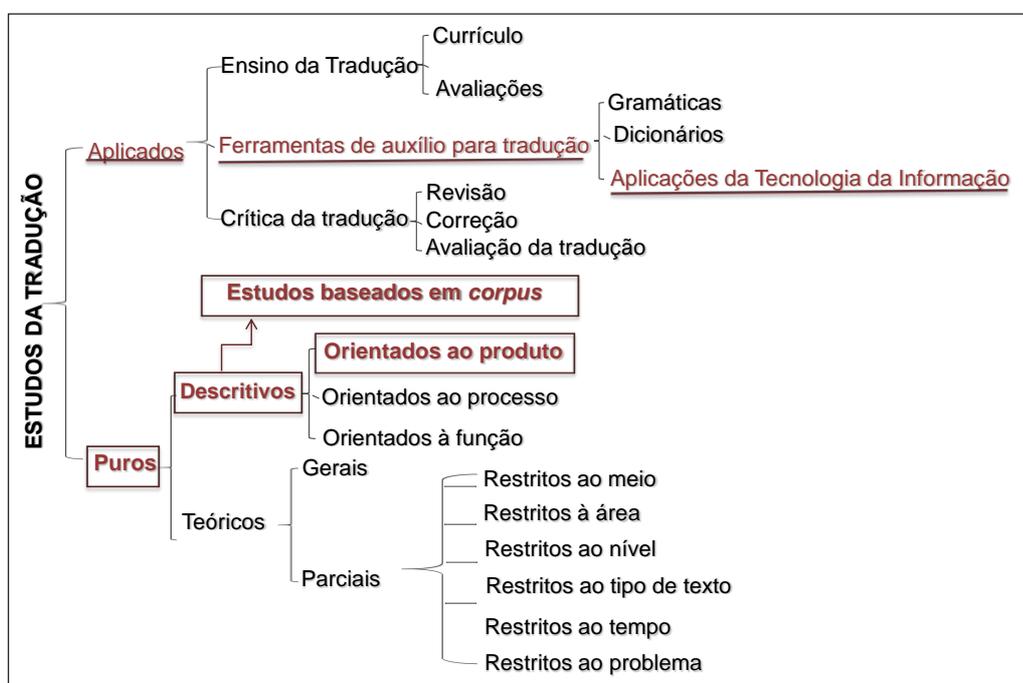


FIGURA 1. Contextualização no Campo Disciplinar

Fonte: adaptação de Pagano e Vasconcelos (2003), com base em Holmes (1972, p. 10), tradução minha.

Além do seu caráter interdisciplinar, a pesquisa em tela é também interinstitucional. Em função da sua inserção no âmbito de um projeto de cooperação entre o LETRA e o NUPAD, foi possível a coleta de textos autênticos para a compilação dos *corpora*. Além disso, na condição de profissional inserida no NUPAD/FM/UFMG como tradutora, a autora desta tese conta com conhecimentos tácitos e com uma interação com diversos autores/autoridades na

temática da doença falciforme, o que vem a contribuir para a validação de resultados e identificação dos padrões de uso linguístico a serem investigados.

Essa perspectiva interdisciplinar e interinstitucional emergiu de discussões entre a autora, a professora da Faculdade de Letras, Profa. Dra. Adriana Pagano, e o diretor geral do NUPAD, Prof. Me. José Nelio Januario. A partir dessas discussões, verificou-se a necessidade de orientar a produção textual de diferentes tipos de textos, dada a crescente internacionalização do Núcleo, que vem desenvolvendo parcerias com os EUA e diversos países da África e da Europa. Essas parcerias implicam uma demanda maior por serviços de tradução de textos sobre a anemia falciforme e, em muitos casos, pela produção simultânea de textos e documentos de diferentes tipos e em línguas distintas. Nesse sentido, a identificação dos padrões de uso linguístico e a descrição metafuncional dos textos têm muito a contribuir para o NUPAD no sentido de viabilizar processos de produção monolíngue/bilíngue que não apenas sejam ágeis, uniformes e automatizados, mas também capazes de atender ao tipo de texto esperado.

Com base em *corpora* comparáveis, esta pesquisa procurou especificamente:

1. Caracterizar os registros a partir de três tipos de texto – artigos acadêmicos, manuais técnicos e cartilhas;
2. Propor a descrição do perfil metafuncional de cada um dos tipos de textos a partir da análise dos sistemas de MODO, TRANSITIVIDADE e TEMA e MENSAGEM; e
3. Elaborar um *template* (modelo) para a produção de um texto do tipo HABILITAR no ambiente multilíngue, com base nos resultados da análise da cartilha e *folder* de divulgação com vistas a sua futura utilização como um primeiro passo para a elaboração de um *pipeline* (percurso ou linha de produção textual) que direcione a tradução e elaboração de diferentes tipos de textos e em diferentes línguas na área da Saúde – mais especificamente na subárea de triagem neonatal para anemia falciforme<sup>3</sup>.

---

Para esta pesquisa, partiu-se do pressuposto de que padrões de uso linguístico mapeados de forma semiautomática (Linguística de Corpus) e identificados a partir de significados metafuncionais (Linguística Sistêmico-funcional) permitem a caracterização de registros. A partir desse pressuposto surgiram duas perguntas de pesquisa:

1. Quais itens lexicais e gramaticais específicos extraídos por anotação morfossintática revelam padrões de uso linguístico?
2. Em que medida a descrição do perfil metafuncional de textos subsidia a elaboração de um *template* de texto vinculado ao processo sociosemiótico HABILITAR que possa ser utilizado para a produção e/ou tradução de textos do tipo cartilha e folder de divulgação?

Para a resposta as perguntas acima e o desenvolvimento da proposta desta Tese, foi fundamental o apoio de outras áreas do conhecimento, como a Linguística Computacional – para a extração e manipulação dos dados dos *corpora* – e a Estatística – para a interpretação dos dados extraídos. Em primeiro lugar, os textos foram anotados e manipulados através da utilização do ambiente de programação R<sup>4</sup>, com o qual foram extraídos padrões de co-ocorrência entre itens lexicais e itens gramaticais específicos. Na sequência, os *corpora* foram anotados em classes de palavras por meio do etiquetador morfossintático Treetagger®. Por fim, excertos de mil palavras de cada um dos tipos de texto em inglês e em português foram classificados de acordo com a tipologia da língua no contexto de cultura (cf. MATTHIESSEN; TERUYA; WU, 2008). Esses excertos foram anotados por meio do programa UAM CorpusTool®, com base nas metafunções ideacional, interpessoal e textual (HALLIDAY; MATTHIESSEN, 2004) e analisados de forma semiautomática. Por meio desse programa é possível anotar, extrair e analisar os dados dos *corpora* estudados.

O presente texto está organizado em quatro capítulos e uma conclusão, além desta Introdução. O **Capítulo 1 – Fundamentação Teórica** apresenta os principais conceitos que nortearam o desenvolvimento deste estudo, além das pesquisas nas quais este texto foi baseado. O **Capítulo 2 – Metodologia** descreve os procedimentos de coleta dos textos e compilação dos *corpora*, o ambiente de programação R e o tratamento dos *corpora*, bem como os procedimentos de análise dos dados. São apresentados também os procedimentos de descrição

---

<sup>4</sup> Faz-se mister ressaltar que o R é um ambiente de programação estatística com o qual é possível programar, por meio de *scripts*, extrair e analisar os dados provenientes dos *corpora*.

sistêmico-funcional do perfil metafuncional dos diferentes tipos de textos que compuseram os *corpora* desta pesquisa, por meio do programa UAM CorpusTool. O **Capítulo 3 – Análise dos Dados** apresenta a análise e interpretação dos dados a partir da interface entre a Linguística de Corpus, Estudos da Tradução e Linguística Sistêmico-Funcional. O **Capítulo 4 – Discussão dos Dados** tece discussões sobre os resultados, buscando um diálogo com a literatura de modo a apresentar a relevância desta pesquisa para a interface entre a Linguística de Corpus e os Estudos da Tradução, com o aporte teórico-metodológico da Linguística Sistêmico-Funcional, conforme explicitados nesta Introdução. Por fim, a **Conclusão** contempla os principais resultados e sua avaliação no que toca aos objetivos propostos nesta tese, além de indicação de pesquisas futuras que podem ser realizadas a partir dos achados e lacunas desta.

# **CAPÍTULO 1**

## **FUNDAMENTAÇÃO TEÓRICA**

## 1 FUNDAMENTAÇÃO TEÓRICA

**E**ste capítulo tem início com as perspectivas da Linguística de Corpus (doravante, LC) (MCENERY; XIAO, 2007; GRANGER, 2008) e a sua interface com os Estudos da Tradução (MUNDAY, 2001). Em seguida, na Seção 1.2, fala-se da Linguística Sistêmico-Funcional (doravante, LSF) e os Estudos da Tradução baseados em *corpora* enquanto abordagens que se complementam no processo de análise tanto de textos originais quanto de (suas) traduções (HALLIDAY; McINTOSH; STREVENSON, 1964; CATFORD, 1965; MATTHIESSEN, 2001; PAGANO; VASCONCELLOS, 2005; MUNDAY, 2001). A Seção 1.3 ilustra o perfil metafuncional dos tipos de textos, caracterizado pelas escolhas experienciais, interpessoais e textuais identificadas em cada um nos sistemas de TRANSITIVIDADE, MODO, e TEMA e MENSAGEM, respectivamente (HALLIDAY; MATTHIESSEN, 1999; FIGUEREDO, 2011). Na Seção 1.4, apresenta-se o novo campo disciplinar que surgiu recentemente no percurso da LSF: os Estudos Multilíngues (MATTHIESSEN; TERUYA; WU, 2008). Na Seção 1.5, descrevem-se os estudos de registro e os estudos de *corpora* (NEUMANN, 2008). Na Seção 1.6, enfocam-se propostas de definição de registro a partir da tipologia de textos baseada nos parâmetros das variáveis do contexto (HALLIDAY; MATTHIESSEN, 1999).

### 1.1 A Linguística de Corpus e Sua Interface com os Estudos da Tradução

A LC é considerada um ramo da Linguística que se baseia em compilações de coleções de textos autênticos em formato eletrônico com o objetivo de consubstanciar suas teorias e descobertas em bases quantitativas/estatísticas (BAKER, 1995). A LC pode ser considerada um campo de investigação empírica que estuda a língua em uso, por meio da compilação e exploração dos *corpora* (BERBER SARDINHA, 2004). Ela foi aqui utilizada por viabilizar uma abordagem metodológica e garantir a exploração da linguagem verbal por intermédio de exemplos reais, tornando possível um estudo empírico de *corpora* comparáveis (KENNING, 2010).

Baker (1995, p. 225) define *corpus* como uma coleção de língua natural em formato eletrônico, de fontes diversas, reunida a partir de critérios e finalidades específicas, passível

de receber tratamento automático e/ou semiautomático. Sob uma perspectiva semelhante, Tognini-Bonelli (2001) ressalta que *corpus* não apenas se refere a uma compilação de língua autêntica, como também a uma seleção de textos com base em determinados critérios explícitos, além de ser passível de análise automática ou semiautomática

Os trabalhos seminais de Baker (1993, 1995, 1996) marcaram o início da interface entre a LC e os Estudos da Tradução. Até então, o potencial de pesquisa representado pelos *corpora* já se encontrava relativamente disseminado entre os pesquisadores dessa área, mas os estudos se restringiam a aspectos da descrição e comparação de línguas, geralmente negligenciando o papel ou a relevância do texto traduzido *per se* ou em comparação com o texto de partida.

Para iniciar uma pesquisa com base na LC, é necessário determinar o tipo e a composição do *corpus*. Em se tratando da composição, os *corpora*, de acordo com Sinclair (2001, p. xi), podem ser divididos em *corpora* de pequenas ou de grandes dimensões. Os *corpora* de pequenas dimensões se distinguem dos *corpora* de grandes dimensões não exatamente por uma questão de tamanho, e sim por questão da representatividade para a pesquisa, considerando-se, dentre outros fatores, os objetivos e a metodologia utilizada no processamento e análise dos dados (*e.g.*, processamento manual, automático e/ou semiautomático). No caso da presente pesquisa, os *corpora* analisados são considerados de pequenas dimensões: compreendem 18 mil palavras distribuídas em três diferentes tipos de texto e em duas línguas (*cf.* Seção 2.1).

Em se tratando dos tipos de *corpora*, Baker (1995, p. 225) propõe a classificação dos *corpora* em três categorias:

1. *Corpora* paralelos, constituídos por textos-fonte em uma língua A e suas traduções em uma língua B;
2. *Corpora* multilíngues, que consistem em um conjunto de *corpora* monolíngues; e
3. *Corpora* comparáveis, formados por dois *subcorpora*, um *corpus* monolíngue de textos originais e um *corpus* monolíngue de textos traduzidos.

Tognini-Bonelli (2001), por sua vez, classifica o *corpus* em quatro tipos:

1. monolíngue simples, com textos produzidos originalmente em uma língua;

2. monolíngue comparável, com textos originalmente escritos e traduzidos para uma mesma língua;
3. paralelo bilíngue ou multilíngue, com um texto e sua(s) tradução(ões); e
4. bilíngue comparável, com grupos de textos selecionados em duas línguas, seguindo um critério de tipologia textual.

Ana Frankenberg-Garcia (2006) propõe quatro tipos de *corpora*: comparáveis bilíngues, comparáveis monolíngues, paralelos unidirecionais e paralelos bidirecionais. A autora explica que os *corpora* paralelos bidirecionais formam dois *corpora* comparáveis monolíngues (traduzidos e não traduzidos na língua A; traduzidos e não traduzidos na língua B) e dois *corpora* comparáveis bilíngues (não traduzidos na língua A e não traduzidos na língua B; traduzidos na língua A e traduzidos na língua B). Este tipo de classificação é denominada por Vela e Hansen-Schirra (2006) de *corpus* combinado paralelo-comparável ou combinado.

Johansson (1998) desenvolveu um dos trabalhos pioneiros utilizando um *corpus* combinado, embora não o tenha classificado dessa forma. Esse autor sugere três classificações:

1. *Corpora* comparáveis bilíngues/multilíngues ou de textos originais comparáveis;
2. *Corpora* de tradução ou de textos originais e respectivas traduções (bilíngue ou multilíngue); e
3. *Corpora* comparáveis monolíngues ou de textos originais e traduzidos em uma mesma língua.

Hansen (2002) utiliza os termos comparável e paralelo, além de descrever dois tipos de *corpus* combinado: o bilíngue e o multilíngue.

Considerando-se o crescente número de pesquisas realizadas com o aporte da Linguística de Corpus desde a década de 1980, McEnery e Xiao (2007) desenvolveram um trabalho em que exploram a importância da utilização de *corpora* comparáveis para os Estudos da Tradução e também para os Estudos Contrastivos, além de apresentarem algumas considerações acerca de questões terminológicas em relação às categorizações dos tipos de *corpora* propostas por alguns pesquisadores. De acordo com McEnery e Xiao (2007, p. 1), os *corpora* comparáveis são recursos úteis às pesquisas de tradução e pesquisas para os Estudos Contrastivos, oferecendo utilidades e possibilidades específicas, como:

- (a) possibilidade de aquisição de novos conhecimentos em relação às diferentes línguas que estão sendo comparadas, conhecimentos esses que provavelmente seriam ignorados através de estudos baseados em *corpora* monolíngues;
- (b) realização de uma variedade de atividades comparativas, além de propiciar um conhecimento maior das diferenças entre as línguas constituintes dos *corpora*;
- (c) estabelecimento das diferenças entre textos-fonte e suas traduções e entre os textos originais e não originais;
- (d) possibilidade de aplicações práticas na lexicografia, ensino de línguas e tradução, por exemplo.<sup>5</sup>

Granger (2003) enfoca a importância do *corpus* eletrônico considerando-se a interface entre os Estudos da Tradução e a Linguística Comparada nas últimas décadas. Em seu estudo, a autora discute as características dos diferentes tipos de *corpora* monolíngues e multilíngues e sugere uma unificação terminológica, ou seja, a autora propõe uma tipologia geral de *corpora*. Essa autora afirma que a utilização de *corpus* aproxima os pesquisadores da Linguística de Corpus e dos Estudos da Tradução, uma vez que assim fazem pesquisa com os mesmos tipos de dados, utilizando os mesmos programas concordanceadores, apesar de terem objetivos diferentes. De acordo com a autora, apesar de toda a utilidade das pesquisas baseadas em *corpora*, é importante apontar que a metodologia baseada em *corpora* tem limitações para os Estudos da Tradução, uma vez que ainda há poucos *corpora* comparáveis de textos traduzidos, há necessidade de manipulação (automática, semiautomática ou manual) dos *corpora* dependendo do objetivo específico de cada pesquisa e é necessária teoria que explique os dados encontrados. Em se tratando dessa última limitação, esta tese se apoia na Linguística Sistêmico-Funcional (cf. Seção 1.2 a seguir) como base teórica para a classificação de alguns dados e explicação dos achados.

Halliday e Webster (2009) também propuseram uma tipologia geral de *corpus*, a qual serviu de base para a classificação dos *corpora* desta pesquisa (cf. FIG. 2). De acordo com esses autores, o *corpus* de uma pesquisa pode ser compilado de acordo com a especificidade dela, englobando uma diversidade de tipos de texto que podem ser contrastados, podendo ser classificados como *corpora* paralelos, comparáveis ou combinados, considerando-se os procedimentos para a compilação do tipo de *corpus* e o contexto de uso.

---

<sup>5</sup> Minha tradução do original:

- “they give new insights into the languages compared - insights that are likely to be unnoticed in studies of monolingual corpora;
- they can be used for a range of comparative purposes and increase our understanding of language-specific, typological and cultural differences, as well as of universal features;
- they illuminate differences between source texts and translations, and between native and non-native texts;
- they can be used for a number of practical applications, e.g. in lexicography, language teaching, and translation.”

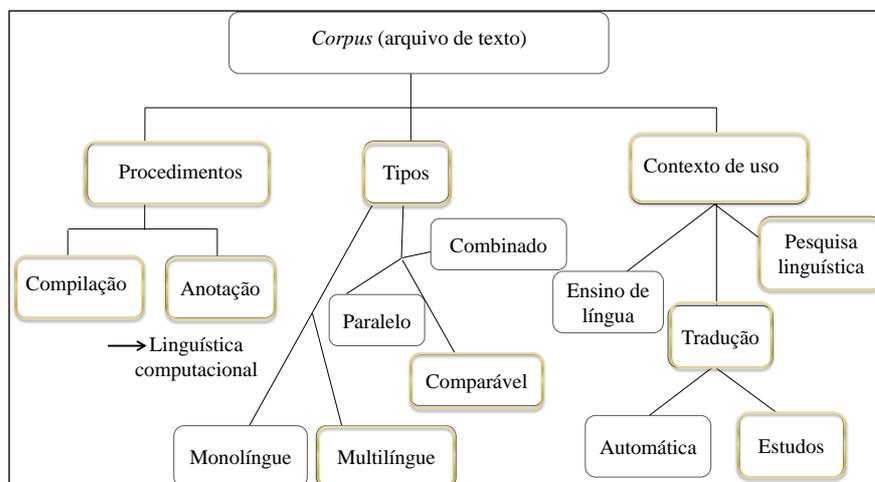


FIGURA 2. Diagrama ilustrando os tipos de *corpus*

Fonte: Halliday e Webster (2009).

O tipo de *corpus* utilizado para a realização desta pesquisa seguiu a categorização dos *corpora* adotada por McEnery e Xiao (2007), as considerações de Granger (2003) e a proposta de Halliday e Webster (2009) e foi classificado como comparável, multilíngue, sendo compilado e anotado por meio de ferramentas computacionais, com o objetivo de investigar padrões de uso linguístico em três tipos de textos (mais especificamente, da medicina – subárea triagem neonatal para anemia falciforme) com vistas à caracterização de cada tipo sob a perspectiva das variáveis do contexto de situação – campo (*field*), sintonia (*tenor*) e modo (*mode*)<sup>6</sup>. Para a realização desta pesquisa foi considerado enquanto contexto de uso do *corpus* a Pesquisa Linguística e os Estudos da Tradução. É importante lembrar que esta pesquisa utiliza o *corpus* como objeto de estudo e se pautou nas vertentes teóricas não excludentes, ou seja, pesquisa de base em *corpus* e conduzida pelo *corpus*. Destaca-se, portanto, a importância da utilização da LC por viabilizar uma abordagem metodológica e garantir a exploração da língua verbal por meio de exemplos reais, o que torna possível um estudo empírico de *corpora* comparáveis.

Destacam-se os trabalhos que vêm sendo desenvolvidos com base em diferentes tipos de *corpus*, em projetos de pesquisa no escopo do Laboratório Experimental de Tradução (LETRA) da FALE-UFMG. No que toca ao projeto “Tradução e Linguística de Corpus, Corpora, Gênero e (Re)textualização: Interfaces nos Estudos da Tradução, Corpora, Cognição e Discurso: uma proposta interdisciplinar para os Estudos da Tradução a partir de bancos eletrônicos de dados e organização temática e tradução”, destacam-se as seguintes pesquisas:

<sup>6</sup> O construto do contexto de situação e de suas variáveis é definido/discutido na Seção 1.6.

Pagano (2000, 2005, 2007); Pagano, Figueredo e Braga (2007); Lacerda e Pagano (2008); Figueredo e Pagano (2009). Vários dos trabalhos com base na Linguística de Corpus citados anteriormente foram elaborados com o aporte teórico-metodológico da Linguística Sistêmico-Funcional. Alguns pressupostos básicos dessa teoria são apresentados a seguir, sobretudo no que toca ao perfil metafuncional dos tipos de texto (HALLIDAY, 1978; CAFFAREL; MARTIN; MATTHIESSEN, 2004), aos estudos de registro e sua interface com os estudos de *corpora* e a tipologia de textos baseada nas variáveis do contexto e a modelagem de base ideacional (HALLIDAY; MATTHIESSEN, 1999).

## 1.2 A Teoria Sistêmico-Funcional

A Linguística Sistêmico-Funcional e os Estudos da Tradução baseados em *corpora* são abordagens que se complementam no processo de análise tanto de textos originais quanto de (suas) traduções (HALLIDAY; McINTOSH; STREVENSON, 1964; CATFORD, 1965; MATTHIESSEN, 2001; PAGANO; VASCONCELLOS, 2005; MUNDAY, 2001). Ambas as abordagens enfatizam a língua em uso e apresentam características probabilísticas. Esse diálogo entre a LC e LSF pode ser considerado proveniente da necessidade de a análise dos textos originais e traduzidos ser realizada por meio de questões linguísticas, incluindo-se o contexto de produção do significado. Em termos mais específicos, esta pesquisa interessa-se pela contribuição que a LC pode oferecer aos Estudos da Tradução por meio de suas ferramentas metodológicas, com o aporte teórico-metodológico da LSF.

Halliday (1978) considera que a teoria da tradução pode ser inserida entre aquelas do campo da Linguística Comparada— mais especificamente, a Linguística Descritiva Comparada, que consiste em primeiro descrever para só então comparar padrões encontrados, contrastando-os. Sabe-se, pois, que a LSF, por ser uma teoria abrangente, permite a análise textual e a descrição textual necessárias para comparar diferentes textos em diferentes línguas.

Cabe ressaltar ainda que apesar da proximidade da relação entre a Linguística descritiva Comparada e a LSF, a descrição não é teoria. No que toca à Linguística, sabe-se que a descrição deve somente pertencer a uma parte da teoria linguística, sendo a teoria a responsável pela forma como a língua funciona. De acordo com Halliday (2002), os métodos de descrição linguística derivam de uma teoria linguística geral. Segundo o autor, descrever implica estabelecer relação entre a teoria e os dados obtidos por meio da própria teoria.

Teich (1999) propõe uma forma de as descrições auxiliarem na análise contrastiva, uma vez que as línguas apresentam uma tendência de se assemelharem em relação a sistemas (paradigmas) e de se diferenciarem no que tange à realização sintática. Em outros termos, os sistemas gramaticais de baixa *delicadeza* tendem a indicar semelhanças através das línguas, ao passo que os sistemas de *delicadeza* mais alta tendem a se diferenciarem. Por conseguinte, a tendência das línguas é de mostrarem-se mais semelhanças nos estratos menos abstratos (no gramatical, por exemplo). Já nos estratos mais abstratos (no estrato semântico, por exemplo), as semelhanças entre as línguas tornam-se menos evidentes.

A teoria Linguística Sistêmico-Funcional hallidayana pode ser considerada como o resultado de uma evolução do pensamento funcionalista desde os trabalhos de Malinowski (1945) e Firth (1957) até Halliday (1978) no que se refere à linguagem e à descrição da linguagem como semiótica social. Essa teoria usa a linguagem para modelar e ilustrar a realidade, por permear a vida social e desempenhar um papel central nos sistemas ideológicos e na formação sociopolítica (HASAN, 1996). Devido à complexidade da organização do sistema linguístico – e de suas relações com o contexto, por conseguinte –, essa teoria examina o sistema linguístico a partir de múltiplas dimensões:

1. Ordem: organização hierárquica da língua em morfemas, palavras, grupos, orações e complexos oracionais;
2. Estratificação: organização da linguagem em estratos fonético, fonológico, léxico-gramatical, semântico e contextual;
3. Metafunções: estabelecimento da construção da experiência humana (ideacional), construção das realizações sociais (interpessoal) e construção do discurso (textual);
4. Realização: relações entre os estratos por meio da expressão (estrato fonológico) e conteúdo (estratos semântico e léxico-gramatical); e
5. Instanciação: relação entre o sistema linguístico e o tipo de texto (oral ou escrito).

As contribuições da sistêmica para esta pesquisa se concentram em mostrar a organização dos textos dos *corpora* com base nos estudos de registro (por meio das variáveis do contexto de situação e seus respectivos parâmetros) e da tipologia de textos baseada nos parâmetros do contexto (MATTHIESSEN; TERUYA; WU, 2008), bem como a descrição do perfil metafuncional (sistemas de TRANSITIVIDADE, MODO, e TEMA e MENSAGEM, com o objetivo de gerar a base para a construção de um *pipeline* que possa orientar a tradução e a

produção textual no ambiente multilíngue. Halliday e Matthiessen (2004, p. 24) definem a linguagem como um sistema semiótico complexo caracterizado pelos estratos citados anteriormente que se distribuem desde um estrato mais abstrato (*i.e.*, a ideologia) a um estrato mais concreto da expressão (*i.e.*, o fonético). Assim, o sistema da linguagem se apresenta organizado em estratos distintos por uma ordem de abstração nos níveis fonológico, léxico-gramatical e semântico. Esses estratos relacionam-se por meio da realização. A FIG. 3 mostra que a gramática é realizada no estrato fonológico. Já o estrato gramatical realiza o estrato semântico que por sua vez realiza o contexto, sendo o contexto realizado na semântica. Cada estrato pode ser considerado como uma rede de opções inter-relacionadas em som, fraseado e significados, as quais opções são realizadas como estruturas da língua (HALLIDAY; MATTHIESSEN, 2004, p. 4-5).

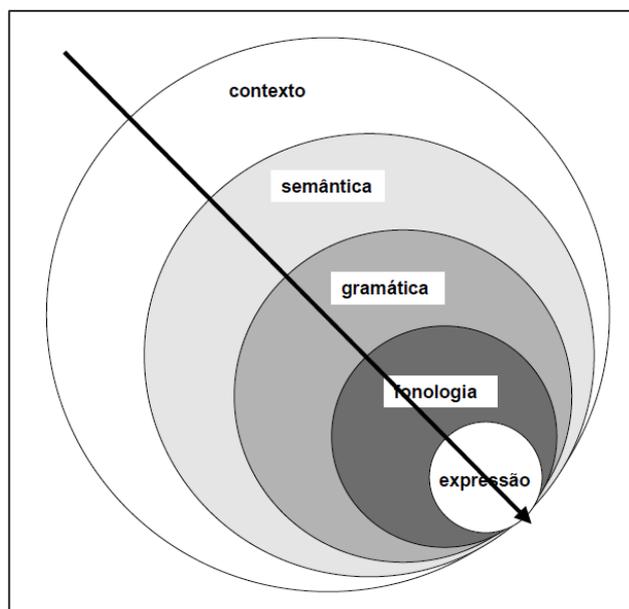


FIGURA 3. Estratificação e realização

Fonte: adaptada de Halliday e Matthiessen (2004, p. 25).

Nota-se, ainda pela FIG. 3, que a expressão realiza a gramática, que realiza a semântica, que, por sua vez, realiza o contexto. Halliday e Matthiessen (1999) afirmam que o conceito de realização refere-se à relação de diferentes tipos de fenômenos, por meio de camadas que estabelecem um conjunto de possibilidades, gerando potenciais de significado (semântica), de fraseado (léxico-gramática) e de expressão.

A língua é, portanto, considerada um sistema semiótico estratificado que engloba dois planos inter-relacionados, isto é: conteúdo e expressão. No primeiro, o estrato da semântica é

realizado pelo estrato da léxico-gramática; no segundo, o estrato da fonologia é realizado pelo estrato da fonética. Os estratos do plano da expressão realizam os do conteúdo.

De acordo com Halliday e Matthiessen (2004, p. 31), a estratificação da língua pode ser analisada sob uma perspectiva trinocular: ao redor, observando uma unidade léxico-gramatical no mesmo nível em que se encontra; de cima, ou seja, a partir do estrato semântico; e de baixo, mais especificamente a partir do nível lexical ou fonológico. A pesquisa em tela pretendeu estabelecer uma correlação entre essas três visões para a construção do significado dos *corpora* analisados. A visão de cima permitiu verificar a construção de significados, por meio da tipologia dos textos e da abordagem de *corpus*, uma vez que a visão de cima revela padrões do tipo de texto (e contexto). A visão ao redor possibilitou mapear o perfil metafuncional dos tipos de textos (explicado mais detalhadamente a seguir), incluindo também a abordagem de *corpus* como um exemplar da gramática mais delicada – “*lexis as most delicate grammar*” (HASAN, 1985). A visão de baixo (estrato da expressão), por meio da abordagem de *corpus*, proporcionou uma análise mais detalhada do nível léxico-gramatical, conforme ilustrado na FIG. 4.

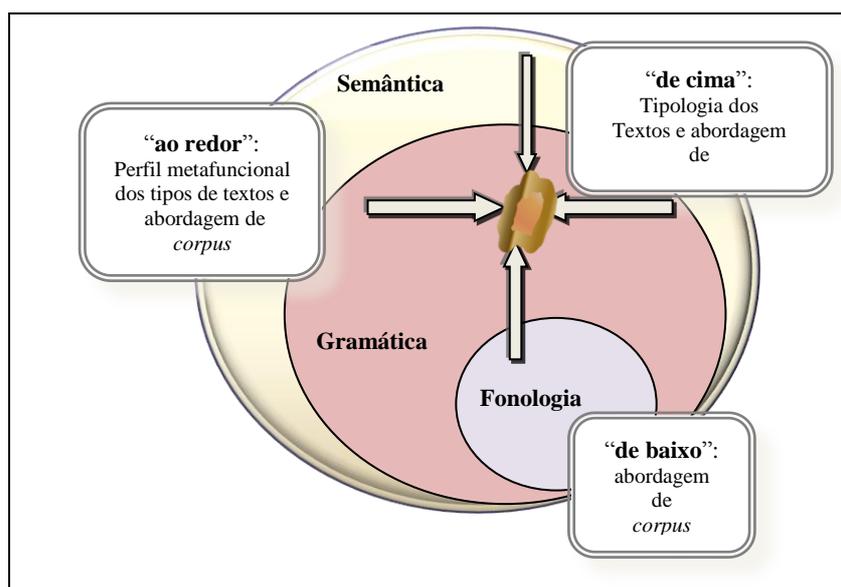


FIGURA 4. Perspectiva trinocular da pesquisa

Fonte: adaptada de Figueredo (2011, p. 77).

Considerando-se, portanto, o estrato semântico, cada uma das variáveis da situação (*i.e.*, campo, sintonia e modo) realiza tipos específicos de significado: ideacionais, interpessoais e textuais, os quais são realizados no nível léxico-gramatical pelos sistemas de TRANSITIVIDADE, MODO e TEMA. A metafunção ideacional é realizada na

TRANSITIVIDADE, que possibilita identificar as atividades sociais e os participantes sociais. A metafunção interpessoal, realizada no MODO, possibilita analisar como os significados sobre a troca de informações e a realização de bens de serviços é realizada entre os participantes. Já a metafunção textual é realizada no TEMA, que possibilita analisar a organização da língua no texto como um todo.

Em cada língua (inglês e português, no caso desta pesquisa), há uma forma de se organizar os sistemas de TRANSITIVIDADE, MODO, e TEMA e MENSAGEM, considerando-se o nível de *delicadeza* (cf. TEICH, 1999). Para se entender como as línguas se organizam dentro desses sistemas é essencial compreender a relação entre as categorias teóricas (metafunção, contexto, semântica, gramática, etc.) e as categorias descritivas (funções discursivas, mensagem, “figuration” etc.), conforme ilustrado na FIG. 5.

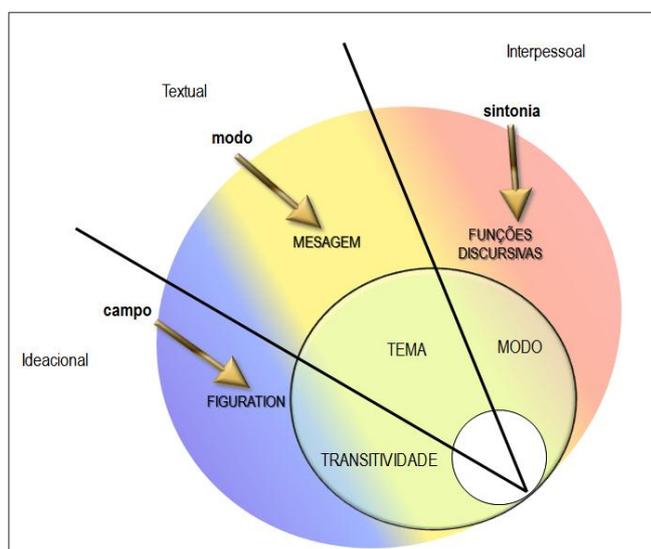


FIGURA 5. Realização entre os estratos semântico, fonológico e léxico-gramatical

Fonte: adaptada de Figueredo (2011, p. 79).

A FIG. 5 mostra a realização, que se refere às relações entre os estratos por meio da expressão (*i.e.*, estrato fonológico) e conteúdo (*i.e.*, estratos semântico e léxico-gramatical), mudando-se o nível de abstração do estrato menos abstrato para o mais abstrato: a léxicogramática realiza a semântica e a semântica é realizada pela gramática (organização física dos significados). Assim, a língua é realizada pelas três variáveis do contexto de situação: campo, sintonia e modo. O contexto é realizado pelas três metafunções (*i.e.*, ideacional, interpessoal e textual),

as quais, por sua vez, são realizadas pelos sistemas de TRANSITIVIDADE, MODO, e TEMA e MENSAGEM. Os sistemas são explicitados a seguir nas Seções 1.2.1 a 1.2.3.

### **1.2.1 O sistema de TRANSITIVIDADE**

O sistema de TRANSITIVIDADE configura a metafunção ideacional-experiencial da linguagem e representa o mundo de experiências por meio dos diferentes tipos de processos (HALLIDAY; MATTHIESSEN, 2004, p. 170). Para esta tese, o estudo desse sistema possibilitou analisar o fluxo de acontecimentos por meio da oração. É na oração onde ocorre a representação de um evento (processo) que se desencadeia com um ou mais participantes envolvidos e as circunstâncias associadas ao processo (“Figuration”, *cf.* HALLIDAY; MATTHIESSEN, 1999). Portanto, do ponto de vista do sistema de TRANSITIVIDADE, a oração apresenta uma estrutura formada por elementos distintos: processos, seus participantes e as circunstâncias nas quais os processos ocorrem.

Halliday e Matthiessen (1999) explicam que esse é um sistema de relação entre componentes que formam uma “figura” (*figure*). De acordo com esses autores, as figuras são configurações compostas de um processo (expressão de um evento) e participantes (quem realiza o quê) e, eventualmente, de circunstâncias associadas ao processo (quando, como, onde, por que etc.).

As figuras são identificadas de acordo com os processos, a saber: (i) figuras de fazer e acontecer; (ii) figuras de sentir; (iii) figuras de dizer; (iv) figuras de ser e ter; (v) figuras de existir; e, na língua inglesa, figuras de comportar-se. Em outras palavras, figuras são os significados realizados pelos processos, sendo eles são os responsáveis por representar eventos que constituem atividades humanas, por meio das experiências no mundo, e por ilustrar os aspectos do mundo das experiências, mental e social. Essa representação perpassa a ideia de mudança por ser realizada por processos, cujos participantes são as entidades envolvidas. Esses participantes podem ser afetados pela transitividade e, dependendo do tipo de processo, recebem diferentes denominações.

Se o processo significar uma figura de fazer ou acontecer (*i.e.*, processo material), tem-se, por exemplo, que o participante que o realiza é denominado Ator e o participante afetado pelo processo é denominado Meta. Se o processo significar uma figura de sentir (*i.e.*, processo mental, realizado em nível de cognição, afeição; enfim no nível do mundo interno, da

consciência), os participantes são denominados Experienciador – que participa conscientemente – e Fenômeno – o que está sendo percebido pelo Experienciador. Se o processo significa uma figura de dizer (*i.e.*, processo verbal), o participante central é denominado Dizente e vem acompanhado do participante denominado Receptor, que recebe o que é dito pelo Dizente. Se o processo significa uma figura de ser e ter (*i.e.*, processo relacional), têm-se participantes distintos dependendo da subdivisão do processo (atributivo ou identificativo). No processo atributivo, o Portador é o participante que tem as características ou descrições e o Atributo é a própria característica; no processo identificativo, o participante que tem sua identidade recebida é o Identificado e o Identificador é a identidade do Identificado. Se o processo significa uma figura de existir (*i.e.*, processo existencial), há apenas um participante, o Existente. Se o processo significa uma figura de comportar-se (*i.e.* processo comportamental), que se refere a comportamentos fisiológicos e psicológicos, o participante central desse tipo de processo é denominado Comportante, havendo ainda o participante Comportamento, que é responsável por estender o processo.

Em suma, processos, participantes e circunstâncias são considerados categorias semânticas que ilustram, de modo geral, a forma que os fenômenos do mundo das experiências são representados na estrutura linguística. A FIG. 6 mostra a representação gráfica do componente semântico relacionado à léxico-gramática e realizado pelo sistema de TRANSITIVIDADE (HALLIDAY; MATTHIESSEN, 1999, p. 55).

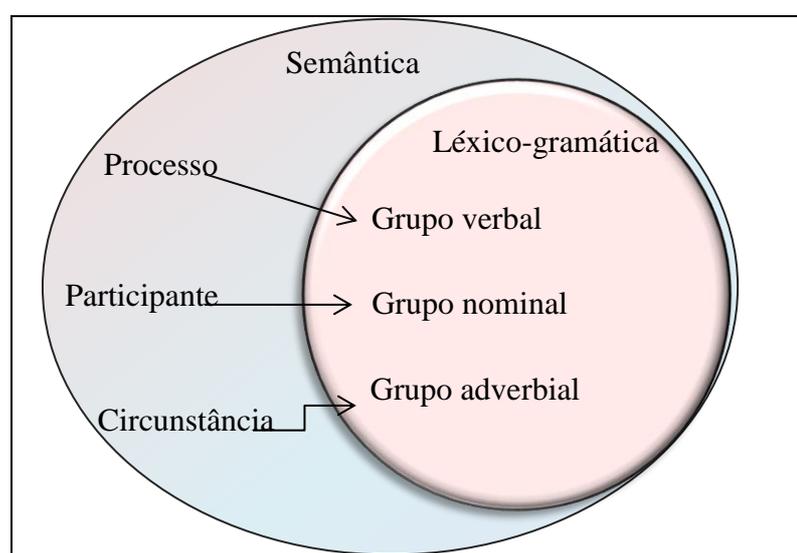


FIGURA 6. Representação gráfica do componente semântico relacionado à léxico-gramática e realizado pelo Sistema de TRANSITIVIDADE

Fonte: Halliday e Matthiessen (1999, p. 55).

De acordo com Halliday e Matthiessen (1999), o componente lógico da metafunção ideacional é instanciado no nível da léxico-gramática, mais especificamente no tipo de interdependência lógico-semântica entre as orações. De acordo com esses autores, as orações que realizam diferentes tipos de processos apresentam contribuições distintas para a construção da experiência nos textos. Halliday (2002) propõe a divisão dos processos em material, mental e relacional (subdivido em identificativo e atributivo), verbal e existencial, havendo, em inglês, também os processos comportamentais, que envolvem comportamento fisiológico e psicológico comportamental (*cf.* FIG. 7).

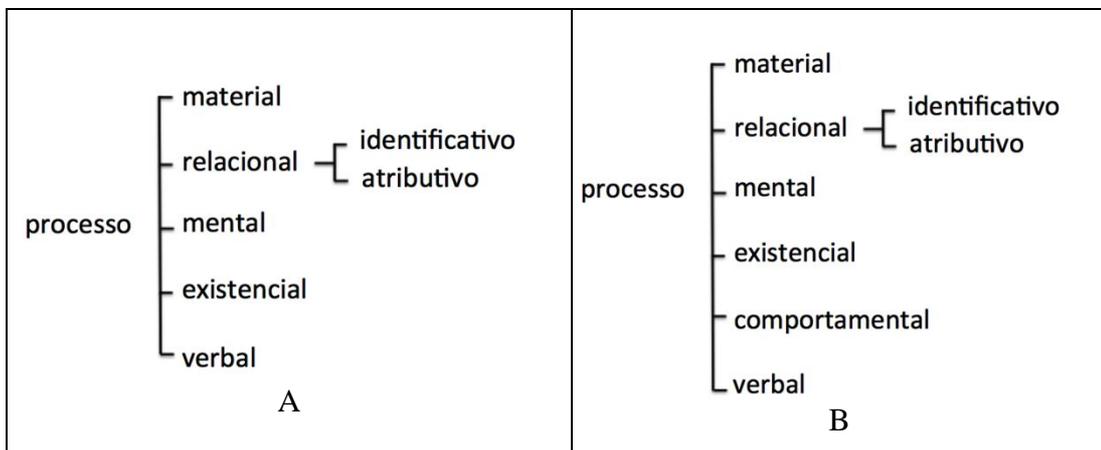


FIGURA 7. Sistemas de TRANSITIVIDADE das línguas portuguesa (A) e inglesa (B)

Fonte: Figueredo (2011) e Halliday e Matthiessen (2004).

O QUADRO 1 ilustra os tipos de processos que compõem o sistema de TRANSITIVIDADE e os exemplifica<sup>7</sup> com partes retiradas dos *corpora* desta pesquisa.

<sup>7</sup> Todos os exemplos desta tese foram extraídos dos *corpora*.

QUADRO 1. Tipos, definição e exemplos dos processos que compõem o sistema de TRANSITIVIDADE

Tipos de Processos	Definição	Denominação do Participante	Exemplos
Materiais	Expressam experiências de mudança no fluxo de eventos no mundo das experiências	Ator Meta	Em casos graves, <b>realiza-se</b> a exsanguíneo transfusão.
Mentais	Referem-se à experiência do mundo gerada a partir da consciência, com participante consciente	Experienciador Fenômeno	Há necessidade de se <b>acreditar</b> nos pacientes quando dizem estar sentindo dor. <i>Não encontrados em inglês</i>
Relacionais	Expressam a necessidade de se relacionar um participante a uma característica ou descrição.	Portador e Atributo	However, research studies <b>are</b> lacking in key areas for children with SCD. <i>Untreated complications <b>can cause</b> long term handicap.</i>
	Servem para caracterizar (processo relacional) e identificar (processo identificativo) os participantes	Identificado e Identificador	A primeira manifestação de dor na maioria das crianças <b>é</b> a dactilite (ou síndrome mão-pé). <i>The liver <b>clears</b> the bilirubin from the body.</i>
Verbais	Constroem o <i>dizer</i> , introduzindo outras orações projetadas em um complexo verbal	Dizente	Os pacientes mais velhos <b>citam</b> que a depressão e exaustão física podem ser fatores precipitantes das crises. <i>Não encontrados em inglês.</i>
Existenciais	Representam o que existe	Existente	<b>There were</b> 321 (10% of abstracts screened) articles selected for full review (Appendix 2), of which 26 were rated as level I evidence.”
Comportamentais (língua inglesa)	Representam formas de comportamento fisiológico e psicológico especificamente humano	Comportante Comportamento	Não encontrados.

Halliday e Matthiessen (2004) salientam que a escolha por uma ou outra estrutura linguística depende do contexto em que os textos são instanciados, principalmente tendo essas escolhas como fenômenos da experiência do mundo das experiências. Logo, com o propósito de verificar como algumas experiências da área da Saúde (em particular a medicina, subárea triagem neonatal para doença falciforme) são construídas no nível da léxico-gramática, realizaram-se os procedimentos para análise das orações de textos que instanciam registros utilizados nessa área. Na seção a seguir, são apresentadas as características do sistema de MODO.

### 1.2.2 O sistema de MODO

A metafunção interpessoal é realizada pelo sistema de MODO, responsável por mostrar as interações sociais localizadas no tempo e no espaço. Neste estudo, o MODO foi analisado a partir da constituição da oração, vista como troca na metafunção interpessoal, com o objetivo de identificar as negociações realizadas durante a escrita dos textos que compuseram os

*corpora*. Cumpre ressaltar que os sistemas de MODO são diferentes para o português brasileiro e para a língua inglesa, dependendo do nível de delicadeza da análise.

O sistema de MODO apresenta recursos que são utilizados para realizar a interação linguística. Nessa interação, o emissor exerce papéis que são mediados pela linguagem (da função semântica para a gramatical). Halliday e Matthiessen (2004) afirmam que, semanticamente, as funções interpessoais servem para (i) designar o que está sendo tratado entre o emissor e o receptor, ou seja, o papel exercido pelos participantes no processo e (ii) a maneira como os falantes transmitem suas mensagens. Esses dois elementos possibilitam identificar os movimentos dentro de um texto, por exemplo, em que os falantes utilizam a língua para interagir e organizar a sociedade. Esses movimentos resultam em categorias funcionais de linguagem da função discursiva (demandar, oferecer) e de mercadorias (bens e serviços e informações), tais como, declaração, questão, oferta e comando. Declaração e questão constituem ofertas, proposições e comandos, além de caracterizarem proposições, todas identificadas pelas opções gramaticais de MODO. As proposições são passíveis de respostas do ouvinte, como: afirmar, negar, duvidar, aceitar com reservas, qualificar, misturar ou retornar o que foi dito (HALLIDAY, 1994, p. 70). Já para ofertas e comandos, por resultarem em efeitos não linguísticos, não há recursos especiais oferecidos pela língua. A FIG. 8 mostra como o sistema geral do MODO foi utilizado para a realização desta pesquisa.

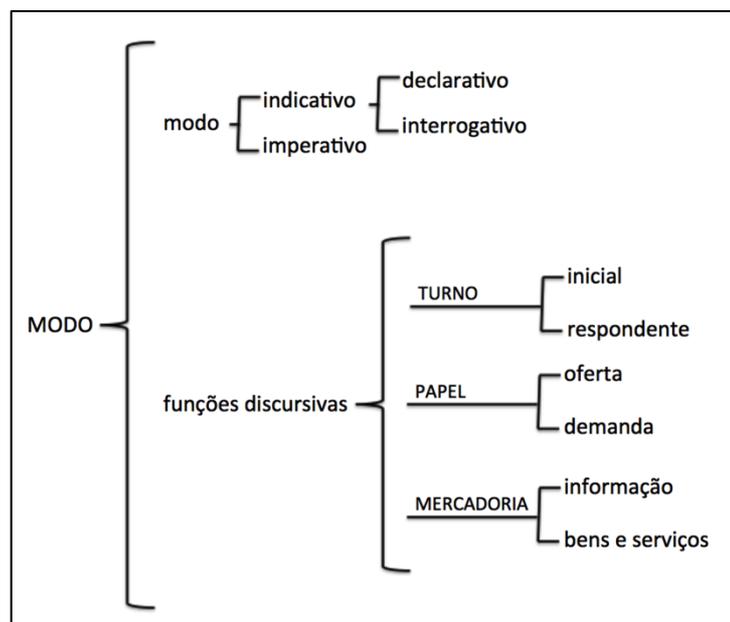


FIGURA 8. Sistema Geral do MODO e Funções discursivas

Fonte: traduzido de Halliday (2002).

Conforme ilustrado na FIG. 8, o sistema geral do MODO é composto pelos Modos Imperativo e Indicativo. O Indicativo se subdivide em declarativo e interrogativo, que, por sua vez, é subdividido em elemental - sim/não. O Modo Imperativo é utilizado quando há a necessidade de uma demanda de bens e serviços, enquanto o Modo Indicativo declarativo é usado quando há uma demanda de informação. Porém, quando há a necessidade de ofertar bens ou serviços, é possível escolher entre os Modos declarativo, interrogativo ou imperativo, pois, de acordo com Halliday e Matthiessen (2004), a oferta não dispõe de gramática própria.

Embora estejam em estratos diferentes, o Modo e as funções discursivas se relacionam por meio da realização. No que toca a um nível mais delicado, no estrato da semântica, observam-se as funções discursivas nos quais os tipos de Modo são realizados: turno, o papel e a mercadoria. O turno (interação entre os falantes) é subdividido entre inicial e respondente; o papel está relacionado com a oferta e a demanda; e a mercadoria refere-se à informação e bens e serviço.

Para uma leitura detalhada sobre o sistema de MODO da língua portuguesa, remete-se o leitor a Figueredo (2011, p. 240, onde está descrita a rede geral do sistema de MODO em português). Os procedimentos de análise desses elementos permitiram descrever o sistema de MODO dos *corpora* desta pesquisa, conforme ilustrado no QUADRO 2.

QUADRO 2. Exemplos de MODO dos tipos de texto

Tipo de Texto	Língua	Exemplo	Função Semântica	Modo	
Artigos acadêmicos	Português	Avaliação e medição criteriosas da dor falciforme – e de qualquer dor – são passos fundamentais para a programação das ações.	Oferta de informação	Indicativo Declarativo	
	Inglês	<i>This article presents a comprehensive review of the literature for the care provided to children with SCD.</i>			
Manuais técnicos	Português	Qualquer tipo de pneumonia exige hospitalização para antibioticoterapia venosa.			
	Inglês	<i>The majority of mild to moderate pain episodes abate spontaneously with bed rest, intake of extra fluids and analgesia and patients manage these at home using pain medications.</i>			
Cartilhas e folders de divulgação	Português	A doença falciforme é uma alteração genética, caracterizada por um tipo de hemoglobina.		Demanda de bens e serviços	Imperativo
	Inglês	<i>When no haemoglobin A is produced the condition is known as sickle beta zero thalassaemia, sometimes written as SBETA Thal.</i>			
		<i>For more information about beta thalassaemia contact the Thalassaemia Society or one of the specialist centres.</i>			

Logo, a análise do sistema de MODO permite identificar as funções semântica e gramatical das orações de cada um dos textos, além de possibilitar uma comparação interlíngua e inter-registro por meio da metafunção interpessoal. Os Sistemas de TEMA e MENSAGEM são abordado na próxima seção.

### 1.2.3 O sistema de TEMA e MENSAGEM

A metafunção textual é realizada no estrato léxico-gramatical pelo sistema de TEMA (estrato da gramática), sendo a responsável pela organização do texto a partir dos significados ideacionais e interpessoais da linguagem. Halliday (2002) descreve a metafunção textual da linguagem como aquela que organiza a mensagem. Halliday e Hasan (1976) apontam que essa organização acontece por meio de recursos da tessitura estrutural (responsável pela

organização do texto) e da tessitura não estrutural ou coesão (que estabelece a relação entre os significados do texto).

Conforme afirmam Halliday e Matthiessen (1999), o texto como um todo pode ser considerado uma unidade de análise da relação direta entre a base ideacional e a base textual, pois o nível semântico do discurso é o nível mais alto da organização do texto. Esse nível está relacionado à coesão do texto (discurso) e o significado do texto (aspecto semântico). Consequentemente, o texto é organizado por partes e as relações semânticas entre essas partes, consideradas uma unidade de significado em um contexto de situação.

No que toca à gramática, essa organização textual é construída pelo sistema de TEMA. Responsável por realizar o sistema de MENSAGEM, o Tema pode ser interpretado de duas formas que se complementam: (i) em relação à oração, considerado a crista de proeminência, o ponto de partida; e (ii) em relação ao discurso, por estabelecer a base para o fluxo de informação que está por vir, possibilitando a interpretação da oração de acordo com a argumentação do autor (HALLIDAY; MATTHIESSEN, 1999, p. 12).

Halliday (1994, p. 37) aponta que a estrutura temática concede à oração característica de mensagem, uma vez que a oração representa uma forma de organização, a qual permite que a oração se encaixe no discurso e contribua para o fluxo discursivo. Assim, o TEMA é responsável por localizar e orientar a oração em seu contexto, conforme afirmam Halliday e Matthiessen (2004).

A FIG. 9 ilustra o sistema de TEMA em português brasileiro.

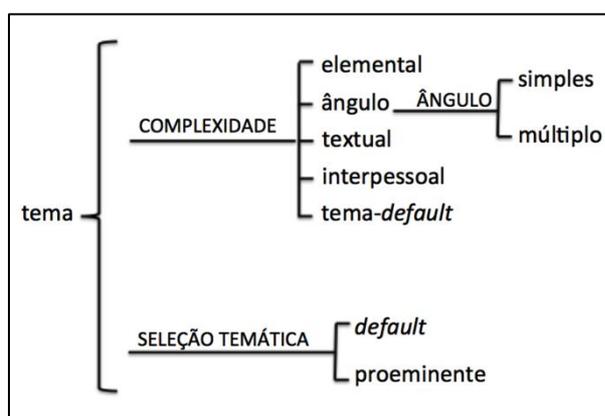


FIGURA 9. Sistema de TEMA em português brasileiro

Fonte: Figueredo (2011).

O sistema de TEMA em português brasileiro, conforme ilustrado na FIG. 9, é dividido em: (i) complexidade, que se subdivide em tema elemental, ângulo (fonte e ponto de vista), textual, interpessoal e tema-*default*; e (ii) seleção temática (*default* e proeminente). A seleção temática engloba duas opções de orientação da oração relativa à variável do contexto *campo*, sendo possível analisar se a oração causa ou não alterações em uma mesma fase do discurso, além de verificar se há alguma forma de descontinuidade ou mudança. A função da seleção *default* é manter a continuidade da orientação também voltada para o *campo*. A seleção proeminente, por sua vez, tem a função de oferecer uma descontinuidade na fase ao apresentar uma subfase, um ângulo ou até mudança de uma fase para outra (FIGUEREDO, 2011).

O QUADRO 3 ilustra os tipos e definições dos temas que pertencem à complexidade no sistema de TEMA em português brasileiro e mostra exemplos retirados dos *corpora* desta pesquisa.

QUADRO 3. Definição e exemplo de TEMA em português brasileiro dos tipos de texto

Tipos de TEMA		Definição	Exemplos
COMPLEXIDADE	Textual	Responsável por ligar a oração ao texto por intermédio dos recursos da tessitura (elos coesivos). Normalmente apresenta uma conjunção no início da oração.	<u>Porém</u> o tratamento pode ser ambulatorial com reavaliação em 48 horas.
	Interpessoal	Responsável por indicar o tipo de relação entre autor/leitor e falante/ouvinte. Confere proeminência temática à avaliação do falante em relação à oração.	<u>Como se sabe</u> , a HbS teve origem no continente africano.
	Elemental	“Compreende os elementos ‘qu-’ que não confluem com o Sujeito. Por isso, esse tipo de Tema se restringe ao ambiente interrogativo: elemental.” (FIGUEREDO, 2011, p. 148)	Não encontrados.
	Ângulo	“Compreende os itens que indicam projeção de pensamento ou fala dentro do ambiente oracional. Estes são realizados por frases preposicionais ou grupos adverbiais.” (FIGUEREDO, 2011, p. 150)	<u>Para o autor</u> os métodos usados para a avaliação devem ser os mesmos utilizados para qualquer síndrome dolorosa.
SELEÇÃO TEMÁTICA	<i>Default</i>	“É a confluência do elemento crítico interpessoal, a partícula de origem no <i>quantum</i> de mudança e a proeminência que mantém a continuidade em uma fase do texto, conservando a orientação para o campo.” (FIGUEREDO, 2011, p. 144)	<u>Os demais pacientes</u> (não graves) devem ser abordados com a mesma propedêutica.
	Proeminente	“Compreende as opções que provocam mudança na interpretação da mensagem; quer para o desenvolvimento do texto (descontinuidade), para o rema da oração, ou ainda, em conjunto com a informação, para a informação com <i>status</i> de nova.” (FIGUEREDO, 2011, p. 151)	<u>Segundo o autor</u> como quase todos os episódios são manejados em casa, a avaliação da dor e das medidas de convivência com ela precisa focalizar seu manejo satisfatório em ambiente domiciliar.

O sistema de TEMA em inglês também é dividido em duas partes: (i) complexidade, que se subdivide em simples e múltiplo (textual e tópico, interpessoal e tópico e textual-interpessoal e tópico); e (ii) seleção temática (não marcado e marcado), conforme ilustrado na FIG. 10.

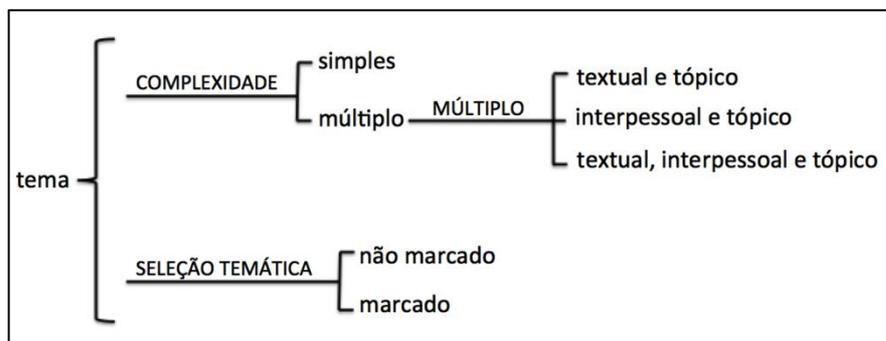


FIGURA 10. Sistema de TEMA em inglês

Fonte: Halliday (2002).

No sistema de Tema em inglês, a complexidade é composta pelos temas simples e múltiplo (subdivido em textual e tópico, interpessoal e tópico e textual, interpessoal e tópico). A seleção temática é composta pelo tema não marcado, localizado na posição esperada na oração, por manifestar a seleção de escolhas em relação à estrutura da língua, ou seja, sujeito/verbo/objeto (SVO), correspondendo ao sujeito em uma oração declarativa. Em uma oração imperativa, o tema marcado representa o processo e, em uma interrogativa, representa uma interrogação com o elemento WH.

O QUADRO 4 ilustra os tipos de temas que constituem a complexidade no sistema TEMA em inglês e mostra exemplos retirados dos *corpora* desta pesquisa.

QUADRO 4. Definição e exemplo de TEMA em inglês dos tipos de texto

Tipos de TEMA		Definição	Exemplos
COMPLEXIDADE	Textual	Responsável por ligar a oração ao texto por intermédio dos recursos da tessitura (elos coesivos). Normalmente apresenta uma conjunção no início da oração. Este tema vem antes do tema tópico para dar proeminência textual aos elementos de ligação.	<i>Although these may precipitate a sickling crisis may occur with no known specific cause or identifiable precipitating factor.</i>
	Tópico (ideacional)	O primeiro elemento ideacional da oração, podendo ser um participante, processo ou circunstância.	<i>Untreated complications can cause long term handicap and can prove fatal.</i>
	Interpessoal	Responsável por indicar o tipo de relação entre autor/leitor e falante/ouvinte (vocativo, comentário, adjuntos e -WH no início de pergunta elemental).	<i>Unfortunately those who inherited two of the sickle cell genes and have sickle cell anaemia were not protected from malaria.</i>
SELEÇÃO TEMÁTICA	Não Marcado	Corresponde à escolha ou seleção mais usual em uma oração no sistema de MODO (declarativa, imperativa ou interrogativa).	<i>The majority of mild to moderate pain episodes abate spontaneously with bed rest.</i>
	Marcado	Corresponde à escolha ou seleção menos usual em uma oração no sistema de MODO (declarativa, imperativa ou interrogativa), o que torna essa seleção proeminente.	<i>If the research methods were unclear in the abstract, the article was reviewed in detail.</i>

A Seção 1.4 traz elucidacões sobre o conceito de perfil metafuncional dos tipos de textos. Até aqui, foram explicados e ilustrados os sistemas de TRANSITIVIDADE, MODO e TEMA, em relação aos quais foi realizada uma descrição sistêmico-funcional (Seção 1.3) com o intuito de mapear o perfil metafuncional de cada um dos três tipos de textos constituintes dos *corpora* compilados com base na tipologia da língua no contexto de cultura (MATTHIESSEN; TERUYA; WU, 2008). Essa descrição possibilita a caracterização dos textos em diferentes registros no ambiente multilíngue.

No estrato semântico, portanto, a MENSAGEM possibilita compreender como os significados são realizados no texto. Neste estudo, a análise desse sistema justifica-se pela possibilidade de descrever os perfis textuais dos textos que compuseram os *corpora* desta pesquisa, por meio de alguns elementos da MENSAGEM, conforme ilustrado na FIG. 11. A partir da análise desses elementos foi possível a descrição dos perfis textuais dos textos desta pesquisa (cf. Seção 3.3).

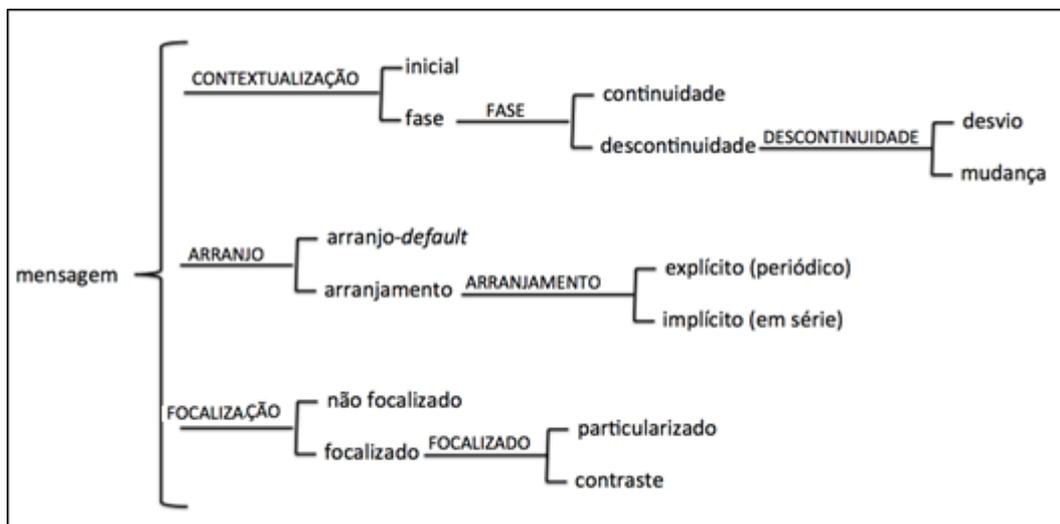


FIGURA 11. Sistema de MENSAGEM

Fonte: Figueredo (2011).

A mensagem é, por conseguinte, a unidade semântica básica do texto, sendo sequenciada pelo fluxo de informação e realizada por uma unidade gramatical: a oração (na maioria das vezes). Uma vez produzida, a mensagem pode dar início ao processo de CONTEXTUALIZAÇÃO (cf. FIG. 11). Esse processo consiste na forma como a organização da mensagem orienta o fluxo da informação até atingir a semiotização dos significados a partir do contexto de situação, abrangendo, assim, desde a estrutura e o tipo de texto até a macrossignificação. Uma vez produzida, a mensagem pode ser a parte inicial da contextualização ou a fase da contextualização anterior. Se a mensagem expande o significado de uma anterior, desempenha a função de continuidade, podendo ser impactada pelo campo ou sintonia. Caso contrário, pode ocorrer a descontinuidade por um desvio em um aspecto específico da mensagem ou por uma mudança no padrão da semiotização dos significados (MARTIN; ROSE, 2007).

De acordo com Martin e Rose (2007), a mensagem organizada em relação à fase (sistema de CONTEXTUALIZAÇÃO) é impactada pelas variáveis do discurso campo e sintonia. Logo, continuidade, desvio e mudança englobam os significados ideacionais e interpessoais que são conservados ao longo do texto e realizam um macrossignificado que trata do mesmo assunto. Entretanto, a fase do discurso sofre alteração quando a organização da mensagem orienta o assunto para o campo (texto mais especializado) ou para a sintonia (texto menos especializado). Daí vem a importância da análise da mensagem neste estudo, cujos textos analisados estão orientados para a sintonia.

Os significados textuais na contextualização estão relacionados diretamente com o sistema de FOCALIZAÇÃO (cf. FIG. 11), que garante proeminência textual para uma mensagem inserida na fase. A focalização pode ser um tipo particularizado da mensagem ou um contraste entre uma mensagem e demais mensagens do texto.

O sistema ARRANJO apresenta a forma como o fluxo discursivo engloba algumas porções de mensagens que guardam semelhanças maiores entre si quando comparadas às demais, culminando nas fases do discurso. As funções desse sistema são divididas em ARRANJAMENTO – subdividido em explícito (periódico) e implícito (em série) – e arranjo default.

O arranjo é responsável por determinar o modo de desenvolvimento do fluxo do discurso, iniciando a onda<sup>8</sup> de informação, ao passo que o arranjo *default* segue esse fluxo, concretizando a onda de informação. Quando o desenvolvimento do fluxo do discurso é limitado ao concatenamento de mensagens e é necessária inferência pelo receptor, diz-se que o arranjo foi realizado em série, implícito (a relação entre a mensagem e o fluxo não é explícita). Quando as ondas formadas pelas mensagens são passíveis de serem identificadas e é possível entender como elas se relacionam, o fluxo é periódico, explícito (não demandando inferência do receptor), sendo hierarquizado por unidades menores.

Em suma, semanticamente, para se tornar parte do texto, a mensagem é produzida pelo sistema de MENSAGEM e, na sequência, ocorre uma seleção de uma função dos sistemas complementares de CONTEXTUALIZAÇÃO, FOCALIZAÇÃO e ARRANJO. Assim, esse modo de organização textual contribui para a construção do texto como fluxo discursivo.

### 1.3 Perfil Metafuncional dos Tipos de Texto

Halliday e Matthiessen (1999) apontam os sistemas MODO, TRANSITIVIDADE, TEMA e MENSAGEM como os sistemas que caracterizam o estrato lexicogramatical. O perfil metafuncional dos tipos de texto é caracterizado pelas escolhas experienciais, interpessoais e textuais identificados em cada um desses sistemas. A partir dessas escolhas é possível

---

<sup>8</sup> O termo metafórico “onda” é usado por Halliday (2002) para descrever o fluxo discursivo do texto, ou seja, a forma de organizar a oração com ênfase no destaque textual para determinados elementos. Esses elementos são denominados de cristas de proeminência. Aqueles elementos que não estão classificados dessa forma ocupam os espaços de não proeminência.

descrever a forma pela qual diferentes tipos de textos são instanciados (por exemplo, por meio da frequência de suas categorias gramaticais).

As escolhas não são aleatórias e seguem uma ordem dentro do estrato semântico. Considerando-se a ordem da oração, por exemplo, essas escolhas referem-se ao sistema de TRANSITIVIDADE (*cf.* Seção 1.2.1) e são realizadas a partir dos processos, listados a seguir, que ilustram o mundo de experiências (HALLIDAY; MATTHIESSEN, 2004):

1. Processos materiais (ilustram as experiências do mundo exterior);
2. Processos mentais (ilustram as experiências do mundo interior);
3. Processos verbais (constroem o *dizer*, ou seja, aquilo que é dito pelo participante em forma de pergunta, declarações etc.);
4. Processos relacionais atributivos (relacionam o participante com suas características gerais ou descrevem esse participante);
5. Processos relacionais identificativos (atribuem identidade ou especificam o papel do participante);
6. processos comportamentais (ilustram o comportamento fisiológico e psicológico); e
7. processos existenciais (apresentam novos participantes).

No sistema de MODO (*cf.* Subseção 1.2.2), as escolhas são realizadas seguindo, obrigatoriamente, o Modo Indicativo (Declarativo ou Interrogativo) ou Imperativo (HALLIDAY, 2002). No sistema de TEMA (*cf.* Subseção 1.2.3), as escolhas possibilitam que a língua se torne um texto, ou seja, uma instanciação do sistema que consiste em uma sequência de escolhas que pode ser estudada de forma a contemplar e organizar as funções selecionadas pelo autor do texto dentro do contexto de cultura (HALLIDAY, 2002).

Figueredo (2011) identificou e propôs uma descrição sistêmico-funcional dos sistemas de TEMA, MODO e TRANSITIVIDADE (processos mentais) do português brasileiro. Essa identificação contribuiu para a interface entre os Estudos da Tradução e os estudos linguísticos de base sistêmico-funcional de orientação tipológica, permitindo, conseqüentemente, compreender a produção linguística no ambiente multilíngue (*cf.* Seção 1.4, a seguir). Seguindo a proposta de Figueredo (2011), esta tese pretendeu descrever o perfil metafuncional dos três diferentes tipos de texto em português e em inglês, produzidos no ambiente multilíngue, de acordo com as três metafunções (*i.e.*, ideacional, interpessoal e

textual). Mais especificamente, a metafunção ideacional (campo) representa as experiências no mundo das experiências, considerando-se a produção de textos da área da saúde (medicina, subárea triagem neonatal); a metafunção interpessoal (sintonia) aponta a relação do autor com o leitor, além de explicitar a forma como os valores da ciência são construídos no texto; e a metafunção textual explicita a forma como os significados interpessoais e ideacionais se distribuem ao longo do texto.

A descrição do perfil metafuncional de cada um dos tipos de textos pode:

1. Informar a forma como os textos dos dois sistemas linguísticos apresentam as variáveis no contexto de cultura (MARTIN, 1992);
2. Subsidiar uma proposta de modelo de sistematização das escolhas metafuncionais dos textos no ambiente multilíngue;
3. Juntamente com a caracterização desses textos em diferentes registros, viabilizar a análise das escolhas metafuncionais de textos do tipo EXPLORAR e HABILITAR (MATTHIESSEN; TERUYA; WU, 2008) em português e inglês; e
4. Subsidiar sugestões de possíveis modificações em textos do tipo HABILITAR com base em pesquisas anteriores, como, por exemplo a de Coulthard (1992).

No artigo “On analyzing and evaluating written text”, Coulthard (1992) analisou alguns excertos de uma cartilha em inglês, *Holidays and travel for diabetics*, produzida pela Associação Britânica do Diabético em 1977 com o objetivo de trazer informações sobre a doença a pacientes diabéticos. Em sua análise, ele explica que as escolhas ideacionais e interpessoais são realizadas ao longo do texto de acordo com o conhecimento de mundo do autor, contudo tais escolhas são de difícil acesso ao leitor, o que dificulta o entendimento do texto. Coulthard (1992) considera que um texto possa ser reescrito (textualizado) infinitas vezes a partir de outro, considerando-se o conteúdo ideacional do texto como um todo e o público-alvo. Nesse sentido, ele propõe a definição de *textualização* da mensagem de um autor, a partir dos conceitos funcionais da língua (HALLIDAY, 1985), na tentativa de evitar o problema da construção do significado em uma linguagem mais acessível. O autor sugere que a Linguística pode auxiliar no processo de produção de texto, facilitando a escolha dos significados de forma a tornar o texto menos denso e com uma linguagem acessível ao público leigo. Em seu estudo, Coulthard (1992) propôs uma análise do bloco ideacional por ser o aspecto que ilustra o sentido geral do texto, a fim de auxiliar autores a produzir textos

direcionados a um público específico. Esse autor sugere uma nova proposta de redação para o excerto da cartilha analisado por ele, no intuito de tornar o texto mais acessível ao público leigo. O mesmo problema apontado por Coulthard (1992) foi observado durante a análise do perfil metafuncional dos textos compilados no corpus desta tese com base no rótulo cartilha e *folder* de divulgação. Observou-se que a linguagem utilizada nesse tipo de texto, supostamente direcionado a um público leigo, guardava características comuns com a linguagem usada no manual técnico, texto direcionado ao técnico da área da Saúde, o que dificulta a caracterização dos textos do tipo HABILITAR, tanto em inglês quanto em português. Textos rotulados como cartilhas, produzidos pelo Ministério da Saúde<sup>9</sup>, por exemplo, apesar de estarem direcionados a um público leigo, apresentam uma linguagem que pode dificultar a interpretação das informações. Nesse sentido, uma análise linguística poderia ajudar o autor a escrever um texto de popularização que realmente seja acessível ao público leigo. Halliday (1978) explica que o processo de se produzir um texto de forma consciente é mais complexo que aquele de produzir um texto de forma inconsciente. No primeiro caso, o autor precisa avaliar cada escolha em cada um dos sistemas, ao passo que, para produzir texto de forma inconsciente, o autor já sabe quais são as opções a serem selecionadas em cada um dos sistemas e a probabilidade de ocorrência dessas opções.

Após a análise do perfil metafuncional dos textos do tipo HABILITAR nesta pesquisa, percebeu-se que as escolhas ideacionais e interpessoais presentes nos textos rotulados como cartilha e *folder* de divulgação podem ser de difícil acesso ao público leigo, por apresentarem uma linguagem mais densa e menos acessível quando comparada aos textos caracterizados como textos de popularização da ciência (*cf.* BRAGA, 2013). Além disso, os dados apontaram que os textos rotulados como cartilha e folder de divulgação e manual técnico, apresentam padrões semelhantes, o que torna menos produtiva sua utilização para a elaboração de um *template*. Assim, com base no perfil metafuncional explicitado na pesquisa de Braga (2013) foi possível criar um *template* (modelo) para a elaboração parcial de cartilhas, baseado no perfil de um texto de popularização da ciência. A escolha de um texto do que geralmente se denomina "popularização da ciência" deve-se ao fato de ser este um texto cuja sintonia é análoga àquela da cartilha e folder de divulgação específica, embora se trate e um texto que pode ser associado ao processo sociosemiótico EXPLICAR, com alguns traços característicos do processo sociosemiótico HABILITAR. O domínio experiencial da variável

---

<sup>9</sup> Exemplos de cartilhas distribuídas à população pelo Ministério da saúde estão disponíveis em meio eletrônico, como: <<http://www.saude.salvador.ba.gov.br/arquivos/coaps/falciforme/impresso.pdf>> e <[http://portal.saude.gov.br/portal/arquivos/pdf/dir\\_tea.pdf](http://portal.saude.gov.br/portal/arquivos/pdf/dir_tea.pdf)>.

campo do texto escolhido também difere daquele da cartilha pretendida: medicina - triagem neonatal para anemia falciforme. Todavia, não são esses fatores que impedem a elaboração do modelo ensejado. Nesta tese, partiu-se do pressuposto de que, dada a inexistência de um texto que atendesse a especificidade do registro cartilha sobre anemia falciforme, um tipo textual novo no contexto de cultura precisava ser modelado, para fins de produção e tradução. A inexistência de um *corpus* paralelo de textos em inglês e suas respectivas traduções para o português brasileiro também foi um fator interveniente, uma vez que não havia um corpus do qual se possam extrair dados passíveis de serem comparados. Cumpre ressaltar que Braga (2013), com base na pesquisa de Figueredo (2011), faz uma proposta de aplicação do perfil metafuncional de textos de popularização da ciência. A autora explicita o perfil metafuncional dos textos analisados em sua pesquisa, a partir de sistematização das escolhas interpessoais, experienciais e textuais. A análise do perfil metafuncional do texto de popularização da ciência “Como funciona a pipoca” em português a partir dos sistemas de MODO, TRANSITIVIDADE e MENSAGEM gerou o perfil metafuncional desse texto (BRAGA, 2013), transcrito na íntegra na FIG. 12.

<b>PERFIL</b>
<b>Dúvida:</b>
<i>Indicativo: interrogativo: elemental / MATERIAL / Inicial</i>
<b>Características gerais da pipoca:</b>
<i>Indicativo: declarativo / RELACIONAL IDENTIFICATIVO / Continuidade</i>
<i>Indicativo: declarativo / MATERIAL / Inicial</i>
<i>Indicativo: declarativo / RELACIONAL ATRIBUTIVO / Descontinuidade: mudança</i>
<i>Indicativo: declarativo / MATERIAL / Inicial</i>
<i>Indicativo: declarativo / MATERIAL / Continuidade</i>
<i>Indicativo: declarativo / EXISTENCIAL / Continuidade</i>
<b>Explicação sobre o funcionamento da pipoca:</b>
<i>Indicativo: declarativo / MATERIAL / Inicial</i>
<i>Indicativo: declarativo / MATERIAL / Descontinuidade: mudança</i>
<i>Indicativo: declarativo / RELACIONAL ATRIBUTIVO / Continuidade</i>
<i>Indicativo: declarativo / MATERIAL / Descontinuidade: desvio</i>
<i>Indicativo: declarativo / MATERIAL / Continuidade</i>
<i>Indicativo: declarativo / MATERIAL / Continuidade</i>
<i>Indicativo: declarativo / RELACIONAL ATRIBUTIVO / Inicial</i>
<i>Indicativo: declarativo / MATERIAL / Continuidade</i>
<i>Indicativo: declarativo / RELACIONAL IDENTIFICATIVO / Continuidade</i>
<i>Indicativo: declarativo / MATERIAL / Continuidade</i>
<i>Indicativo: declarativo / MATERIAL / Descontinuidade: mudança</i>
<i>Indicativo: declarativo / MATERIAL / Continuidade</i>
<i>Indicativo: declarativo / MATERIAL / Continuidade</i>
<b>Experimento:</b>
<i>Imperativo / MENTAL / Inicial</i>
<i>Imperativo / MATERIAL / Continuidade</i>
<i>Indicativo: declarativo / MATERIAL / Continuidade</i>
<i>Imperativo / MATERIAL / Continuidade</i>
<i>Indicativo: declarativo / EXISTENCIAL / Continuidade</i>
<i>Imperativo / MATERIAL / Continuidade</i>
<i>Indicativo: declarativo / MATERIAL / Continuidade</i>

FIGURA 12. Perfil metafuncional do texto de popularização da ciência  
Fonte: Braga (2013).

Assim, com base no perfil metafuncional explicitado na pesquisa de Braga (2013) foi possível criar um *template* (modelo) para a elaboração de cartilha, tendo como base no perfil de um texto de popularização da ciência.

A seguir alguns pressupostos básicos da Linguística Sistêmico-Funcional, sobretudo no que toca aos estudos multilíngues (Seção 1.4), aos estudos de registro e interface com os estudos de *corpora*, que apontada a forma como a LSF complementa a Linguística de Corpus (Seção 1.5) e a tipologia de textos baseada nos parâmetros do contexto (Seção 1.6).

#### 1.4 Estudos Multilíngues

A proposta do campo disciplinar Estudos Multilíngues, elaborada por Matthiessen, Teruya e Wu (2008) e propiciada pela teoria Sistêmico-Funcional, tem por objetivo articular diferentes campos disciplinares e subáreas considerando-se o enfoque e o interesse de cada um deles através do estudo de contato linguístico. Esses autores apresentam o novo campo disciplinar como agregador de pesquisas dos Estudos da Tradução, do Ensino ou Aprendizagem de Línguas, da Tipologia Linguística, da Linguística Descritiva, da Linguística Contrastiva, nos Estudos de Multilinguismo, dentre outros.

No escopo do campo disciplinar delineado por Matthiessen, Teruya e Wu (2008), Figueredo (2011, 2012<sup>10</sup>, 2013<sup>11</sup>) propõe o estudo do ambiente multilíngue entendido como espaço no qual pode ser mapeado “o conjunto de configurações distintas de produção de significados cuja instanciação demanda a participação de um número variável de sistemas linguísticos” (FIGUEREDO, 2013).

De acordo com Figueredo (2013), pesquisas como as de Teich (1999) e Matthiessen *et al.* (2008) revelam que o estudo do ambiente multilíngue é potencializado quando realizado a partir de dois pontos que se complementam, a saber:

---

<sup>10</sup> FIGUEREDO, G. P. Modelagem da equivalência em tradução: parâmetros para a descrição do ambiente multilíngue. In: SEMANA DAS LETRAS. Pluridade da Memória, Departamento de Letras, Universidade Federal de Ouro Preto, 23 a 26 de outubro de 2012. Comunicação Verbal.

<sup>11</sup> Comunicação a ser apresentada em: FIGUEREDO, G. P. Developing tools to model meaning in multilingual environments. In: SIMPÓSIO INTERNACIONAL DE LINGUÍSTICA FUNCIONAL, 2, UFSCar, São Carlos, 14-16 de agosto de 2013. Comunicação Pessoal.

1. Por meio da análise de textos produzidos em diferentes línguas em contato;
2. Através da modelagem do ambiente multilíngue, considerando-se os sistemas em contato (dentro das línguas) e a descrição, incluindo as probabilidades dos recursos linguísticos, além da história sistêmica dos textos.

Nessa perspectiva, a pesquisa em tela está inserida no escopo daquelas que visam contribuir para a elaboração de um modelo que torne possível a produção e tradução de textos no ambiente multilíngue, onde as configurações distintas de produção de significados no contexto de cultura podem ser comparadas em diferentes línguas, em particular nos contatos entre a língua portuguesa e a língua inglesa, no caso deste estudo. Utilizou-se a abordagem metodológica da Linguística de Corpus para a compilação, armazenamento e extração dos dados analisados e, como base teórica, foi utilizada a Linguística Sistêmico-Funcional.

Sabe-se que a LSF modela a língua como um sistema semiótico imbricado no contexto de cultura, que por sua vez também é modelado como sistema semiótico. Ambos são mapeados e correlacionados ao longo do contínuo da instanciação (*cline of instantiation*). Levando-se em consideração, portanto, a organização da língua nesse contínuo, observa-se que o texto está localizado em um dos polos, conforme ilustrado na FIG. 13. O conceito de multilíngue refere-se a uma noção de se movimentar um determinado texto na direção do polo do potencial no contínuo da instanciação do sistema linguístico de acordo com a disposição topológica das categorias nos sistemas de TRANSITIVIDADE, MODO, e TEMA e MENSAGEM.

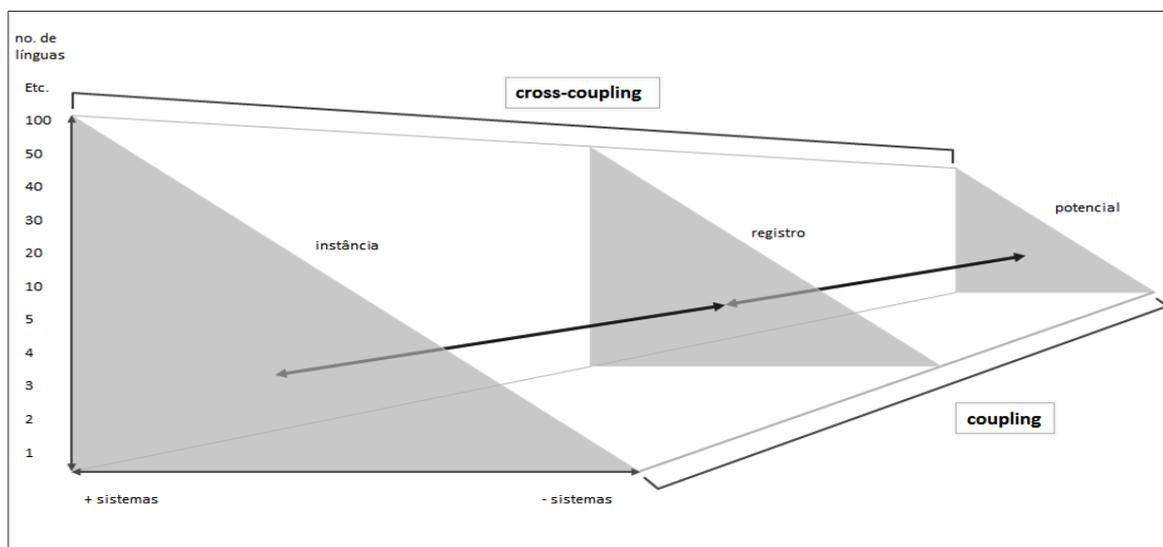


FIGURA 13. Ambiente multilíngue no contínuo da instanciação do sistema linguístico

Fonte: Figueredo (2012).<sup>12</sup>

A FIG. 13 ilustra o modelo do ambiente multilíngue em termos de número de línguas, contato linguístico e a realização de *coupling* e *cross-coupling* ao longo do contínuo da instanciação. Martin (2008, p. 39) entende *coupling* enquanto a maneira como ocorre combinações de significados através de sistemas. De acordo com este autor, o ponto principal da realização de *coupling* é entender que há significados interativos que configuram os registros no contínuo da instanciação. Pode-se dizer que os diferentes sistemas da língua fazem o *coupling* baseados na variação do registro (MATTHIESSEN, 1993) nessa mesma língua. Já o *cross-coupling* pode ser entendido como a combinação do *coupling* de duas línguas ou mais línguas diferentes. Nesse sentido, não há apenas *coupling* dos diferentes sistemas na mesma língua, mas sim a combinação do *coupling* de uma determinada língua com o *coupling* de outra língua.

No modelo do ambiente multilíngue, observa-se que o registro é um ponto intermediário entre o polo do potencial e a instância, passível de análise para os fenômenos linguísticos. Considera-se que o termo multilíngue aponte diretamente para os sistemas linguísticos resultantes do processo de contraste. Nesse sentido, o termo multilíngue não é definido pela comparação entre os textos em uma ou em mais línguas, mas pela abordagem ao estudo do objeto e de seus fenômenos, buscando explicar como a produção de significados em

<sup>12</sup> FIGUEREDO, G. P. Modelagem da equivalência em tradução: parâmetros para a descrição do ambiente multilíngue. In: SEMANA DAS LETRAS. Pluraridade da Memória, Departamento de Letras, Universidade Federal de Ouro Preto, 23-26 out. 2012. Comunicação Verbal.

diferentes contextos interlinguísticos ocorre e fazer generalizações sobre os registros produzidos no ambiente multilíngue.

### 1.5 Estudos de Registro e Estudos de Corpora

A abordagem teórico-metodológica proporcionada pela Linguística de Corpus permite observar padrões de frequências a partir de *corpora* que constituem uma amostra autêntica de língua (BIBER, 1988, 1995). A investigação linguística que tem o *corpus* como objeto de estudo comporta duas vertentes teóricas não excludentes (TOGNINI-BONELLI, 2001):

1. Baseada em *corpus*: de natureza confirmatória, em que a evidência dos dados pode ocupar a posição de modelo para uma teoria já formulada; e
2. Conduzida pelo *corpus*: de natureza exploratória e empírica, na qual o pesquisador busca padrões e distinções entre os dados constituintes de um *corpus* para o embasamento de formulações de generalizações linguísticas.

A proposta de Biber (1988, 1995; BIBER; CONRAD; REPPEN, 1998; BIBER *et al.*, 1999), adotada nesta pesquisa, segue a vertente baseada em *corpus*, na medida em que utiliza conceitos e categorias já existentes para a escolha das variáveis de estudo e para chegar a dimensões da linguagem verbal e a tipos de texto por meio de dados disponíveis. Assim, é possível extrair dimensões da linguagem verbal, consideradas construtos teóricos, a partir da exploração dos dados dos *corpora* da pesquisa em tela por meio de métodos empíricos e técnicas estatísticas.

Biber et al (1999) desenvolveram a gramática da língua inglesa *Longman grammar of spoken and written English* (LGSWE). Essa obra é baseada em um *corpus* de 40 milhões de palavras e descreve a gramática inglesa por meio de exemplos autênticos com enfoque tanto no inglês escrito quanto no inglês oral. Essa gramática considera também diferenças dialetais (variação da língua de acordo com o usuário) e diferenças entre o que denominam registros (variação da língua de acordo com o uso). Os autores exploram quatro tipos de variação ou registro, a saber: conversação (registro oral), ficção, notícias de jornal e prosa acadêmica (registros escritos). Segundo eles, cada amostra de língua inserida a partir de um registro específico constitui um texto que pode ser organizado a partir do meio escrito (*e.g.*, artigos científicos,

notícia e livro) ou oral (*e.g.*, palestras e diálogos). A gramática adota também os modelos de descrição privilegiados pela Linguística de Corpus (computação da frequência, descrição da co-ocorrência, além da indução de padrões léxico-gramaticais). A perspectiva de registro explorada neste tese, como será detalhado a seguir, se pauta por um entendimento do conceito de registro diferente daquele dos autores mencionados no parágrafo anterior. Trata-se do conceito de registro postulado pela teoria sistêmico funcional (MATTHIESSEN, 1993).

A interface entre os Estudos da Tradução, a LC e a LSF tem sido mapeada por diversos autores. De acordo com Pagano (2003, p. 1), “o diálogo entre os Estudos da Tradução e a Linguística Sistêmico-Funcional, conforme desenvolvida por M. A. K. Halliday e discípulos, conta atualmente com uma tradição de praticamente quatro décadas”. A autora afirma que, no período em questão, “tanto as teorias da tradução como a própria gramática sistêmica têm sido objeto de revisões e expansões em sintonia com novas teorias linguísticas e culturais que foram emergindo” (PAGANO 2003, p. 1). O caráter interdisciplinar dos Estudos da Tradução permite, dentre outras abordagens, o estabelecimento de uma interface entre os Estudos da Tradução, a Linguística de Corpus e a LSF.

Para subsidiar uma reflexão específica acerca do registro, examinam-se, a seguir, linhas de análise na área da Linguística Aplicada. Dentre elas, destacam-se, com base na LSF, (i) o estudo de Neumann (2008) em complementaridade com aquela sugerida por Biber *et al.* (1999) e o estudo de Eggins (2004), e (ii) Matthiessen, Teruya e Wu (2008), principalmente no que toca à possibilidade da classificação de textos baseada nos parâmetros do contexto (MATTHIESSEN; TERUYA; WU, 2008; FIGUEREDO, 2011).

O estudo de Neumann (2008) se pautou na complementaridade entre: (i) a Linguística de Corpus, para a descrição linguística por meio de métodos empíricos e técnicas estatísticas, e (ii) a Linguística Sistêmico-Funcional, para o entendimento das variações do registro. Nessa perspectiva de complementaridade, a autora se pautou nos estudos de Biber (1988, 1995) e Biber *et al.* (1999) da variação do registro.

Para a compilação de *corpora*, Neumann (2008) adotou os parâmetros definidos por Biber (1990, 1993), o qual sugere que, para fins de representatividade, é necessária a análise de uma amostra de dez textos do mesmo registro contendo trechos com, no mínimo, 1.000 palavras cada. Para tal, Neumann (2008) compilou em, primeiro lugar, um *corpus* de referência

composto por blocos de textos de 1.000 palavras cada, sendo 16 registros diferentes: reportagens de imprensa, editoriais, críticas, textos religiosos, histórias folclóricas, biografias, textos políticos, textos científicos, ficção, histórias de mistério, discursos preparados, receitas, romances, licitações, guias turísticos e decisões judiciais. Em seguida, a autora compilou um *corpus* com base na variabilidade funcional de registro (HALLIDAY; HASAN, 1976), englobando textos de popularização da ciência, de propaganda turística, de discursos preparados, de ensaios políticos sobre economia, de ficção, de comunicação empresarial, de manuais de instrução e de páginas da Internet (NEUMANN, 2005, p. 4). A presente tese seguiu a proposta de compilação de Neumann (2008), que será detalhada no Capítulo 2.

Para complementar a análise dos dados provenientes dos *corpora*, Neumann (2008) se pautou nas considerações de Halliday e Hasan (1976) sobre registro. De acordo com esses autores, o registro compreende uma variação linguística que ocorre em função de uma variação do contexto de situação. A complementaridade entre a LC e a LSF nos estudo de Neumann (2005) possibilitou investigar a natureza da explicitação que ocorre devido à diferença entre sistemas linguísticos, além de informar acerca de propriedades da tradução, como a simplificação, a normalização, a estabilização (*levelling out*) e a visibilidade da tradução (*shining through*) (NEUMANN, 2005, p. 2).

De acordo com Neumann (2008), a definição de registro surgiu com a estrutura social para a descrição do contexto de situação ou contexto imediato de enunciação, tendo como base as variáveis campo (*field*), sintonia (*tenor*) e modo (*mode*), o que remete respectivamente a cada uma das três metafunções da linguagem verbal: (i) ideacional (abrange a experiência humana); (ii) interpessoal (reflete as relações pessoais e sociais); (iii) textual (organiza o fluxo discursivo) (HALLIDAY; MATTHIESSEN, 2004). Neste momento faz-se mister explicitar as principais características de cada uma dessas variáveis, além de apontar recursos léxico-gramaticais passíveis de serem utilizados na análise textual de cada uma das variáveis e seus parâmetros, conforme sintetiza a FIG. 14.

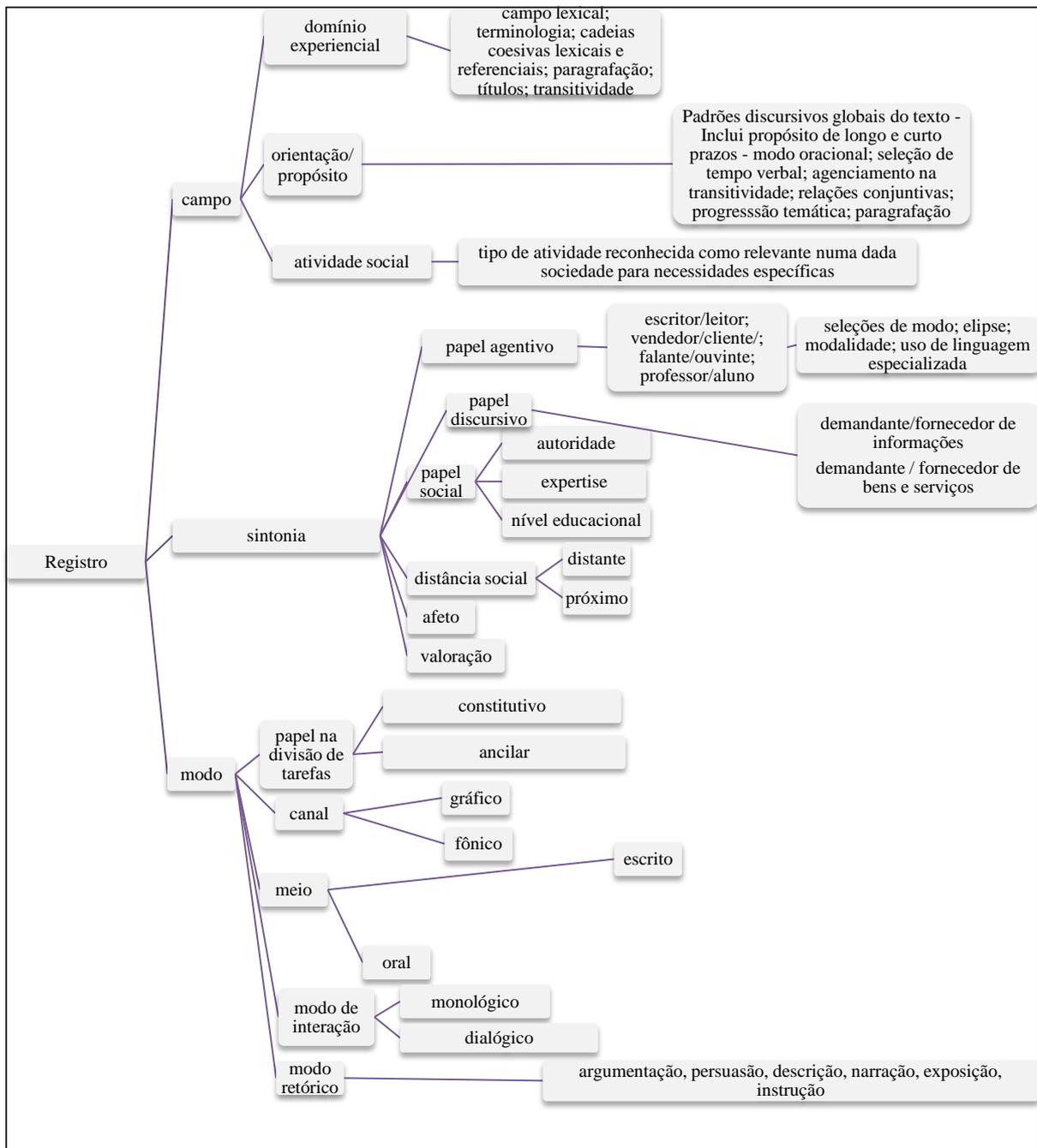


FIGURA 14. Variáveis do contexto de situação e seus respectivos parâmetros

Fonte: Pagano (2012, p. 20).

De acordo com a FIG. 14, o registro é definido pelas variáveis campo, sintonia e modo. Retomando o que foi dito anteriormente, o campo refere-se a um tópico específico da troca linguística em uma dada situação. A sintonia do discurso caracteriza o tipo de relação social entre os participantes na situação, destacando o papel discursivo, o papel social, o papel agentivo. Além desses parâmetros, há as relações de poder entre os participantes, que são caracterizadas pelo gênero social, nível de *expertise* e classe social. Já o afeto significa o

envolvimento dos participantes em uma dada situação e, por fim, a valoração, que consiste na atribuição de valor dentre os participantes. Finalmente, o modo do discurso descreve a maneira como a troca é realizada, determinando o papel da língua no contexto (orientada ao campo ou orientada à sintonia), considerando-se a interação entre os textos enquanto monológicos ou dialógicos, os meios falado ou escrito, os canais fônico, gráfico e eletrônico e o modo retórico (*i.e.*, argumentação, persuasão, descrição, narração, exposição ou instrução).

O estudo de Eggins (2004) explora os princípios e técnicas da abordagem sistêmica e funcionalista do uso da língua (*i.e.*, o que é a língua, como ela funciona e qual é a relação da língua com o contexto), pautando-se nas considerações de Halliday (1994). De acordo com a autora, essa abordagem propicia os meios necessários para entender como os significados na negociação dos diferentes tipos de texto são construídos. A FIG. 15 ilustra o modelo de estratificação da língua realizada com base nas três variáveis do contexto de situação (*i.e.*, campo, modo e sintonia), relacionadas às três metafunções (*i.e.*, ideacional, textual e interpessoal, respectivamente).

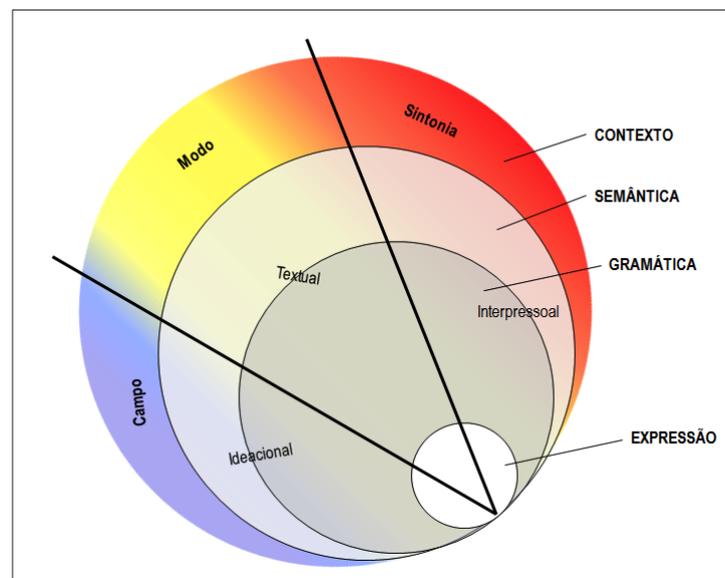


FIGURA 15. A realização da língua através das três variáveis do contexto de situação

Fonte: adaptada de Figueredo (2011, p. 79).

Em sua pesquisa, Eggins (2004) utiliza três tipos de texto de diferentes gêneros sobre o mesmo assunto, *crying baby*, a saber: (i) texto de popularização; (ii) texto acadêmico e (iii) registro oral (conversação). A autora enumera esses textos como 1.1, 1.2 e 1.3,

respectivamente, sendo os textos 1.1 e 1.2 classificados como pertencentes ao gênero explicativo/informativo e o texto 1.3, ao gênero interativo (diálogo).

O texto 1.1, um artigo de uma revista popular, foi escrito com o objetivo de dar conselhos aos pais que acabaram de ter um bebê; apesar de ser um texto escrito, podem ser observadas características de interação entre os participantes, principalmente por meio da utilização de pronomes pessoais, como “*you*”, “*he*”, “*they*” ou pelo uso de vocabulário do dia a dia. O texto 1.2, um texto acadêmico, não enseja interações explícitas entre os participantes. Esses dois textos dizem respeito a bebês que choram e o que fazer com eles nessa situação.

O primeiro texto diz que os bebês sempre choram por alguma razão específica, enquanto o segundo texto sugere que o choro dos bebês não tem um motivo claro ou uma explicação. Essa diferença é retrata pela autora em termos da ideologia do texto. Para a autora, a ideologia está acima do contexto de cultura, sendo esse o nível em que se percebe como os participantes constroem os momentos sociais e como se posicionam culturalmente. Nesse sentido, o texto 1.1 estimula mais os pais a saberem como lidar com o choro dos bebês, ao passo que o texto 1.2 os desestimula. O primeiro texto conduz os pais a se sentirem culpados pelo choro do bebê, e o segundo texto culpa o próprio bebê por chorar sem um motivo aparente. Já o texto 1.3 retrata um exemplo de linguagem oral, um diálogo que envolve vários participantes, fragmentos de frases, sendo claramente interativo. A língua em cada um dos textos é estruturada para se entender a mensagem de cada um deles a partir dos significados ideacionais (a partir do mundo das experiências), interpessoais (o papel de cada um dos participantes, proximidade entre autor-leitor, relações de poder) e textuais (escrita ou fala). Todos esses significados atuam simultaneamente.

Cada um dos três textos foi analisado por Eggins (2004) de acordo com:

1. Modo (relacionado à variável sintonia, metafunção interpessoal);
2. Transitividade (relacionada à variável campo, metafunção ideacional);
3. Complexo oracional;
4. Tema (relacionada à variável modo, metafunção textual);
5. Relações conjuntivas;
6. Referência;
7. Relações lexicais; e
8. Estrutura esquemática.

Eggins (2004) iniciou sua análise caracterizando os textos a partir do estrato da léxico-gramática (Modo, Transitividade, Relações Táticas e Lógico-Semânticas, e Tema); em seguida, realizou a análise coesiva desses textos; por fim, com as informações obtidas por meio dessas duas análises, interpretou e comparou os textos enquanto realizações do contexto de situação. No que tange à análise do estrato da léxico-gramática, Eggins (2004) relata que o texto 1.2 (artigo acadêmico) é orientado à construção de significados sobre ações e identidade. Esse texto apresentou número muito maior de nominalizações em comparação com o texto 1.1 (texto de popularização). Segundo Halliday (1994, p. 353), o texto se torna lexicalmente mais denso à medida que aumenta o número de nominalizações, sendo estas mais comuns na língua escrita, principalmente em textos acadêmicos.

Eggins (2004) destaca o conceito de texto segundo a LSF como unidade semântica na qual os significados de noção de mundo (ideacional), interação entre os participantes (interpessoal) e organização do texto (textual) são realizados por meio dos sistemas TRANSITIVIDADE, MODO e TEMA, respectivamente. A autora se pauta na concepção dos linguistas sistêmicos de que é preciso analisar como o contexto se insere no texto e define o registro como o “contexto de situação no texto”<sup>13</sup> (EGGINS, 2004, p. 85), ou seja, a forma como o contexto de situação se realiza no texto – forma essa passível de ser descrita pelas variáveis campo, sintonia e modo. Para Eggins (2004) o registro delinea o impacto das diferentes dimensões do contexto de situação, sendo essas dimensões as três variáveis do contexto de situação. A autora se baseia nessas três variáveis para identificar as diferenças e semelhanças entre os textos, pautando-se nas informações obtidas por meio da análise do estrato da léxico-gramática hallidayana até chegar ao nível do registro, perpassando pelos aspectos culturais aos quais os textos estão inseridos (contexto de cultura).

Como é apontado a seguir, diante das propostas de definição de registro com o aporte da LSF, é possível classificar cada um dos três tipos de textos específicos constituintes dos *corpora* compilados para a realização da presente pesquisa, considerando-se as especificidades das variáveis campo (*field*), sintonia (*tenor*) e modo (*mode*), conforme ilustrado na FIG. 15. Além disso, também é possível definir o lugar dos textos na tipologia dos textos com base nos parâmetros do contexto (MATTHIESSEN; TERUYA; WU, 2008).

---

<sup>13</sup> Minha tradução do original: “Register: context of situation in text”.

## 1.6 A Tipologia de Textos Baseada nos Parâmetros do Contexto

A LSF entende a língua como um sistema de escolhas, que viabiliza a construção de textos, nos quais podem ser encontrados padrões diferentes das escolhas sistêmicas. Ao contrário de outros modelos, a abordagem hallidayana investiga regularidades por meio da análise das escolhas feitas nos textos e disponibilizadas no contexto, do registro, dos recursos e da potencialidade de significados, em vez de analisar formas, sentenças isoladas, regras e gramaticalidade.

O modelo hallidayano é funcional no sentido de que é desenvolvido para explicar como a língua é usada, ou seja, como constrói a experiência do mundo exterior e do mundo interior por meio das relações sociais viabilizadas por textos coesos e coerentes. Para a pesquisa em tela, a produção textual foi observada a partir da tipologia de textos (MATTHIESSEN; TERUYA; WU, 2008), baseada no contexto e pautada por parâmetros das variáveis campo, sintonia e modo. Os dois tipos de texto selecionados para compor os corpora analisados, com base nos rótulos a eles atribuídos pela comunidade de usuários, foram caracterizados como registros diferenciados a partir dos processos sociosemióticos EXPLORAR (*exploring*) e HABILITAR (*enabling*), explicados a seguir mais detalhadamente.

Matthiessen, Teruya e Wu (2008) propuseram uma tipologia de registro baseada nas variáveis contexto, com base na pesquisa de Ure (1989). A FIG. 16 ilustra a tipologia dos textos, enfocando a linguagem no contexto de cultura (núcleo da roda).

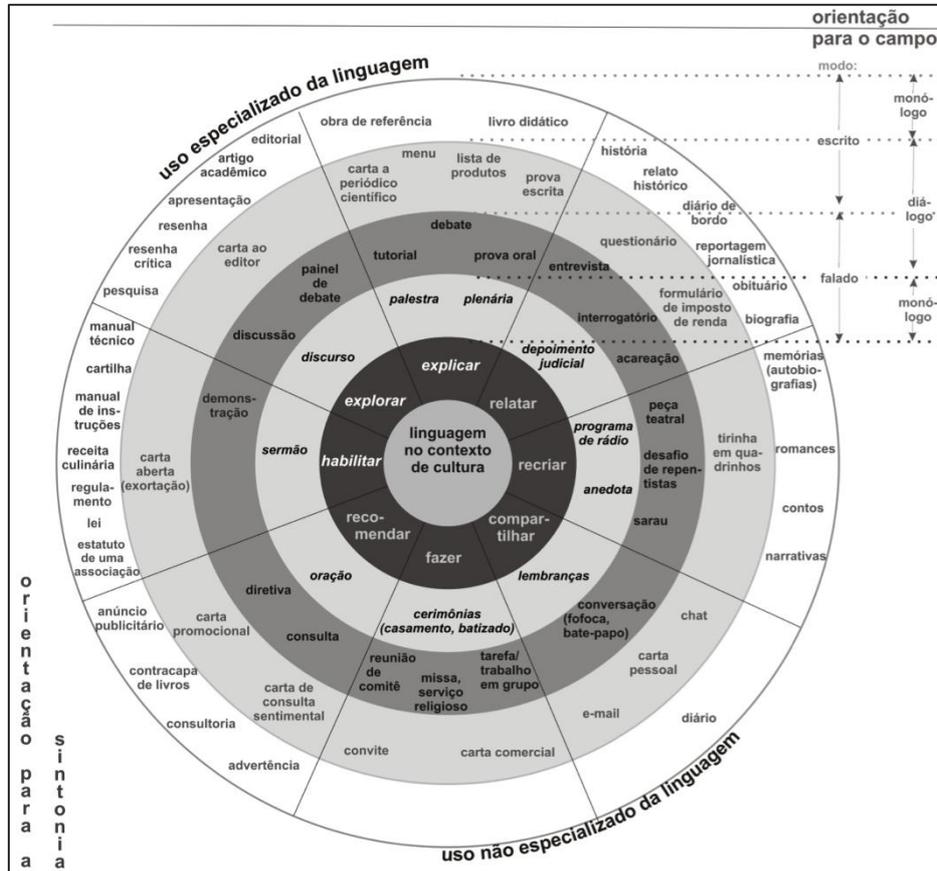


FIGURA 16. Tipologia de textos

Fonte: traduzido de Matthiessen, Teruya e Lam (2010, p. 221) por Pagano (2012).

Nessa figura, estão contemplados os oito processos sociossemióticos, considerando-se a variável campo:

1. **EXPLICAR** (*expounding*): elaborar teorias para a construção do conhecimento por meio de livros didáticos, palestras, entre outros;
2. **RELATAR** (*reporting*): construir linguisticamente um acontecimento por meio de reportagens jornalísticas, relato históricos, entre outros tipos de texto;
3. **RECRIAR** (*recreating*): criar um evento que aconteceu no mundo, de forma ficcional, considerando-se os aspectos da vida social, por meio, por exemplo, de romances e tirinha em quadrinhos;
4. **COMPARTILHAR** (*sharing*): estreitar os laços sociais e negociação de valores, considerando-se a proximidade e a distância entre os participantes. Têm-se como exemplos o diário e a conversa (*e.g.*, fofoca e bate-papo);

5. FAZER (*doing*): facilitar o desempenho de uma atividade não linguística por meio, por exemplo, de cerimônias (*e.g.*, casamento e batizado) ou missas; nesse processo, a língua não apresenta um papel de destaque;
6. RECOMENDAR (*recommending*): buscar controlar as ações e procedimentos dos participantes a partir da língua por meio de diretivas, consultorias, anúncios publicitários, entre outros;
7. HABILITAR (*enabling*): facilitar o comportamento dos participantes em uma dada situação por meio, por exemplo, de leis ou manuais de instrução, cartilhas;
8. EXPLORAR (*exploring*): criar significados para a construção de negociação de posições e valores com outros participantes dentro da comunidade por meio de pesquisas, painéis de debate, discussões, entre outros.

Esses oito processos, conforme ilustrado na FIG. 16, resumem a forma como a língua pode ser instanciada a partir do contexto de cultura. Nesta pesquisa, os textos selecionados foram classificados considerando-se os parâmetros das variáveis do contexto. Dentro dos processos sociosemióticos, as configurações de campo, sintonia e modo variam de acordo com cada um dos registros. Consequentemente, neste estudo, cada um dos tipos de texto foi caracterizado de acordo com cada uma dessas variáveis (*cf.* FIG. 16). O artigo acadêmico está inserido no processo sociosemiótico EXPLORAR, por meio da transmissão de conhecimento. Essa transmissão pode ser realizada entre os próprios pares ou entre especialistas e agentes da área da saúde, em diferentes tipos de texto (interação especialista – especialista). Tanto o manual técnico quanto (interação especialista – técnico da área da saúde) e a cartilha e *folders* de divulgação (interação especialista – leigo) estão inseridos no processo sociosemiótico HABILITAR, que diz respeito à língua em uso por meio de capacitação do leitor a realizar com sucesso uma ação ou um procedimento.

A seguir, está a apresentado o Capítulo de Metodologia, que apresenta os procedimentos de coleta e análise de dados para a realização desta pesquisa.

# **CAPÍTULO 2**

## **METODOLOGIA**

## 2 METODOLOGIA

**E**ste capítulo apresenta os procedimentos de coleta e análise de dados. No caso específico da coleta, justifica-se a seleção dos três tipos de texto que compõem os *corpora* comparáveis (MCENERY; XIAO, 2007; GRANGER, 2003) deste estudo. Na sequência, descrevem-se os procedimentos de extração dos dados dos *corpora*, por meio de ferramentas metodológicas como o ambiente de programação R e de anotação morfossintática como o Treetagger, que subsidiam a busca por padrões de uso linguístico (SINCLAIR, 1991; BERBER SARDINHA, 2004), para fins de caracterização dos registros. Por fim, apresenta-se a extração dos dados por meio do programa UAM CorpusTool, com o objetivo de investigar frequências relativas dos sistemas de TRANSITIVIDADE, MODO, e TEMA e MENSAGEM, para propor uma descrição sistêmico-funcional dos tipos de texto (FIGUEREDO, 2011).

No que diz respeito à metodologia de análise dos dados, são apresentados os procedimentos de análise dos dados empíricos extraídos dos *corpora*. Em seguida, os procedimentos de análise para a classificação dos tipos de textos a partir dos parâmetros das variáveis do contexto a partir dos processos sociossemióticos e o procedimento de análise da descrição do perfil metafuncional dos textos a partir dos sistemas de MODO, TRANSITIVIDADE e TEMA e MENSAGEM (HALLIDAY; MATTHIESSEN, 1999). Por último, apresenta-se a forma como o *template* foi construído neste estudo como produto passível de ser incorporado em um *pipeline* semiautomático direcionado para a elaboração de cartilha (com base no perfil de um texto de popularização da ciência).

### 2.1 Metodologia de Coleta de Dados

Os procedimentos de coleta de dados desta tese consistiram na adoção de duas abordagens metodológicas complementares, quais sejam: (i) Linguística de Corpus (MCENERY; XIAO, 2007; GRANGER, 2003), com vistas à compilação e à extração de co-ocorrências de itens lexicais e itens gramaticais dos *corpora* e ao etiquetamento desses itens em classes de palavras por tipo de texto e por língua; e (ii) Linguística Sistêmico-Funcional, tendo como

base o trabalho de Figueredo (2011), no intuito de extrair dados para a descrição do perfil metafuncional dos textos em português e em inglês.

### 2.1.1 *Compilação dos corpora*

A compilação dos *corpora* comparáveis desta pesquisa resultou da coleta de textos autênticos e originais da medicina (subárea triagem neonatal para anemia falciforme) provenientes do Núcleo de Ações em Pesquisa em Apoio Diagnóstico da Faculdade de Medicina da Universidade Federal de Minas Gerais (NUPAD/FM/UFMG).

Esse Núcleo conta com uma biblioteca virtual gerenciada por meio do programa Personal Home Library (PHL)<sup>14</sup>. Toda a produção acadêmica e técnico-científica gerada e recebida pelo Núcleo é cadastrada nesse programa por uma bibliotecária, utilizando-se palavras-chaves da área da saúde, conforme consta nos Descritores em Ciências da Saúde (DeCS), criados pelo Centro Latino-Americano e do Caribe de Informação em Ciências da Saúde (SciELO), para servir como uma linguagem única na indexação de produções acadêmicas e técnico-científicas. Para a obtenção dos textos, foi realizada uma pesquisa nessa biblioteca, utilizando-se a ferramenta de busca do PHL e inserindo-se os descritores “anemia falciforme” / “doença falciforme” para os textos em português e “*sickle cell disease*” / “*sickle cell anemia/anaemia*” para os textos em inglês, para o período de 2010 a 2012, com o objetivo de restringir o número de documentos a serem encontrados. As siglas DF (doença falciforme) ou SCD (*sickle cell disease*) não foram utilizadas nas buscas pelos textos.

Todos os textos selecionados estavam em formato .pdf (*portable document format*) e precisaram ser convertidos na íntegra para o formato .docx, arquivo do programa Microsoft Word (a partir da versão 2007), para que os *corpora* pudessem ser organizados cada qual em um *corpus* específico de acordo com o tipo de texto ou rótulo a eles atribuído, denominado nesta análise de “registro”, e língua. Nesse sentido, investiga-se se dados empíricos provenientes dos *corpora* facilitam o processo de busca por padrões de uso linguístico nesses textos, com o seguinte objetivo:

- Extrair palavras e combinações de palavras no intuito de identificar padrões de uso linguístico específicos para cada tipo de *corpus*.

---

<sup>14</sup> Para mais informações sobre o PHL, cf.: <<http://glima.eci.ufmg.br/guiaprograma/pages/biblioteca/programas-livres/phl.php>>. Último acesso em: 15 set. 2012.

Para a seleção da amostra para o estudo, adotou-se a metodologia descrita em Biber (1990), retomada por Neumann (2005). De acordo com Biber (1990 *apud* NUNES, 2010, p. 60), o termo amostra é entendido como “a escolha (aleatória ou não) de um subconjunto de elementos pertencentes a uma população”. Biber (1990) propõe que cada texto dos diferentes *corpora* seja identificado por um cabeçalho (*header*) que contenha informações como: classificação do tipo de texto, autoria, título, local e ano de publicação, editora, edição, número de páginas, *status* do texto (original ou tradução) e tipo de amostra (parcial ou integral).

A amostra foi selecionada aleatoriamente, aplicando-se a fórmula “randbetween” do programa Microsoft Excel©. Essa fórmula permite que seja escolhida uma página aleatória dentro de um intervalo entre o número da primeira página e o número da última página de cada um dos registros, não se considerando a capa, contracapa e páginas contendo informações sobre as referências.

Após a seleção aleatória do número da página, foram contadas aproximadamente três mil palavras correntes de cada um dos três tipos de texto, certificando-se de que foram selecionados parágrafos completos (*cf.* NUNES, 2010). Assim, os textos selecionados foram copiados e colados em um novo arquivo .docx. Esse novo arquivo foi salvo ainda em formato .docx e passou a ser considerado o *corpus* da pesquisa.

Com esse procedimento, os textos analisados não foram selecionados na íntegra. A escolha do número de *tokens* inclusos em cada *corpus* foi planejada inicialmente com a finalidade de contemplar cerca de 1.000 palavras para cada um dos tipos de texto. Esse número de *tokens* foi o mesmo proposto no estudo de Biber (1990), que tece algumas considerações sobre tendências de repetição de alguns padrões representativos em um mesmo tipo textual, após analisar dez textos do mesmo tipo de texto com trechos de pelo menos 1.000 *tokens*. Como o propósito desta pesquisa é identificar padrões de uso linguístico mais frequentes, optou-se por compilar os *corpora* mais extensos de forma a contemplar 3.000 para cada um dos três tipos de textos (artigos acadêmicos; manuais técnicos; e cartilhas e *folders* de divulgação) e para cada uma das línguas (português e inglês). Cumpre lembrar que para tal compilação foram selecionados textos aos quais se atribui esses rótulos nos respectivos contextos pelas respectivas comunidades de usuários. Esses *corpora* foram compilados conforme mostra o QUADRO 5 a seguir.

QUADRO 5. Compilação dos *corpora* comparáveis bilíngue por tipo de texto conforme processos sociossemióticos

Processo socios-semiótico	Rótulo	Língua	Total de <i>tokens</i> (aproximado)
EXPLORAR	Artigos acadêmicos (interação especialista – especialista)	Português	3.000
		Inglês	3.000
HABILITAR	Manuais técnicos (interação especialista – técnico da área da saúde)	Português	3.000
		Inglês	3.000
HABILITAR	Cartilhas e <i>folders</i> de divulgação (interação especialista – leigo)	Português	3.000
		Inglês	3.000

Em seguida, os textos que compuseram os *corpora* foram copiados e colados no Bloco de Notas, devidamente salvos em formato .txt e, posteriormente, tratados de forma rigorosa para que pudessem ser lidos pelo ambiente de programação R, por meio de *scripts*. A maneira de preparar os *corpora* e o desenvolvimento desses *scripts* são abordados mais detidamente na Seção 2.2, a seguir.

A FIG. 17 ilustra o cabeçalho (*header*) de um dos textos comparáveis utilizados. Vale ressaltar que cada um dos seis textos analisados durante o desenvolvimento deste estudo foi identificado por um cabeçalho como este.

<KIKUCHI, Berenice. São Paulo: OPAS/OMS, 2012. Trecho do livro: Anemia Falciforme manual para trabalhadores de saúde e educação nas Américas, p. 53-61. Original. Manual técnico. Parcial. Português brasileiro. Brasil.>

FIGURA 17. Exemplo de cabeçalho de um dos textos comparáveis para o tipo de texto manual técnico

Por fim, para a análise com o aporte da Linguística Sistemico-Funcional a partir da aplicação dos perfis metafuncionais, foi selecionado um texto do tipo EXPLORAR e dois textos do tipo HABILITAR, com a finalidade de fazer uma diferenciação entre os registros da área da medicina, subárea triagem neonatal para anemia falciforme.

### 2.1.2 Coleta de dados

Apresentam-se, a seguir, as ferramentas metodológicas utilizadas para a extração de dados e tratamento estatístico, a partir da aplicação dos *scripts* do ambiente de programação R, anotação morfossintática por meio do programa Treetagger e anotação de frequências relativas através do programa UAM CorpusTool®.

### 2.1.2.1 Extração de dados através do ambiente de programação R

Conforme consta no manual, o R<sup>15</sup> é uma linguagem computacional desenvolvida a partir da linguagem S e S-Plus e se destina, dentre outras finalidades, à manipulação e análise estatística de dados numéricos e categóricos. O R foi desenvolvido por dois professores do Departamento de Estatística da Universidade de Auckland (Nova Zelândia) – Ross Ihaka e Robert Gentleman – e é distribuído gratuitamente pela Internet. A utilização dos recursos desse programa permite a criação de funções e modelos estatísticos diversos com vários níveis de complexidade, possibilitando a geração de gráficos, tabelas e matrizes. Além disso, o R facilita a manipulação e a análise de forma exploratória tanto de dados numéricos quanto categóricos, sendo também utilizado em pesquisas de *corpora* devido ao seu potencial de manipulação de caracteres. É possível utilizar o R em inúmeras etapas de análise dos dados, pois ele, além de permitir análises estatísticas e armazenar grandes quantidades de textos, possibilita a busca de itens lexicais e gramaticais e suas frequências de ocorrências nos *corpora*.

Por se tratar de uma linguagem computacional, a manipulação de dados é facilitada através da criação de *scripts*, ou seja, sequências de comandos específicos para cada pesquisa. No que toca à pesquisa em tela, o primeiro *script* desenvolvido se destinou ao tratamento/adequação dos *corpora*, para que pudessem ser lidos pelo R.

Após a seleção aleatória de cerca de 3.000 palavras – por meio da função “randbetween” no Microsoft Excel, conforme já apontado –, organizou-se o texto em sentenças completas, cada sentença em uma linha, com espaçamento simples e sem marcas de parágrafos entre as linhas. Uma vez que o ambiente de programação R faz distinção entre caracteres maiúsculos e minúsculos, todos os caracteres foram transformados em minúsculos, a fim de evitar que uma mesma palavra fosse contabilizada duas vezes apenas em razão de diferenças grafológicas.

Cada um dos textos foi, então, transferido para um arquivo do bloco de notas no formato .txt (cf. FIG. 18).

---

<sup>15</sup> Para informações mais detalhadas acerca do programa, cf.: <<http://www.r-project.org/>>Último acesso em: 15 set. 2012.

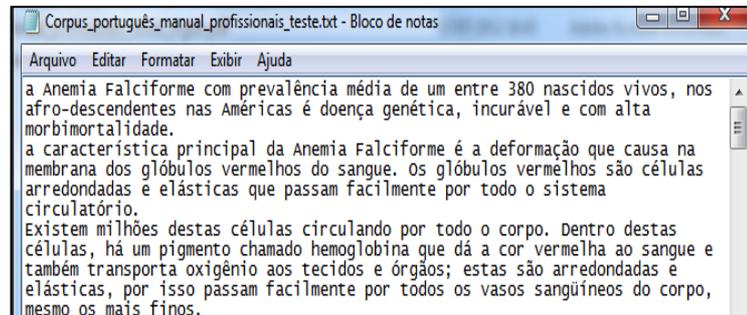


FIGURA 18. Exemplo do *corpus* compilado

A FIG. 18 mostra a compilação de um do *corpus* do estudo, que serviu como modelo para a compilação de todos os *corpora*. Observou-se que alguns caracteres são considerados parte de uma palavra pelo ambiente de programação R (e.g., a palavra “anemia:”, seguida de dois pontos, é diferente de “anemia,” seguida de vírgula, e ambas são diferentes de “anemia” sem qualquer pontuação). Sendo assim, antes de proceder à contagem de frequências, fez-se a limpeza dos *corpora* no que diz respeito a caracteres especiais, incluindo sinais de pontuação.

Para importar o *corpus* para o R, foi utilizado o seguinte comando 1:

`x=scan(file=choose.files(),what="char",sep="\n",skip=0,quiet=F)` Comando (1)

Esse comando (1) deve ser digitado na linha *prompt* da área de trabalho do R. Ao digitar esse comando no programa e pressionar a tecla “enter”, uma janela abre automaticamente, e o usuário pode localizar o arquivo que deseja importar para o ambiente de programação R. Esse comando apenas roda na plataforma Windows.

A FIG. 19 ilustra como os dados foram lidos pelo R.

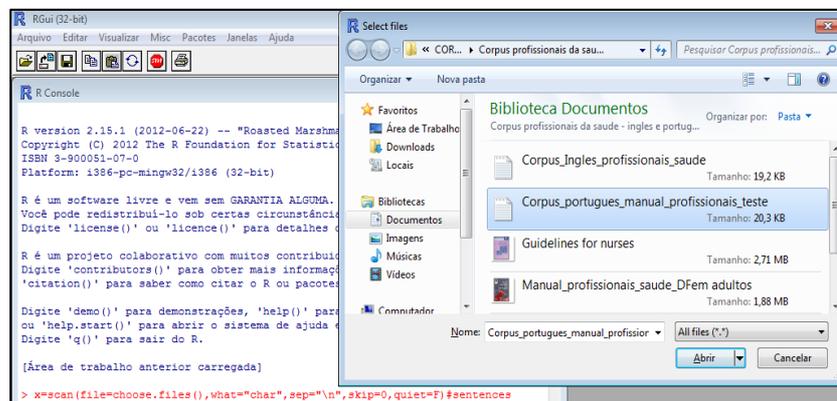
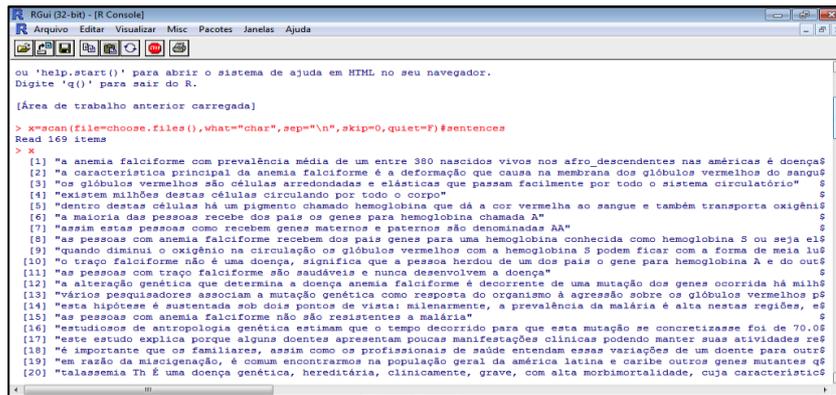


FIGURA 19. Leitura do *corpus* a partir do ambiente de programação R

O “x” do comando (1) significa que foi criado um objeto (vetor), cujo conteúdo é aquele proveniente do arquivo .txt. Conforme mostrado na FIG. 20, ao digitar o x na linha *prompt* do programa, verifica-se que as 169 linhas (*i.e.*, sentenças) referentes ao *corpus* de exemplo foram lidas e, a partir desse momento, estão prontas para serem analisadas com base em *scripts* específicos, conforme a demanda da pesquisa.



```

RGui (32-bit) - R Console
Arquivo Editar Visualizar Misc Pacotes Janelas Ajuda

ou 'help.start()' para abrir o sistema de ajuda em HTML no seu navegador.
Digite 'q()' para sair do R.

[Área de trabalho anterior carregada]

> x = scan(file=choose.files(), what="char", sep="\n", skip=0, quiet=F) #sentences
Read 169 items
> x
[1] "a anemia falciforme com prevalência média de um entre 380 nascidos vivos nos afro_descendentes nas américas é doença?"
[2] "a característica principal da anemia falciforme é a deformação que causa na membrana dos glóbulos vermelhos do sangue"
[3] "os glóbulos vermelhos são células arredondadas e elásticas que passam facilmente por todo o sistema circulatório"
[4] "existem milhões destas células circulando por todo o corpo"
[5] "dentro destas células há um pigmento chamado hemoglobina que dá a cor vermelha ao sangue e também transporta oxigênio"
[6] "a maioria das pessoas recebe dos pais os genes para hemoglobina chamada A"
[7] "assim estas pessoas como recebem genes maternos e paternos são denominadas AA"
[8] "as pessoas com anemia falciforme recebem dos pais genes para uma hemoglobina conhecida como hemoglobina S ou seja eis"
[9] "quando diminui o oxigênio na circulação os glóbulos vermelhos com a hemoglobina S podem ficar com a forma de meia lua"
[10] "o traço falciforme não é uma doença, significa que a pessoa herdou de um dos pais o gene para hemoglobina A e do out"
[11] "as pessoas com traço falciforme são saudáveis e nunca desenvolvem a doença"
[12] "a alteração genética que determina a doença anemia falciforme é decorrente de uma mutação dos genes ocorrida há milh"
[13] "vários pesquisadores associam a mutação genética como resposta do organismo à agressão sobre os glóbulos vermelhos p"
[14] "esta hipótese é sustentada sob dois pontos de vista: milenarmente, a prevalência da malária é alta nestas regiões, e"
[15] "as pessoas com anemia falciforme não são resistentes a malária"
[16] "estudiosos de antropologia genética estimam que o tempo decorrido para que esta mutação se concretizasse foi de 70.00"
[17] "este estudo explica porque alguns doentes apresentam poucas manifestações clínicas podendo manter suas atividades re"
[18] "é importante que os familiares, assim como os profissionais de saúde entendam essas variações de um doente para outr"
[19] "em razão da miscigenação, é comum encontrarmos na população geral da américa latina e caribe outros genes mutantes q"
[20] "talassemia Th É uma doença genética, hereditária, clinicamente, grave, com alta morbimortalidade, cuja característic

```

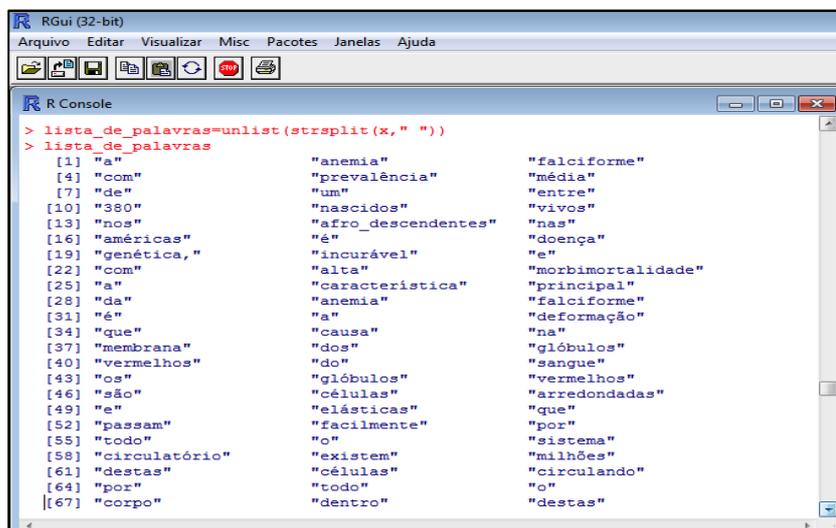
FIGURA 20. Tela inicial do ambiente de programação R, mostrando o comando para importar documentos em formato .txt

Para fazer uma lista de frequência, é preciso que o R separe cada palavra de cada frase como objeto único. Para tanto, utiliza-se o comando 2:

`unlist(strsplit(x, " "))`

Comando (2)

A FIG. 21 ilustra o uso desse comando (2).



```

RGui (32-bit)
Arquivo Editar Visualizar Misc Pacotes Janelas Ajuda

R Console

> lista_de_palavras=unlist(strsplit(x, " "))
> lista_de_palavras
[1] "a" "anemia" "falciforme"
[4] "com" "prevalência" "média"
[7] "de" "um" "entre"
[10] "380" "nascidos" "vivos"
[13] "nos" "afro_descendentes" "nas"
[16] "américas" "é" "doença"
[19] "genética," "incurável" "e"
[22] "com" "alta" "morbimortalidade"
[25] "a" "característica" "principal"
[28] "da" "anemia" "falciforme"
[31] "é" "a" "deformação"
[34] "que" "causa" "na"
[37] "membrana" "dos" "glóbulos"
[40] "vermelhos" "do" "sangue"
[43] "os" "glóbulos" "vermelhos"
[46] "são" "células" "arredondadas"
[49] "e" "elásticas" "que"
[52] "passam" "facilmente" "por"
[55] "todo" "o" "sistema"
[58] "circulatório" "existem" "milhões"
[61] "destas" "células" "circulando"
[64] "por" "todo" "o"
[67] "corpo" "dentro" "destas"

```

FIGURA 21. Criação da lista de palavras

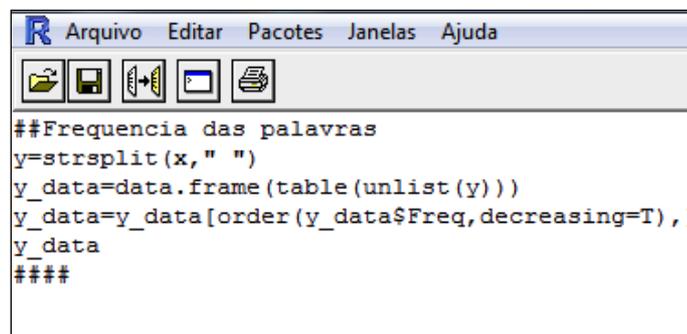
Após a compilação da lista de palavras, observou-se que algumas palavras estavam seguidas ainda de alguns sinais de pontuação. Foi criado um *script* para eliminar/limpar todos esses caracteres especiais e não relevantes para a presente pesquisa.

Apresentam-se a seguir, a aplicação dos *scripts* desenvolvidos para a extração dos dados dos *corpora*.

Após preparados para serem lidos pelo R (formato .txt), os *corpora* receberam rótulos específicos, conforme listado a seguir, totalizando seis arquivos (3 tipos de texto x 2 línguas):

1. artigo\_ingles.txt;
2. artigo\_portugues.txt;
3. manual\_ingles.txt;
4. manual\_portugues.txt;
5. cartilha\_ingles.txt; e
6. cartilha\_portugues.txt.

Para a leitura desses arquivos, extração automática dos dados e aplicação dos testes estatísticos, foram desenvolvidos *scripts* específicos para esta pesquisa. O primeiro *script* processado para a extração automática de dados foi aquele desenvolvido para mostrar as características gerais dos *corpora* compilados em relação à frequência absoluta do total de *tokens* e lista de *types*, conforme ilustrado na FIG. 22.



```

R  Arquivo  Editar  Pacotes  Janelas  Ajuda
[Icons: File Explorer, Save, Undo, Redo, Print]
##Frequencia das palavras
y=strsplit(x, " ")
y_data=data.frame(table(unlist(y)))
y_data=y_data[order(y_data$Freq,decreasing=T),]
y_data
####

```

FIGURA 22. Lista de *types* e frequência de *tokens*

Esse *script* reproduzido na FIG. 22 foi aplicado em cada um dos *corpora*. A partir dos *tokens* foi possível contabilizar todas as palavras presentes no *corpus* e visualizá-las por ordem alfabética ou por frequência de ocorrência. Os números provenientes da aplicação do *script*

têm a função de descrever o tamanho de cada um dos *corpora*. A partir dos números obtidos em cada um dos arquivos, foi possível analisar a relação *type/token* simples e padronizada<sup>16</sup> para cada um dos tipos de texto que compuseram os *corpora* nas línguas inglês e português. Para tanto, também foi desenvolvido um segundo *script*, conforme ilustra a FIG. 23.

```
##Frequência das palavras
y=unlist(strsplit(x," "))
p=seq(1,length(y),500)
p1=p+499
ratio_block=c()
for(i in 1:length(p)){
  block=y[p[i]:p1[i]]
  type_token=length(unique(na.omit(block)))/length(na.omit(block))
  ratio_block=append(ratio_block,type_token)
}
mean(ratio_block) ## Standardized
length(unique(y))/length(y) # Non-standardized
```

FIGURA 23. Relação *type/token* simples e padronizada

Logo, após o ambiente de programação R ter lido os textos e gerado a lista de *types* e de frequência de *tokens*, bastou programar outros *script* que atendessem às necessidades da pesquisa em tela. Para a obtenção de dados por meio dos arquivos em formato .txt, observou-se o seguinte fluxo:

1. Extração da média das palavras distribuídas por sentença;
2. Extração de itens lexicais a partir de nódulos específicos;
3. Criação de gráficos e tabelas; e
4. Aplicação do teste de associação  $\chi^2$ .

Cada um dos comandos gerados para a extração dos dados dos *corpora* foi testado em cada um dos *corpora* e está disponível no Anexo 1.

Vale lembrar que, de acordo com o que foi explicado na Subseção 2.1.1, os textos que compõem os *corpora* foram provenientes da base de dados do NUPAD/FM/UFMG, utilizando-se para a obtenção deles os descritores da área da saúde “anemia falciforme” /

<sup>16</sup> Berber Sardinha (2006, p. 175) define a razão *type/token* padronizada como “uma razão type-token média, calculada em blocos do texto. O tamanho de cada bloco é definido pelo usuário (no menu ‘Settings / Min-Max Frequencies / Type-Token Basis’), mas o valor padrão (‘default’) é 500 palavras”.

“doença falciforme” para os textos em português e “*sickle cell disease*” / “*sickle cell anemia/anaemia*” para os textos em inglês. Esses mesmos descritores foram utilizados como os primeiros nós identificados para serem analisados neste estudo, com a finalidade de testar os *scripts* desenvolvidos para a extração dos dados, aplicação dos testes estatísticos e identificação dos itens lexicais a serem utilizados como nós para a busca de padrões de uso linguístico. Esses padrões sugerem indícios de como os significados estão sendo construídos nos diferentes registros.

Cabe explicar, que conceitualmente, “anemia falciforme” e “doença falciforme” não são sinônimos. A mesma afirmação pode ser feita em relação a “*sickle cell disease*” e “*sickle cell anemia/anaemia*”. A anemia falciforme acomete o indivíduo que apresenta o perfil hemoglobínico HbSS. Entretanto, em alguns indivíduos ocorre a união de um gene S com outro gene doente (C, D, Talassemia), levando ao aparecimento da hemoglobinopatia SC, SD, S - Talassemia. Todas as hemoglobinopatias juntas, incluindo a anemia falciforme (SS), são chamadas doenças falciformes (SS, SD, SC, S - Talassemia ) (SERJEANT, G; SERJEANT, B., 1991, p. 31).

Assim, é importante pesquisa que mostre sistematicamente como tais nós são apresentados de modo a construir uma unidade semântica nos diferentes tipos de texto. Além disso, faz-necessário também analisar quais itens lexicais e gramaticais que acompanham esses nós e, então, extrair os padrões de uso linguístico em cada um dos *corpora*.

Foi criado um *script*, ilustrado na FIG. 24, que gera linhas de concordância para cada um dos nós (“anemia falciforme” / “doença falciforme” e “*sickle cell disease*” / “*sickle cell anemia/anaemia*”).

```

R Console
[3166] "cuidados"      "de"
[3169] "e"              "proteção"
[3172] "microrganismos" "que"
[3175] "infecção"
> x=scan(file=file.choose(),what="char")
Read 169 items
> x=gsub("[.,:;()]", "", x, ignore.case=T)
> x=gsub("[&]", "por_cento", x, ignore.cas
> x=gsub(" ", " ", x, perl=T) # this gets
> y=strsplit(x, " ")
> data.frame(table(unlist(y)))
  Var1 Freq
1     -     2
2     1    10
3     2    12
4    120    1
5 14por_cento 1
6     15     2
7 150000     1
8    2001     1
9     24     1
10    3000     1
11 30por_cento 1
12    380     1

Editor R
k=scan(file=file.choose(), what="char", sep="\n", skip=0, quiet=F) # sentences
x=gsub("[.,:;()]", "", x, ignore.case=T, perl=T)
x=gsub("&", "por_cento", x, ignore.case=T, perl=T)
x=gsub(" ", " ", x, perl=T) # this gets rid of EXTRA spaces
y=strsplit(x, " ")
data.frame(table(unlist(y)))
list.frames=c()
for(i in 1:length(y)){ # go through the list
  current.line=y[[i]] # access each line
  for(j in 1:(length(current.line)-1)){ #go up to the antepenultimate words
    list.frames=append(list.frames,paste(current.line[j],current.line[
  ]#end of loop go through the line
}#end of loop go through the list
match=grep("(\\w+\\W\\w+) (\\Wfalciforme\\W) (\\w+\\W\\w+)", list.frames, perl=T, v
target=gsub("(\\w+\\W\\w+ (\\Wfalciforme\\W) \\w+\\W\\w+", "\\1", match, perl=T)
A1=gsub("(\\w+\\W\\w+ (\\Wfalciforme\\W) \\w+\\W\\w+", "\\1", match, perl=T)
A2=gsub("(\\w+\\W\\w+\\W\\w+\\W\\w+\\W\\w+", "\\1", match, perl=T)
D1=gsub("(\\w+\\W\\w+\\W\\w+\\W\\w+\\W\\w+\\W\\w+", "\\1", match, perl=T)
D2=gsub("(\\w+\\W\\w+\\W\\w+\\W\\w+\\W\\w+\\W\\w+", "\\1", match, perl=T)
concordance=data.frame(A2,A1,target,D1,D2)
concordance

```

FIGURA 24. *Script* para localizar as concordâncias por meio dos nódulos ou itens lexicais

Com esse *script*, pode-se visualizar, a partir de um nódulo, os itens lexicais e gramaticais que ocorrem antes e depois desse nódulo. Esse *script* foi aplicado em todos os outros *corpora* compilados. Vale ressaltar que, a partir desse *script*, é possível localizar qualquer outro nódulo, bastando apenas substituir as palavras já utilizadas por outras.

Para verificar a frequência de ocorrência dos nódulos “doença falciforme” e “anemia falciforme” para cada um dos *corpora* em português, bem como a frequência de ocorrência dos nódulos “*sickle cell disease*” e “*sickle cell anemia/anaemia*” para cada um dos *corpora* em inglês, foi aplicado o *script* ilustrado na FIG. 25, a seguir. Essa figura mostra o *script* aplicado para a seleção e identificação das palavras mais frequentes em cada um dos *corpora*.

```

R Console
Digite 'contributors()' para obter mais
'citation()' para saber como citar o R.
Digite 'demo()' para demonstrações, 'he
ou 'help.start()' para abrir o sistema
Digite 'q()' para sair do R.
> x=scan(file=file.choose(),what="char")
Read 167 items
> x=gsub("[.,:;()]", "", x, ignore.case=T)
> x=gsub("[&]", "por_cento", x, ignore.cas
> x=gsub(" ", " ", x, perl=T) # this gets
> y=strsplit(x, " ")
> data.frame(table(unlist(y)))
  Var1 Freq
1     1    16
2     1     1

Editor R
x=scan(file=file.choose(), what="char", sep="\n", skip=0, quiet=F) # sentences
x=gsub("[.,:;()]", "", x, ignore.case=T, perl=T)
x=gsub("&", "por_cento", x, ignore.case=T, perl=T)
x=gsub(" ", " ", x, perl=T) # this gets rid of EXTRA spaces
y=strsplit(x, " ")
data.frame(table(unlist(y)))
data.frame(sort(table(unlist(y))))

```

FIGURA 25. *Script* para criar lista de frequência das palavras

Portanto, ao rodar o *script* no R, surge uma lista de palavras e o número de vezes que cada palavra ocorre no texto (FIG. 26). Os dados provenientes dessa lista foram copiados e salvos em planilhas do Microsoft Excel®.

The screenshot shows the RGui interface. The R Console window displays an error message: "Erro: ')' inesperado em "Y=table(unlist(y))". Below the error, a list of words is shown with their frequencies: 'the' (207), 'of' (166), 'and' (141), 'in' (104), 'to' (84), 'for' (76), 'pain' (76), 'with' (68), 'was' (56), 'a' (54), 'were' (51), 'children' (46), 'days' (36), 'sod' (35), 'on' (34), 'patients' (34), 'episodes' (30), 'as' (26), 'or' (26), 'that' (26), 'diary' (22), and 'not' (22). The script editor window shows the following R code:

```
x=scan(file=file.choose(),what="char",sep="\n",skip=0,quiet=F) # sentences
x=gsub("[.,:()]", "", x, ignore.case=T, perl=T)
x=gsub("[%]", "por_cento", x, ignore.case=T, perl=T)
x=gsub(" ", " ", x, perl=T) # this gets rid of EXTRA spaces
y=strsplit(x, " ")
data.frame(table(unlist(y)))
Y=table(unlist(y))
data.frame(sort(table(unlist(y)), TRUE))
AA=c(2, 1, 3, 6, 3, 14, 2)
sort(AA, decreasing = TRUE)
```

FIGURA 26. Lista de palavras mais frequentes em ordem decrescente

A partir de uma lista de palavras (*cf.* FIG. 26), cada item lexical pode ser analisado individualmente, considerando-se aquele cuja frequência foi significativa em termos estatísticos. Para tanto, basta replicar o *script* a seguir, transcrito na íntegra, substituindo o item lexical a ser localizado. Os Exemplos 1 e 2 a seguir exemplificam esse procedimento. No Exemplo 1, observa-se a busca pelo item lexical “falciforme”.

### Exemplo 1:

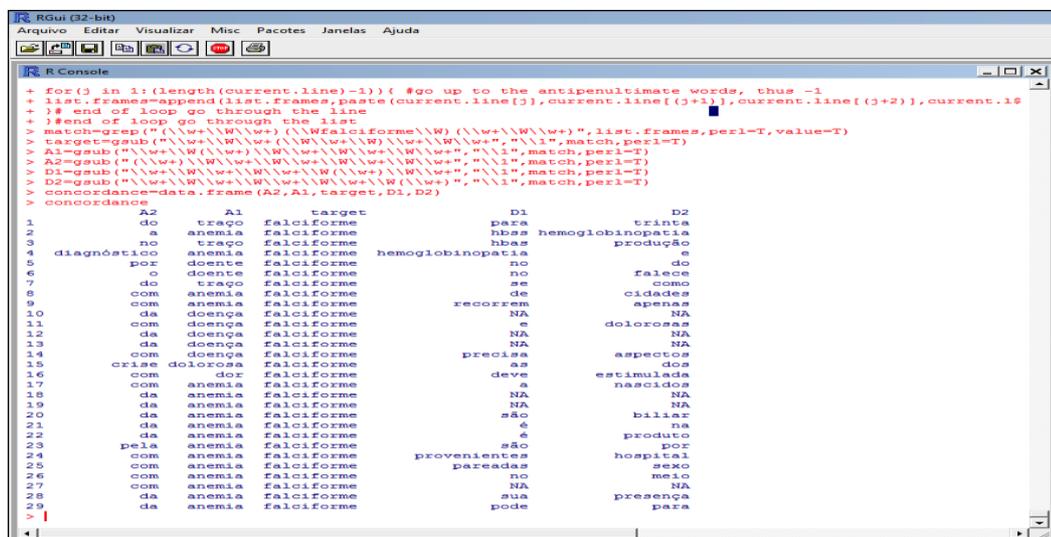
```
x=scan(file=file.choose(),what="char",sep="\n",skip=0,quiet=F) # sentences
x=gsub("[.,:()]", "", x, ignore.case=T, perl=T)
x=gsub("[%]", "por_cento", x, ignore.case=T, perl=T)
x=gsub(" ", " ", x, perl=T) # this gets rid of EXTRA spaces
y=strsplit(x, " ")
data.frame(table(unlist(y)))
list.frames=c()
for(i in 1:length(y)){# go through the list
current.line=y[[i]] # access each line
for(j in 1:(length(current.line)-1)){#go up to the antipenultimate words, thus -1
list.frames=append(list.frames,paste(current.line[j],current.line[(j+1)],current.line[(j+2)],current.line[(j+3)],current.line[(j+5)],collapse=" ")) # print into a file
}# end of loop go through the line
}#end of loop go through the list
match=grep("(\\w+\\W\\w+)(\\Wfalciforme\\W)(\\w+\\W\\w+)",list.frames,perl=T,value=T)
target=gsub("(\\w+\\W\\w+(\\W\\w+\\W)\\w+\\W\\w+", "\\1", match, perl=T)
A1=gsub("(\\w+\\W(\\w+)\\W\\w+\\W\\w+\\W\\w+", "\\1", match, perl=T)
A2=gsub("(\\w+)\\W\\w+\\W\\w+\\W\\w+\\W\\w+", "\\1", match, perl=T)
D1=gsub("(\\w+\\W\\w+\\W\\w+\\W(\\w+)\\W\\w+", "\\1", match, perl=T)
D2=gsub("(\\w+\\W\\w+\\W\\w+\\W\\w+\\W(\\w+)", "\\1", match, perl=T)
concordance=data.frame(A2,A1,target,D1,D2)
concordance
```

No Exemplo 2, a busca por “doença”, ambos destacados em negritos apenas para fins de visualização.

### Exemplo 2:

```
x=scan(file=file.choose(),what="char",sep="\n",skip=0,quiet=F) # sentences
x=gsub("[.,:;()]", "", x, ignore.case=T, perl=T)
x=gsub("[%]", "por_cento", x, ignore.case=T, perl=T)
x=gsub(" ", " ", x, perl=T) # this gets rid of EXTRA spaces
y=strsplit(x, " ")
data.frame(table(unlist(y)))
list.frames=c()
for(i in 1:length(y)){# go through the list
current.line=y[[i]] # access each line
for(j in 1:(length(current.line)-1)){#go up to the antipenultimate words, thus -1
list.frames=append(list.frames,paste(current.line[j],current.line[(j+1)],current.line[(j+2)],current.line[(j+3)],current.line[(j+5)],collapse=" ")) # print into a file
}# end of loop go through the line
}#end of loop go through the list
match=grep("(\\w+\\W\\w+)(\\Wdoença\\W)(\\w+\\W\\w+)",list.frames,perl=T,value=T)
target=gsub("(\\w+\\W\\w+(\\W\\w+\\W)\\w+\\W\\w+", "\\1", match, perl=T)
A1=gsub("(\\w+\\W(\\w+)\\W\\w+\\W\\w+\\W\\w+", "\\1", match, perl=T)
A2=gsub("(\\w+)\\W\\w+\\W\\w+\\W\\w+\\W\\w+", "\\1", match, perl=T)
D1=gsub("(\\w+\\W\\w+\\W\\w+\\W(\\w+)\\W\\w+", "\\1", match, perl=T)
D2=gsub("(\\w+\\W\\w+\\W\\w+\\W\\w+\\W(\\w+)", "\\1", match, perl=T)
concordance=data.frame(A2,A1,target,D1,D2)
concordance
```

Ao rodar o *script* anterior, obtêm-se excertos do texto no qual tais nódulos ocorrem, conforme se pode verificar na FIG. 27, a seguir.



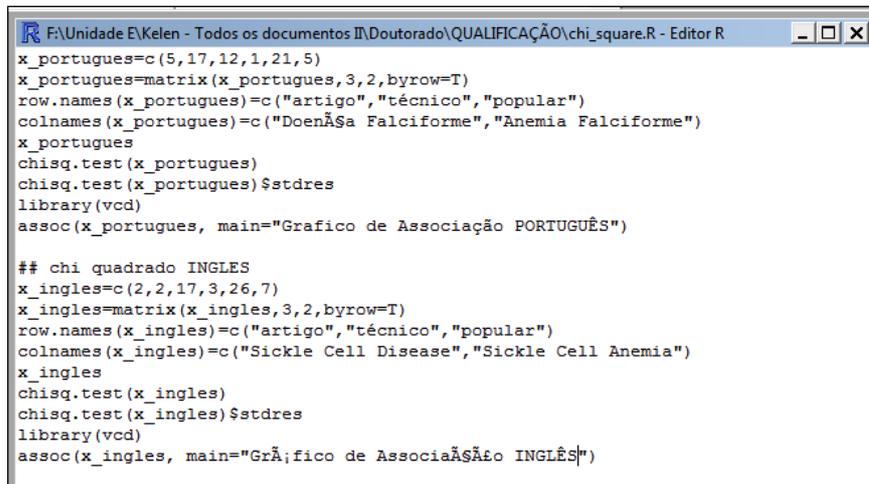
```
RGui (32-bit)
Arquivo Editar Visualizar Misc Pacotes Janelas Ajuda

R Console
+ for(j in 1:(length(current.line)-1)){ #go up to the antipenultimate words, thus -1
+ list.frames=append(list.frames,paste(current.line[j],current.line[(j+1)],current.line[(j+2)],current.line[(j+3)],current.line[(j+5)],collapse=" ")) # print into a file
+ }# end of loop go through the line
+ }#end of loop go through the list
> match=grep("(\\w+\\W\\w+)(\\Wdoença\\W)(\\w+\\W\\w+)",list.frames,perl=T,value=T)
> target=gsub("(\\w+\\W\\w+(\\W\\w+\\W)\\w+\\W\\w+", "\\1", match, perl=T)
> A1=gsub("(\\w+\\W(\\w+)\\W\\w+\\W\\w+\\W\\w+", "\\1", match, perl=T)
> A2=gsub("(\\w+)\\W\\w+\\W\\w+\\W\\w+\\W\\w+", "\\1", match, perl=T)
> D1=gsub("(\\w+\\W\\w+\\W\\w+\\W(\\w+)\\W\\w+", "\\1", match, perl=T)
> D2=gsub("(\\w+\\W\\w+\\W\\w+\\W\\w+\\W(\\w+)", "\\1", match, perl=T)
> concordance=data.frame(A2,A1,target,D1,D2)
> concordance
  A2      A1      target      D1      D2
1  do traço falciforme      para trinta
2  a anemia falciforme      hbaa hemoglobinopatia
3  no traço falciforme      hbaa produção
4  diagnóstico anemia falciforme hemoglobinopatia e
5  por doente falciforme      no do
6  o doente falciforme      no falce
7  do traço falciforme      se como
8  com anemia falciforme      de cidades
9  anemia falciforme      recorrem apenas
10 da doença falciforme      NA NA
11 com doença falciforme      e dolorosas
12 da doença falciforme      NA NA
13 da doença falciforme      NA NA
14 com doença falciforme      precisa aspectos
15 crise dolorosa falciforme      as dos
16 com dor falciforme      deve estimulada
17 com anemia falciforme      s nascidos
18 da anemia falciforme      NA NA
19 da anemia falciforme      NA NA
20 da anemia falciforme      são biliar
21 da anemia falciforme      é na
22 da anemia falciforme      é produto
23 pela anemia falciforme      são por
24 com anemia falciforme      provenientes hospital
25 anemia falciforme      pareadas sexo
26 com anemia falciforme      no meio
27 com anemia falciforme      NA NA
28 da anemia falciforme      sua presença
29 da anemia falciforme      pode para
```

FIGURA 27. Localização do item lexical no *corpus*

Após rodar o *script*, é possível identificar as sentenças em que os nódulos ocorrem.

Para verificar se existe associação estatisticamente significativa entre os nódulos, a língua e os tipos dos tipos de textos, foi aplicado o teste de associação qui-quadrado ( $\chi^2$ ), conforme ilustrado na FIG. 28, a seguir.



```

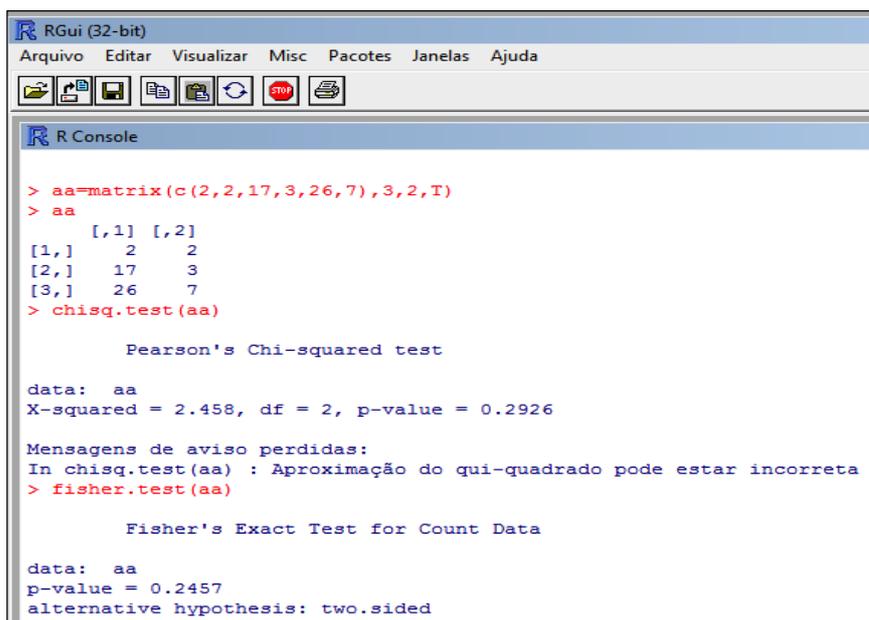
R F:\Unidade E\Kelen - Todos os documentos IX\Doutorado\QUALIFICAÇÃO\chi_square.R - Editor R
x_portugues=c(5,17,12,1,21,5)
x_portugues=matrix(x_portugues,3,2,byrow=T)
row.names(x_portugues)=c("artigo","técnico","popular")
colnames(x_portugues)=c("Doença Falciforme","Anemia Falciforme")
x_portugues
chisq.test(x_portugues)
chisq.test(x_portugues)$stdres
library(vcd)
assoc(x_portugues, main="Grafico de Associação PORTUGUÊS")

## chi quadrado INGLES
x_ingles=c(2,2,17,3,26,7)
x_ingles=matrix(x_ingles,3,2,byrow=T)
row.names(x_ingles)=c("artigo","técnico","popular")
colnames(x_ingles)=c("Sickle Cell Disease","Sickle Cell Anemia")
x_ingles
chisq.test(x_ingles)
chisq.test(x_ingles)$stdres
library(vcd)
assoc(x_ingles, main="Gráfico de Associação INGLÊS")

```

FIGURA 28. *Script* para aplicação do teste  $\chi^2$

A FIG. 29 ilustra o resultado obtido após aplicar o *script* utilizado para o cálculo do  $\chi^2$ .



```

RGui (32-bit)
Arquivo Editar Visualizar Misc Pacotes Janelas Ajuda
R Console
> aa=matrix(c(2,2,17,3,26,7),3,2,T)
> aa
      [,1] [,2]
[1,]    2    2
[2,]   17    3
[3,]   26    7
> chisq.test(aa)

      Pearson's Chi-squared test

data: aa
X-squared = 2.458, df = 2, p-value = 0.2926

Mensagens de aviso perdidas:
In chisq.test(aa) : Aproximação do qui-quadrado pode estar incorreta
> fisher.test(aa)

      Fisher's Exact Test for Count Data

data: aa
p-value = 0.2457
alternative hypothesis: two.sided

```

FIGURA 29. Resultado da aplicação do teste  $\chi^2$

A seguir estão relacionados os outros *scripts* programados para a extração dos dados obtidos a partir do etiquetamento morfossintático dos *corpora* por meio do programa Treetagger.

#### 2.1.2.2 Anotação dos *corpora* em classes de palavras: etiquetador morfossintático Treetagger

O Treetagger faz uso de um modelo probabilístico a partir de trigramas, ou seja, a ferramenta capta sequências de três palavras encontradas no *corpus* e estabelece uma relação entre as classes gramaticais (SCHMID, 1994). Para anotar as palavras do texto em inglês, o etiquetador possui um léxico desenvolvido a partir de uma parte do *Penn Treebank Corpus*<sup>17</sup> – parte essa composta por dois milhões de palavras etiquetadas. Para a anotação dos textos em português brasileiro, o Treetagger utiliza o conjunto de etiquetas proposto por pesquisadores da Universidade de Santiago da Compostela, Espanha<sup>18</sup>. Embora os pesquisadores da Espanha utilizem a análise morfossintática para o português europeu e galego, não há diferença na classificação das classes das palavras para o português brasileiro, não havendo, portanto, comprometimento na validade dos dados desta pesquisa. De acordo com Schmid (1994), o Treetagger alcança em torno de 96% de acurácia, sendo considerado bem eficiente, para etiquetar tanto os textos em português quanto os textos em inglês<sup>19</sup>.

A seguir, o QUADRO 6 mostra todas as etiquetas (*tagsets*), para a anotação dos textos em inglês e em português, identificadas pelo etiquetador morfossintático Treetagger.

---

<sup>17</sup> O Penn Treebank Corpus, desenvolvido por pesquisadores da Universidade da Pensilvânia, Estados Unidos está disponível na página dessa universidade. Disponível em: <<http://www.cis.upenn.edu/~treebank/>>. Último acesso em: 25 jan. 2013.

<sup>18</sup> Para mais informações sobre o Treetagger em português, basta acessar a página específica, disponível em: <[www.gramatica.usc.es/~gamallo/](http://www.gramatica.usc.es/~gamallo/)>, último acesso em 25/01/2013.

<sup>19</sup> Vale lembrar que as etiquetas seguem a categorização da gramática tradicional, pois não há conjunto de etiquetas organizadas sob a ótica da gramática sistêmico-funcional.

QUADRO 6. Etiquetas (*tagset*) para anotação dos textos em inglês e em português identificadas pelo programa Treetagger (continua)

<i>Tagset do inglês</i>		<i>Tagset do português</i>	
Etiqueta	Descrição	Etiqueta	Descrição
CC	Conjunção	ADJ	Adjetivo
CD	Número cardinal	ADV	Advérbio
DT	Determinante	CARD	Número cardinal/ordinal
EX	<i>There</i> existencial	CONJ	Conjunção
FW	Termo estrangeiro	DET	Determinante
IN	Preposição e conjunção subordinada	NOM	Substantivo/Nome próprio
IN/that	<i>That</i> subordinado	P	Pronome
JJ	Adjetivo	PR	Pronome relativo
JJR	Adjetivo comparativo	PRP	Preposição
JJS	Adjetivo superlativo	PRP + DET	Preposição seguida de determinante
LS	Marcador de lista	V	Verbo
MD	Verbo modal	V + P	Verbo seguido de pronome
NN	Substantivo singular ou incontável	I	Interjeição
NNS	Substantivo plural	VIRG	Separadores dentro da oração
NP	Nome próprio singular	SENT	Separadores entre orações
NPS	Nome próprio plural	-	-
PDT	Predeterminante	-	-
POS	Apóstrofo + s (possessivo)	-	-
PP	Pronome pessoal	-	-
PP\$	Pronome possessivo	-	-
RB	Advérbio	-	-
RBR	Advérbio comparativo	-	-
RBS	Advérbio superlativo	-	-
RP	Partícula verbal/adverbial	-	-
SENT	Separador de oração	-	-
SYM	Símbolo	-	-
TO	<i>to</i> precedente a verbo no infinitivo	-	-
UH	Interjeição	-	-
VB	Verbo <i>to be</i> : forma básica	-	-
VBD	Verbo <i>to be</i> : pretérito	-	-
VBG	Verbo <i>to be</i> : gerúndio/particípio presente	-	-
VBN	Verbo <i>to be</i> : particípio passado	-	-
VBP	Verbo <i>to be</i> : primeira pessoa do singular e segunda pessoa do singular e do plural	-	-
VBZ	Verbo <i>to be</i> : terceira pessoa do singular	-	-
VH	Verbo <i>to have</i> : forma básica	-	-
VHD	Verbo <i>to have</i> : pretérito	-	-
VHG	Verbo <i>to have</i> : gerúndio/particípio presente	-	-
VHN	Verbo <i>to have</i> : particípio passado	-	-
VHP	Verbo <i>to have</i> : primeira pessoa do singular e segunda pessoa do singular e do plural	-	-
VHZ	Verbo <i>to have</i> : terceira pessoa do singular	-	-
VV	Verbo: forma básica	-	-
VVD	Verbo: pretérito	-	-
VVG	Verbo: gerúndio/particípio presente	-	-
VVN	Verbo: particípio passado	-	-

QUADRO 6. Etiquetas (*tagset*) para anotação dos textos em inglês e em português identificadas pelo programa Treetagger (continuação)

<i>Tagset</i> do inglês		<i>Tagset</i> do português	
Etiqueta	Descrição	Etiqueta	Descrição
VVP	Verbo: primeira pessoa do singular e segunda pessoa do singular e plural	-	-
VVZ	Verbo: terceira pessoa do singular	-	-
WDT	Pronome determinante iniciado por <i>wh</i> <sup>20</sup>	-	-
WP	Pronome iniciado por <i>wh</i> <sup>21</sup>	-	-
WP\$	Pronome possessivo iniciado por <i>wh</i> <sup>22</sup>	-	-
WRB	Advérbio iniciado por <i>wh</i> <sup>23</sup>	-	-
#	#	-	-
\$	Cifrão	-	-
“	Aspas	-	-
(	Abre parênteses	-	-
)	Fecha parênteses	-	-
,	Vírgula	-	-
:	Separador dentro de oração	-	-

Fonte: Tradução e organização de Nunes (2013).<sup>24</sup>

Retomando o QUADRO 6, nota-se que há uma diferença entre a especificidade e a descrição das etiquetas quando o *tagset* da língua portuguesa é comparado com aquele da língua inglesa. Os *tagsets* da língua inglesa e os da língua portuguesa utilizados foram ilustrados no QUADRO 6. Esses *tagsets* são todos aqueles reconhecidos pela ferramenta Treetagger no momento do processamento automático para etiquetar morfossintaticamente os textos em inglês e em português, conforme ilustrado na FIG.30, a seguir.

<sup>20</sup> E.g., “which”.

<sup>21</sup> E.g., “who” e “what”.

<sup>22</sup> E.g., “whose”.

<sup>23</sup> E.g., “when” e “where”.

<sup>24</sup> Os *tagsets* em inglês estão disponíveis em: <<http://www.cis.upenn.edu/~treebank/>>. Último acesso em: 25 jan. 2013. Trata-se de um conjunto de etiquetas morfossintáticas que atribuem a cada palavra dos *corpus* uma determinada classe de palavras.

<p>gramatica.usc.es/~gamallo/php/tagger/Tagg</p> <p>Envia ao tagger</p> <p><b>Análise morfo-sintática:</b></p> <p>a DET a  doença NOM doença  falciforme ADJ falciforme  é V ser  um DET um  dos PRP+DET de  distúrbios NOM distúrbio  genéticos ADJ genético  mais ADV mais  frequentes V frequentar  no PRP+DET em  brasil NOM brasil  e CONJ e  no PRP+DET em</p>	<p>web4u.setsunan.ac.jp/Website/cmd/tt_win.cgi</p> <table border="0"> <tr><td>sickle</td><td>NN</td><td>sickle</td><td></td></tr> <tr><td>cell</td><td>NN</td><td>cell</td><td></td></tr> <tr><td>disease</td><td>NN</td><td>disease</td><td></td></tr> <tr><td>is</td><td>VBZ</td><td>be</td><td></td></tr> <tr><td>a</td><td>DT</td><td>a</td><td></td></tr> <tr><td>term</td><td>NN</td><td>term</td><td></td></tr> <tr><td>covering</td><td>VVG</td><td>cover</td><td></td></tr> <tr><td>a</td><td>DT</td><td>a</td><td></td></tr> <tr><td>number</td><td>NN</td><td>number</td><td></td></tr> <tr><td>of</td><td>IN</td><td>of</td><td></td></tr> <tr><td>different</td><td>JJ</td><td>different</td><td></td></tr> <tr><td>but</td><td>CC</td><td>but</td><td></td></tr> <tr><td>similar</td><td>JJ</td><td>similar</td><td></td></tr> <tr><td>conditions</td><td>NNS</td><td>condition</td><td></td></tr> <tr><td>that</td><td>WDT</td><td>that</td><td></td></tr> <tr><td>affect</td><td>VVP</td><td>affect</td><td></td></tr> <tr><td>haemoglobin</td><td>NN</td><td>haemoglobin</td><td></td></tr> <tr><td>haemoglobin</td><td>NN</td><td>haemoglobin</td><td></td></tr> </table>	sickle	NN	sickle		cell	NN	cell		disease	NN	disease		is	VBZ	be		a	DT	a		term	NN	term		covering	VVG	cover		a	DT	a		number	NN	number		of	IN	of		different	JJ	different		but	CC	but		similar	JJ	similar		conditions	NNS	condition		that	WDT	that		affect	VVP	affect		haemoglobin	NN	haemoglobin		haemoglobin	NN	haemoglobin	
sickle	NN	sickle																																																																							
cell	NN	cell																																																																							
disease	NN	disease																																																																							
is	VBZ	be																																																																							
a	DT	a																																																																							
term	NN	term																																																																							
covering	VVG	cover																																																																							
a	DT	a																																																																							
number	NN	number																																																																							
of	IN	of																																																																							
different	JJ	different																																																																							
but	CC	but																																																																							
similar	JJ	similar																																																																							
conditions	NNS	condition																																																																							
that	WDT	that																																																																							
affect	VVP	affect																																																																							
haemoglobin	NN	haemoglobin																																																																							
haemoglobin	NN	haemoglobin																																																																							

FIGURA 30. *Corpora* etiquetados por meio do etiquetador automático Treetagger

Na FIG. 30, há uma parte de um *corpus* em português e de um em inglês, ambos do artigo acadêmico. O excerto de cada um dos *corpora* é exibido em três colunas: na primeira, estão as palavras conforme aparecem no texto; na segunda, as etiquetas morfológicas utilizadas como parâmetros para a classificação das classes de palavras; na terceira, a forma canônica da palavra, como o infinitivo do verbo e o substantivo no singular e sem a identificação de feminino ou masculino. Para fins de comparação entre os dados anotados a partir do Treetagger, os *tagsets* em inglês e em português precisaram ser compilados de forma tal que atendessem à finalidade da pesquisa em tela. Para tanto, foi necessária a organização desses *tagsets* de modo que ficassem evidentes as categorias gramaticais comuns entre as duas línguas e que fosse possível subsidiar a categorização dos registros à luz da Linguística Sistêmico-Funcional, com o aporte do ambiente de programação R. Consequentemente, algumas etiquetas do tagset do inglês foram descartadas (*i.e.*, EX, LS, POS, RP, SENT, SYM e LS), e os símbolos # e \$ e os elementos de pontuação foram eliminados, por não serem classificados como classe de palavra, além de não serem lidos pelo ambiente de programação R. No que tange ao *tagset* da língua portuguesa, todos os sinais de pontuação também foram descartados, em virtude do ambiente de programação R (*cf.* NUNES, 2010).

O QUADRO 7 ilustra o agrupamento dos dois *tagsets* (inglês e português), que contemplou as dez classes de palavras: adjetivo, advérbio, conjunção, determinante, interjeição, numeral, preposição, pronome, substantivo e verbo.

QUADRO 7. Agrupamento das etiquetas dos tagsets do inglês e do português

Classe de palavra	Etiquetas do inglês	Etiquetas do Português
Adjetivo	JJ JJR JJS	ADJ
Advérbio	RB RBR RBS WRB	ADV
Conjunção	CC	CONJ
Determinante	DT	DET
Interjeição	UH	I
Numeral (cardinal e ordinal)	CD	CARD
Preposição	IN IN/that TO	PRP PRP+DET
Pronome	PP PP\$ WDT WP WP\$	P PR
Substantivo	FW NN NNS NP NPS	NOM
Verbo	MD VB VBD VBG VBN VBP VBZ VH VHD VHG VHN VHP VHZ VV VVD VVG VVN VVP VVZ	V V+P

Fonte: organizado por Nunes (2013).

Todos os seis *corpora* deste estudo foram etiquetados com base nos *tagsets* do inglês e português, conforme agrupamento das etiquetas para as duas línguas (cf. QUADRO 7), com vistas a:

1. Anotar, por meio do Treetagger, as classes de palavras de cada um dos tipos de texto; e
2. Mapear itens lexicais e itens gramaticais.

Conforme explicado no capítulo anterior, para a anotação dos *corpora* por meio do etiquetador automático Treetagger, foram criados outros comandos específicos para o ambiente de programação R. Cumpre ressaltar que a anotação dos *corpora* por meio do etiquetador morfossintático gera dados passíveis de serem armazenados em planilhas do Excel em formato .CSV, tipo de arquivo lido pelo ambiente de programação R. A seguir estão listados os comandos desenvolvidos para obtenção dos dados referentes às classes de palavras, etiquetadas a partir do Treetagger:

1. Extração de dados de cada uma das seis planilhas (três tipos de texto e duas línguas) referentes às classes de palavras;
2. Extração da frequência absoluta de *tokens* e *types* de cada classe gramatical por tipo de texto e por língua;
3. Extração da frequência absoluta de *tokens* e *types* das classes de palavras “substantivo”, “conjunção” e “pronome” por tipo de texto e por língua por essas classes apresentarem importância para a caracterização dos tipos de texto com o aporte da Linguística Sistemico-Funcional; e
4. Aplicação do teste de significância estatística  $\chi^2$ .

A seguir estão descritos os procedimentos de coleta por meio do programa UAM CorpusTool.

#### 2.1.2.3 Anotação e etiquetamento dos *corpora* por meio do UAM CorpusTool

O UAM CorpusTool foi desenvolvido por O’Donnell (2008) com a finalidade de anotar tanto textos quanto imagens. Para este estudo, foram selecionados excertos dos textos que compuseram os *corpora* com mil palavras (Anexo 3), utilizando-se a fórmula “randbetween” do programa Microsoft Excel© para selecionar uma página aleatória dentro de um intervalo entre o número da primeira página e o número da última página de cada um dos *corpora*.

Após a seleção aleatória do número da página, foram contadas aproximadamente mil palavras correntes de cada um dos três tipos de texto, em inglês e português (*cf.* BIBER, 1990), certificando-se de que foram selecionados parágrafos completos (*cf.* NUNES, 2010). Assim, os textos selecionados foram copiados e colados em um novo arquivo .txt, para que pudessem ser lidos e armazenados pelo programa UAM CorpusTool, passando a ser considerados os *corpora* para esta etapa de análise.

Após a inserção dos textos no programa, foi possível criar os esquemas necessários para a análise de cada um deles. Para que esses esquemas sejam gerados, é necessária intervenção manual, observando-se até qual nível da delicadeza da teoria sistêmico-funcional se pretende analisar. No caso deste estudo, os esquemas foram gerados de acordo com os sistemas de TRANSITIVIDADE (processos), MODO (Modo e função discursiva), TEMA (complexidade e seleção temática) (HALLIDAY; MATTHIESSEN, 2004) e MENSAGEM (contextualização, arranjo e focalização) (MARTIN; ROSE, 2007) de modo a contemplarem a descrição do perfil metafuncional dos textos. A seguir, cada um dos esquemas criados para anotação semiautomática dos textos foi ilustrado de acordo com os quatro sistemas, quais sejam: 1) TRANSITIVIDADE; 2) MODO e 3) TEMA e 4) MENSAGEM:

#### 1) TRANSITIVIDADE

Conforme explicado anteriormente, a metafunção ideacional (experiencial) é realizada pela TRANSITIVIDADE por meio de processos, participantes e circunstâncias (*cf.* Subseção 1.3.1). Assim, por ser o processo o elemento central em uma figura, os esquemas de análise foram criados de acordo com os tipos de processos: material, relacional (atributivo e identificativo), mental, existencial, comportamental e existencial.

A FIG. 31 mostra o esquema do sistema gramatical de TRANSITIVIDADE criado no programa UAM CorpusTool.

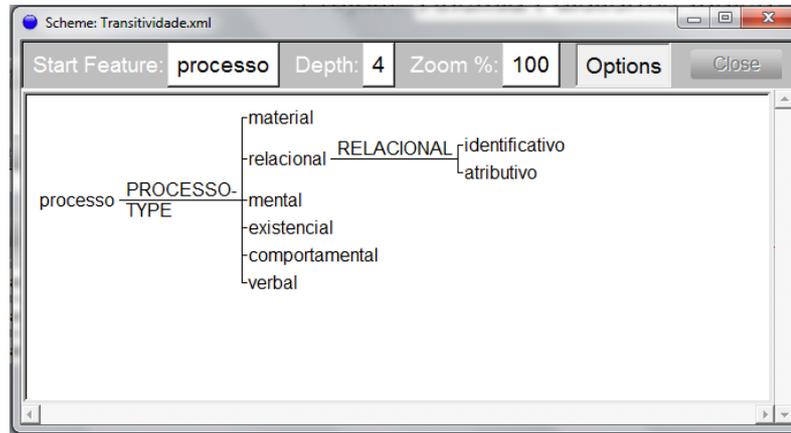


FIGURA 31. Exemplo do Esquema do Sistema Gramatical de TRANSITIVIDADE no UAM CorpusTool

## 2) MODO

A metafunção interpessoal estabelece a relação entre autor e leitor e é realizada pelo sistema de MODO. Neste estudo, para ilustrar a interação especialista-especialista (artigo acadêmico), especialista-técnico da área da saúde (manuais técnicos) e especialista-leigo (cartilhas e *folders* de divulgação) foram criados esquemas de análise de acordo com:

1. Funções discursivas, classificadas em: (i) turno (turnos inicial ou respondente); (ii) papel do falante (inicial ou respondente) e (iii) mercadoria (informações de bens e serviços);
2. Modo: classificado em Indicativo Declarativo, Indicativo Interrogativo ou Imperativo.

A FIG. 32 mostra o esquema do sistema gramatical de MODO criado no programa UAM CorpusTool®:

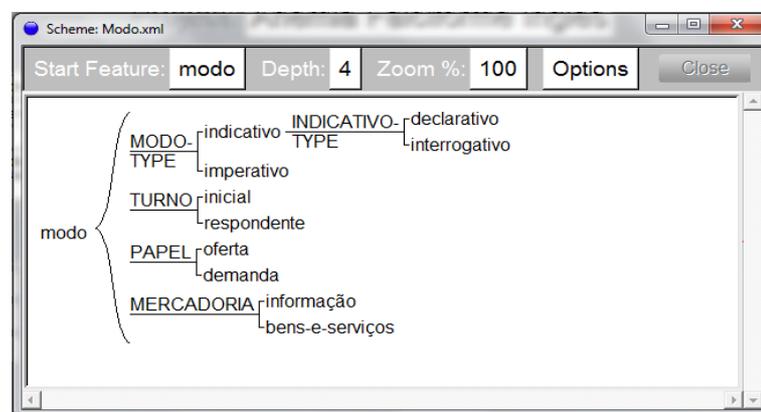


FIGURA 32. Exemplo do Esquema do Sistema Gramatical de MODO no UAM CorpusTool

### 3) TEMA e MENSAGEM

A metafunção textual é a responsável por organizar as metafunções ideacional e interpessoal ao longo do texto. Para analisar essa organização, foram criados três esquemas com base nos sistemas de TEMA do inglês (HALLIDAY; MATTHIESSEN, 2004) e do português (FIGUEREDO, 2011) e, também, no sistema de MENSAGEM (MARTIN; ROSE, 2007).

O esquema com base no sistema de TEMA em inglês foi criado de acordo com:

1. Complexidade: dividida em temas simples e múltiplo, sendo o múltiplo subdividido em: (i) textual e tópico; (ii) interpessoal e tópico; ou (iii) textual, interpessoal e tópico; e
2. Seleção temática: subdividida em tema não marcado e tema marcado.

A FIG. 33 mostra o esquema do sistema gramatical de TEMA em inglês criado no programa UAM CorpusTool.

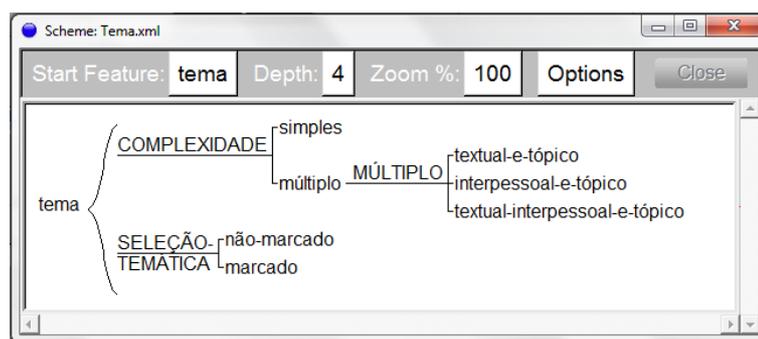


FIGURA 33. Exemplo do Esquema do Sistema Gramatical de TEMA em inglês no UAM CorpusTool

Já o sistema de TEMA em português foi criado com base na:

1. Complexidade: subdividida em temas elemental, ângulo, textual, interpessoal e *tema-default*; e
2. Seleção temática: subdividida em temas *default* e proeminente.

A FIG. 34 ilustra o esquema do sistema gramatical de TEMA em português criado no programa UAM CorpusTool.

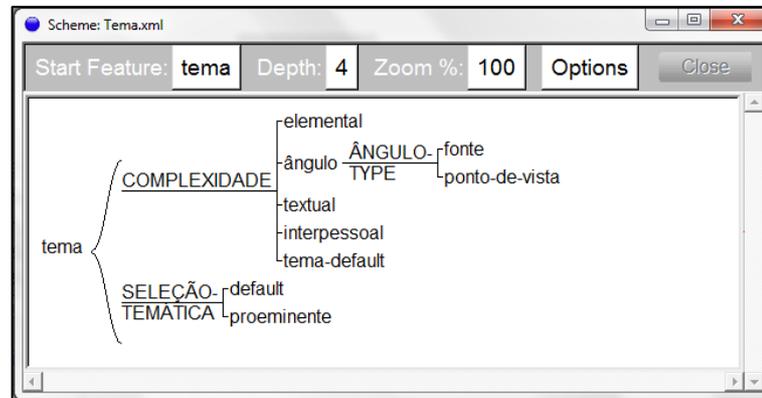


FIGURA 34. Exemplo do Esquema do Sistema Gramatical de TEMA em português no UAM CorpusTool

Por fim, o esquema para a análise do sistema de MENSAGEM foi criado a partir dos seguintes parâmetros:

1. Contextualização: dividida em inicial e fase. A fase se subdivide em: (i) continuidade e (ii) descontinuidade (desvio ou mudança);
2. Arranjo: subdividido em arranjo-*default* e arranjo (explícito ou implícito); e
3. Focalização: focalizado (particularizado ou contraste) e não focalizado.

A FIG. 35 apresenta o esquema do sistema gramatical de MENSAGEM criado no programa UAM CorpusTool.

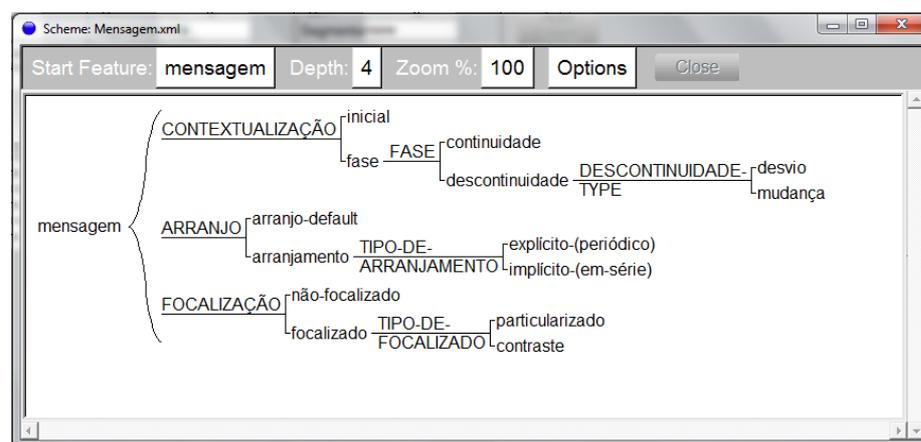


FIGURA 35. Exemplo do Esquema do Sistema Gramatical de TEMA em português no UAM CorpusTool

Após a criação de todos esses esquemas, iniciaram-se os procedimentos para a anotação semiautomática a partir dos *corpora* já inseridos no programa. Para a anotação semiautomática dos dados, foram organizados dois projetos, sendo um específico para os

textos em língua portuguesa e o outro para os textos em língua inglesa. Cada um dos projetos recebeu um título de acordo com a língua dos textos. A FIG. 36 exemplifica a tela de um dos projetos criado no UAM CorpusTool, intitulado “Project Anemia Falciforme Inglês”. Para fins de visualização, os esquemas com base nos sistemas de TRANSITIVIDADE, MODO, e TEMA e MENSAGEM estão destacados na cor vermelha; os *corpora* prontos para a análise estão destacados na cor azul.

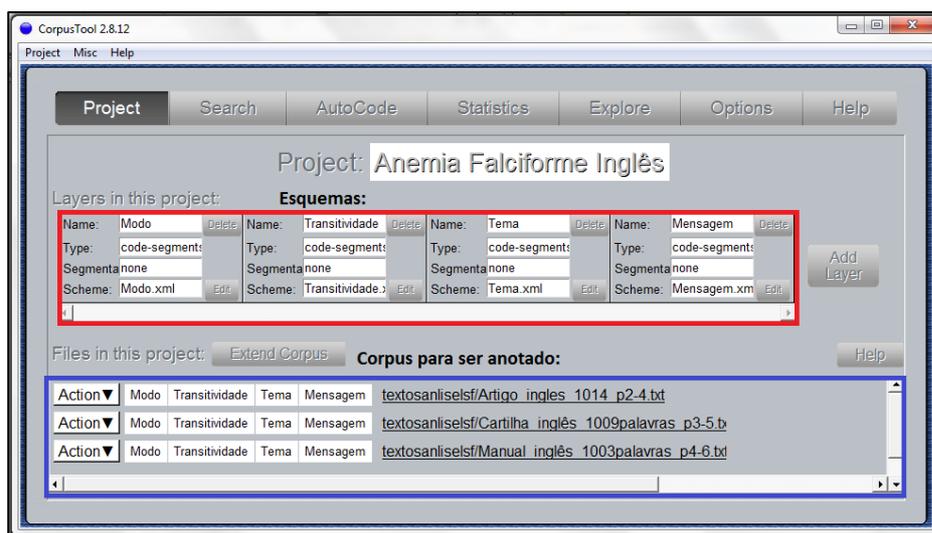


FIGURA 36. Exemplo de tela de projeto do UAM CorpusTool

Em seguida, os seguimentos dos textos foram anotados de acordo com cada um dos sistemas.

A FIG. 37 mostra um exemplo de análise feita no programa UAM CorpusTool.

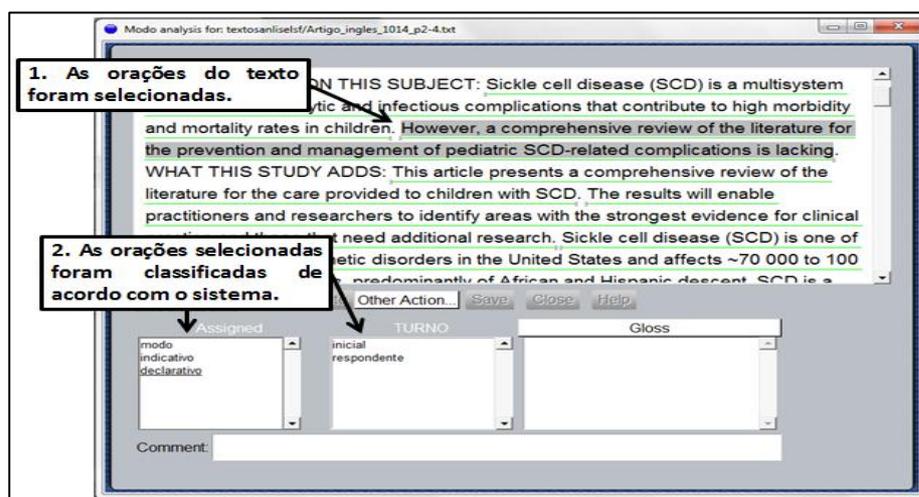


FIGURA 37. Exemplo de análise realizada no UAM CorpusTool por meio de esquema previamente criado

Terminadas as análises, os dados estatísticos de cada um dos tipos de textos foram extraídos de forma comparada automaticamente, como apresenta a FIG. 38.

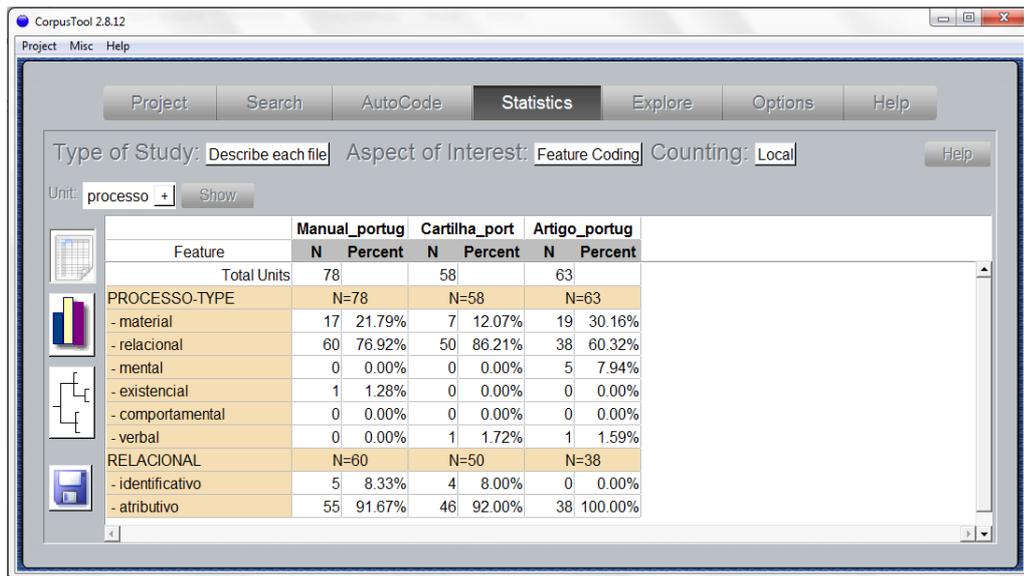


FIGURA 38. Exemplo da extração dos dados estatísticos dos tipos de textos de forma comparada UAM CorpusTool

Para ilustrar graficamente os dados extraídos (transformá-los em tabelas, gráficos ou dendrogramas), foram atribuídas cores às funções dos sistemas, ou seja, MODO = X, cor azul; TRANSITIVIDADE = Y, cor vermelha; TEMA = Z, cor verde e MENSAGEM = W, cor lilás.

Além disso, foram atribuídos valores às categorias (1,2, 3, 4 etc.). Esses valores são correlacionados de acordo com a disposição topológica de cada categoria no sistema linguístico, conforme ilustrado na FIG. 39.

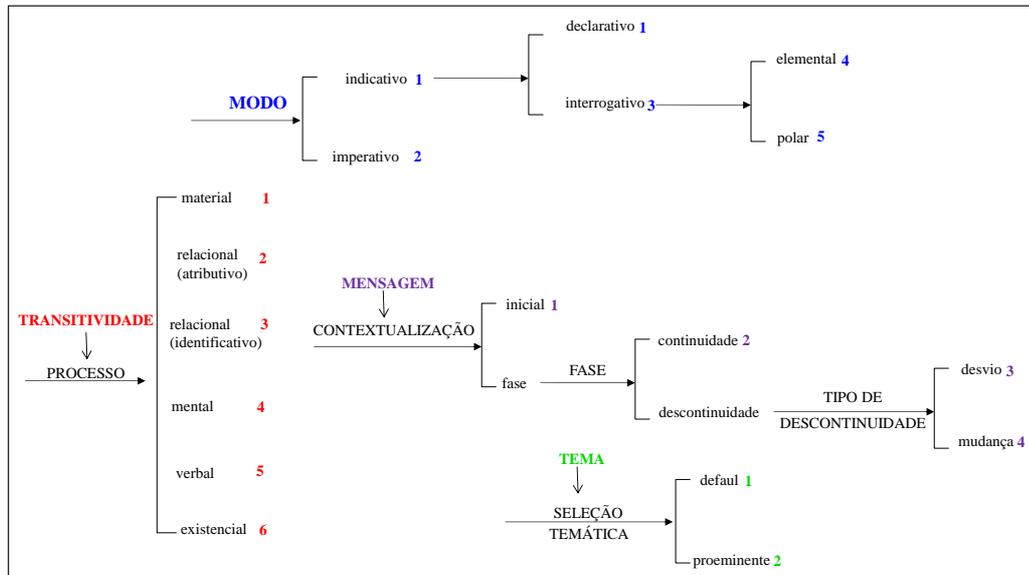


FIGURA 39. Disposição topológica das categorias

Os dados obtidos após as análises por meio dos esquemas foram organizados em quadros (cf. QUADRO 8).

QUADRO 8. Exemplo de tabela gerada a partir dos dados extraídos pelo UAM CorpusTool

Oração	Modo X	Transitividade Y	Tema Z	Mensagem W
1	1	3	1	1
2	1	2	1	2
3	2	1	2	4
4	1	1	1	2
5	1	2	1	2

A parte destacada no QUADRO 8 representa a terceira oração do texto do tipo cartilha e *folder* de divulgação:<sup>25</sup> “*For more information about beta thalassaemia contact the Thalassaemia Society or one of the specialist centres*”. A classificação dessa oração pode ser descrita de acordo com os dados da FIG. 39, ou seja, Modo X = 2 (imperativo); Transitividade Y = 1 (processo material); Tema Z = 2 (simples marcado) e Mensagem W = 4 (fase descontinuidade:desvio). Para a descrição dos dados, foram criadas duas legendas que especificam cada uma das categorias de acordo com as funções do sistema que foram evidenciadas durante a análise dos *corpora* desta pesquisa. Essas legendas foram organizadas de acordo com os dados que foram passíveis de análise nos textos em português e nos textos em inglês:

<sup>25</sup> Os exemplos desta seção foram selecionados de forma aleatória, apenas no intuito de ilustrar os dados obtidos por meio do programa UAM CorpusTool.

## 1) Legenda para os textos em português (legenda inserida também no Anexo 2)

**Modo**

- 1 – Modo Indicativo Declarativo - Oferta de Informação
- 2 – Imperativo - Demanda de Bens e Serviços

**Transitividade**

- 1 – Processo Material
- 2 – Processo Relacional Atributivo
- 3 – Processo Relacional Identificativo
- 4 – Processo Mental
- 5 – Processo Verbal
- 6 – Processo Existencial

**Tema**

- 1 – Tema *Default - Default*
- 2 – Ângulo: Fonte - *Default*
- 3 – Textual - *Default*
- 4 – Interpessoal - *Default*

**Mensagem**

- 1 – Inicial Arrançamento Explícito Não Focalizado
- 2 – Fase Continuidade Arranjo-*Default* Não Focalizado
- 3 – Fase Descontinuidade Mudança Arranjo *Default* Não Focalizado
- 4 – Fase Descontinuidade Desvio Arranjo *Default* Não Focalizado
- 5 – Inicial Arrançamento Implícito Não Focalizado

## 2) Legenda para os textos em inglês(legenda inserida também no Anexo 2)

**Modo**

- 1 – Modo Indicativo Declarativo - Oferta de Informação
- 2 – Imperativo - Demanda de Bens e Serviços

**Transitividade**

- 1 – Processo Material
- 2 – Processo Relacional Atributivo
- 3 – Processo Relacional Identificativo
- 4 – Processo Existencial

**Tema**

- 1 – Simple – Não Marcado
- 2 – Simple – Marcado
- 3 – Múltiplo – Textual e Tópico – Não Marcado
- 4 – Múltiplo – Textual e Tópico – Marcado
- 5 – Múltiplo – Interpessoal e Tópico – Não Marcado

**Mensagem**

- 1 – Inicial Arrançamento Explícito Não Focalizado
- 2 – Fase Continuidade Arranjo-*Default* Não Focalizado
- 3 – Fase Descontinuidade Mudança Arranjo *Default* Não Focalizado
- 4 – Fase Descontinuidade Desvio Arranjo *Default* Não Focalizado

Por fim, essas tabelas possibilitaram organizar os dados de cada um dos tipos de texto para que fossem transformados em gráficos tridimensionais (*scatterplot 3D*) e dendrogramas a partir de *scripts* específicos (*cf.* Anexo 1) no ambiente de programação R, conforme ilustrado

na FIG. 40. Para a construção dos gráficos e dendrogramas, não foram utilizados os dados do sistema de MENSAGEM, apenas os dados dos obtidos por meio da análise dos sistemas de TRANSITIVIDADE, MODO e TEMA. A Mensagem faz parte do sistema de MENSAGEM que é um sistema semântico e não gramatical. Sendo assim, não foi considerada para a criação do gráfico tridimensional e nem para a criação do dendrograma. Ela foi analisada juntamente com o sistema de TEMA por ser responsável pela estrutura do tipo de texto e permitir entender as fases de cada um dos textos.

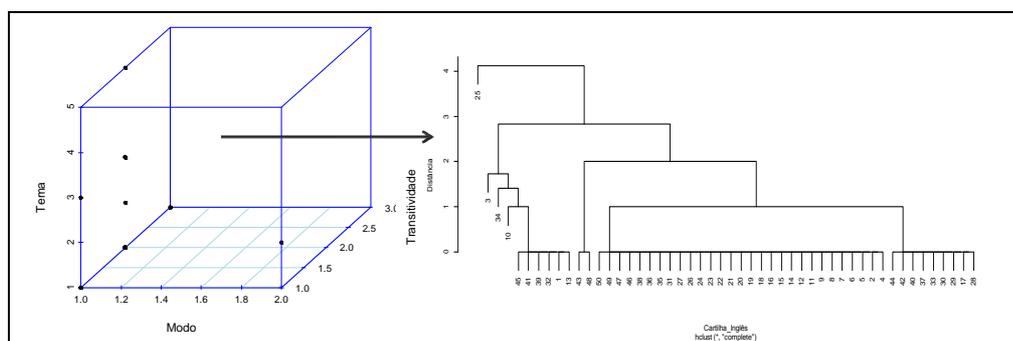


FIGURA 40. Gráfico tridimensional e dendrograma a partir dos dados obtidos pelo programa UAM CorpusTool

O diagrama tridimensional é um gráfico que apresenta os pontos no espaço cartesiano XYZ. Esses pontos são usados para representar simultaneamente os valores de três variáveis quantitativas em cada elemento do conjunto de dados (*cf.* valores categóricos na FIG. 39). No caso desta pesquisa, as variáveis são os dados provenientes dos sistemas de TRANSITIVIDADE, MODO e TEMA para fins de visualização da configuração do texto no contínuo de instanciação. Já os elementos do conjunto de dados são as orações que estão tabuladas nas tabelas extraídas a partir do programa UAM CorpusTool (Anexo 2).

Para uma melhor visualização da distribuição das orações, os dados categóricos foram organizados também em dendrogramas. Os pontos ilustrados nos gráficos coincidem, neste estudo, com o número total de agrupamentos criados a partir de cada texto. No caso da FIG. 40, por exemplo, há oito pontos no espaço e cada um desses pontos está relacionado com um agrupamento do dendrograma.

O dendrograma é uma estrutura de apresentação de dados que se assemelha a uma árvore cujos ramos podem formar grupos semelhantes (agrupamentos) – ou *clusters* – (GRIES, 2013) de acordo com o grau de similaridade entre as variáveis. As variáveis são dadas pelos

ramos da árvore, que é construída a partir de ramos individuais que se juntam de acordo com a distância Euclidiana<sup>26</sup>, de forma a agrupá-las em classes, até chegar à raiz da árvore. Toda vez que a sequência de um grupo é quebrada tem-se um indício de redução dessa similaridade. De acordo com Gries (2013, p. 338), a análise de *clusters* ou agrupamento aglomerativo hierárquico (*hierarchical agglomerative cluster analysis*) é usada para detectar padrões de semelhança entre os dados obtidos, mas não fornece resultados passíveis de interpretação, mas sim resultados exploratórios.

Para analisar o dendrograma, é necessário iniciar pela raiz e ir em direção às folhas. À medida que se move para a direita, aumentam-se as distâncias e agrupamentos cada vez maiores são formados. Esses agrupamentos apresentam as diversidades entre eles, ilustrando a divisão dos dados em grupos ou classes. Tanto os gráficos ou dendrogramas contribuíram para a caracterização dos tipos de textos analisados nesta pesquisa.

Na sequência, descrevem-se os procedimentos de análises dos dados extraídos dos *corpora* comparáveis desta pesquisa.

## 2.2 Metodologia de Análise dos Dados

Os procedimentos de análises também seguiram duas abordagens metodológicas complementares: (i) a Linguística de Corpus e (ii) a Linguística Sistêmico-Funcional. Essas duas abordagens enfatizam a língua em uso, são probabilísticas e focam-se nos textos.

A Linguística de Corpus serviu de embasamento para a abordagem metodológica deste estudo por garantir a exploração da linguagem verbal por meio de exemplos reais, possibilitando um estudo empírico das línguas. A Linguística de Corpus possibilita a extração dos dados dos *corpora* pelas ferramentas de análise por ela disponibilizadas (no caso desta pesquisa, o ambiente de programação R e programa UAM CorpusTool). Logo, para a pesquisa em tela, a LC propiciou a extração de dados empíricos dos *corpora* e a posterior análise de alguns desses dados com o aporte da Linguística Sistêmico-Funcional.

---

<sup>26</sup> A distância Euclidiana refere-se à distância espacial entre dois pontos, o que a torna um espaço métrico que pode ser medida por uma régua (GRIES, 2013, p. 342).

A Linguística Sistêmico-Funcional, responsável por explicar teoricamente os dados provenientes dos *corpora*, por meio de um modelo de língua como um sistema estratificado, organizado pelas metafunções, foi aplicada como a abordagem metodológica de coleta e análise por ocupar-se da língua em uso real, ou seja, o texto, considerado a unidade de funcionamento linguístico. Assim, a caracterização dos tipos de texto por meio das variáveis do contexto de situação mostra evidências de como os tipos dos tipos de textos são instanciadas no e pelo texto, o que se faz com o texto, como se constrói os significados por meio das relações estabelecidas entre os participantes e o valor dessas relações e do texto *per se*. Portanto, a análise dos dados busca mostrar as relações construídas na sociedade sobre cada um dos registros considerados isoladamente em inglês e em português. Essa abordagem propiciou também a descrição do perfil metafuncional dos textos no ambiente multilíngue.

Apresentam-se em, primeiro lugar, os procedimentos de análise adotados com o aporte do ambiente de programação R e o etiquetador morfossintático automático Treetagger. Na sequência, descrevem-se os procedimentos de análise com base na síntese das variáveis do contexto de situação e as suas respectivas variáveis. Por fim, ilustram-se os procedimentos de análise, por meio da anotação semiautomática no programa UAM CorpusTool, para a descrição metafuncional dos textos, mais especificamente com enfoque nas escolhas metafuncionais, no intuito de caracterizar cada um dos registros sob a óptica da Linguística Sistêmico-Funcional, (O'DONNELL, 2008).

### **2.2.1 Procedimentos de análise a partir dos dados proveniente dos corpora**

Os procedimentos de análise dos dados provenientes dos *corpora* foram aplicados no intuito de caracterizar o desenho desses *corpora* em relação à sua dimensão, extrair as frequências de co-ocorrências de itens lexicais e itens gramaticais a fim buscar os padrões de uso linguístico, além de etiquetar esses itens em classes de palavras por tipo de texto e por língua.

Inicialmente, foram calculadas a razão *type/token* simples e a razão *type/token* padronizada a partir da extração automática do número total de *types* e *tokens*, que ilustram os dados gerais dos *corpora*, bem como sua dimensão no que toca à densidade lexical dos três tipos de texto. O resultado da densidade lexical é aproximado e simplificado; para procedimentos mais robustos, porém mais complexos, *cf.* Steiner (2005, p. 37). Para esta pesquisa, o cálculo da densidade lexical foi realizado para efeitos ilustrativos a partir da razão da classe de



A técnica de nuvem de palavras auxilia na caracterização de cada um dos *corpora* a partir das palavras mais frequentes. Assim, é possível estudar as variações dos registros utilizados para construir significados sobre a anemia falciforme. A análise das nuvens de palavras mostra de forma gráfica a frequência de ocorrência das palavras em cada um dos *corpora*, sem, contudo, considerar a disposição dessas palavras em sentença ou a classe de palavras a que cada palavra pertence.

Para elucidar ainda mais a forma como os tipos de texto são instanciados, procedeu-se à identificação das classes de palavras mais frequentes, após o etiquetamento dos *types* por meio do Treetagger por tipo de texto e por idioma. Essa parte da análise traz indícios da densidade lexical dos diferentes tipos de textos no que toca a frequência de verbos e substantivos. Essa análise possibilita elucubrações sobre a assertiva de que a densidade lexical aumenta à medida que os textos se tornam mais especializados. Para tal, considera-se que menos orações e mais nominalizações em cada um dos tipos de texto estão atrelados a um menor número de verbos e a um maior número de substantivos, respectivamente. A presença de menos verbos e mais substantivos em um tipo de texto mostra que a informação desse tipo de texto foi condensada. Halliday e Matthiessen (1999) assumem que a nominalização dos verbos é uma característica do uso da linguagem na ciência. Nesta pesquisa, partiu-se do pressuposto de que os artigos científicos apresentam maior densidade lexical, seguidos dos manuais técnicos e, por fim, as cartilhas e *folders* de divulgação nas duas línguas.

O teste do  $\chi^2$  foi realizado para investigar se a frequência total dos itens lexicais (substantivos, adjetivos, advérbios e verbos) e dos itens gramaticais (pronomes e conjunções) observada em cada tipo de texto e por língua se desviou ou não da frequência esperada em relação ao número total de *tokens*. Para identificar qual classe de palavra está associada com qual língua (em qual das duas línguas a classe de palavra ocorre em maior porcentagem), utilizou-se o teste Z de comparação de proporção (porcentagem), considerando-se um nível de significância de 0,05. Dessa forma, se para uma determinada classe de palavra testada o p-valor fosse menor que 0,05, rejeitar-se-ia a hipótese nula de que as proporções (ou porcentagens) são iguais. Se o p-valor<sup>27</sup> fosse 0,05, não se rejeitaria a hipótese nula de que as proporções (ou porcentagens) são iguais.

---

<sup>27</sup> Entende-se por valor p a probabilidade de se encontrar, por um acaso, uma configuração específica nos dados se a hipótese nula for verdadeira (SIQUEIRA; TIBÚRCIO, 2011, p. 241). Já o nível de significância é

Após analisadas as frequências de ocorrência de cada uma das classes de palavras, partiu-se para uma análise mais detalhada dos itens lexicais mais frequentes por tipo de texto e por idioma, considerando-se exemplos das classes de substantivos, pronomes e conjunções. Portanto, para este estudo, foi realizada, primeiramente, a análise dos quatro nódulos identificados por meio dos DeCS<sup>28</sup> e MeSH<sup>29</sup>, previamente selecionados, são eles: “doença falciforme” e “anemia falciforme” (em português), bem como “*sickle cell disease*” e “*sickle cell anemia/anaemia*” (em inglês). Observaram-se esses nódulos juntamente com a parte do texto onde ocorreram, sendo possível geral uma lista de palavras e combinações de palavras que ocorreram próximas a esses nódulos. Na sequência, escrutinizaram-se os nódulos com as palavras “dor”, “*pain*”, “paciente”, “*patient*” Essas palavras foram selecionadas por serem frequentes. Neste estudo, foi analisado mais detidamente o pronome “*you*”, por ser um indicativo de caracterização específica da cartilha e *folder* de divulgação em inglês, considerado um aspecto do impacto da sintonia no modo e campo, o correspondente em português “você” não foi encontrado na cartilha em português. Na sequência, levantaram-se as combinações de palavras (*clusters*) mais frequentes, selecionadas a partir dos nódulos analisados. Entende-se por padrões a sequência fixa de palavras, como “tem por objetivo”, “sabe-se que”, etc. (SINCLAIR, 1991). Para verificar se há associação significativa entre a língua, o tipo de texto e os nódulos, foi realizado o teste  $\chi^2$ . Analisaram-se esses padrões de uso linguístico para identificar indícios que permitam caracterizar os diferentes tipos de texto.

A Subseção 2.2.2 a seguir apresenta os procedimentos de análise com o aporte da Linguística Sistêmico-Funcional.

### **2.2.2 Procedimentos de análise sistêmico-funcional**

A aplicação da Linguística Sistêmico-Funcional implica ocupar-se da língua em uso, ou seja, o texto, considerado a unidade de funcionamento linguístico. Para tanto, observa-se cada um dos tipos de texto a partir da tipologia de textos (MATTHIESSEN; TERUYA; WU, 2008),

---

interpretado como a probabilidade de se rejeitar essa hipótese nula quando ela for verdadeira (erro Tipo I). O valor de referência utilizado nessa análise foi de 0,05<sup>27</sup>. Assim, se o valor p fosse maior que o nível de significância ( $p > 0,05$ ) usado como referência no presente estudo, a hipótese nula não seria rejeitada; se o valor p fosse menor, rejeitar-se-ia essa hipótese.

<sup>28</sup> Descritores em Ciências da Saúde (DeCS). Disponível em: <<http://decs.bvs.br/P/decsweb2012.htm>>. Último acesso em: 15 set. 2012.

<sup>29</sup> Medical Subject Headings (MeSH). Disponível em: <<http://www.ncbi.nlm.nih.gov/mesh>>. Último acesso em: 15 set. 2012.

baseada no contexto e pautada por parâmetros das variáveis campo, sintonia e modo, como o objetivo de classificar os registros conforme tipologia de textos.

A interface entre a Linguística de Corpus e os Estudos da Tradução para a extração de dados passíveis de análise sob a perspectiva da Linguística Sistêmico-Funcional propiciou a investigação dos padrões de uso linguístico dos textos que compuseram os *corpora* extraídos pelo ambiente de programação R e anotados pelo programa Treetagger. A seguir, estão descritos os procedimentos de análise de textos em contexto de uso com base nos estudos de Eggins (2004).

#### 2.2.2.1 Procedimento de análise de textos em contexto de uso

Este procedimento de análise foi pautado no estudo de Eggins (2004), com base em Halliday e Hasan (1989), os quais postulam que todo texto é realizado em um contexto de situação, considerando a interação estreita entre o ambiente social e a língua enquanto um sistema funcional. Eggins (2004) destaca o conceito de texto segundo a LSF enquanto unidade semântica a partir do qual os significados de noção de mundo (ideacional), interação entre os participantes (interpessoal) e a organização do texto (textual) são realizados por meio dos sistemas TRANSITIVIDADE, MODO e TEMA, respectivamente, explicados mais detidamente na Seção 2.2.2.2. A autora se pauta na concepção dos linguistas sistêmicos em analisar como o contexto se insere no texto, definindo o registro enquanto o “contexto de situação no texto”<sup>30</sup> (EGGINS, 2004, p. 85), ou seja, a forma como o contexto de situação se realiza no texto, descrita pelas variáveis campo, sintonia e modo. Neste estudo, investiga-se o pressuposto de que há uma diferença, em função do registro, entre os padrões de uso linguístico – padrões esses identificados sob a perspectiva dos parâmetros das variáveis do contexto campo, sintonia e modo, explicadas mais detidamente a seguir:

- **Campo:** está organizado no texto de forma que construa/reflita a natureza da atividade social, incluindo os tipos das ações, os participantes e a maneira como percebem, sentem, experienciam e representam o mundo, além das circunstâncias. No caso dos *corpora* desta pesquisa, o domínio experiencial é o mesmo para os três tipos de texto analisados. A atividade social divide-se em, por um lado, apresentar uma visão sobre a doença e argumentar a favor dessa visão (artigo) e orientar o

---

<sup>30</sup> Minha tradução do original: “Register: context of situation in text”.

comportamento do leitor (manual técnico e cartilhas e *folders* de divulgação). No que toca à orientação/propósito há diferenças significativas, pois cada registro é realizado de forma a atingir um objetivo, considerando-se a posição do leitor em relação ao autor. Para entender como essas diferenças foram construídas, foram analisadas as itens lexicais (substantivos, adjetivos, advérbios e verbos) mais frequentes. Alguns exemplos da classe de substantivos foram utilizados como como nódulos para a busca por padrões de uso linguístico em cada um dos tipos de texto, no intuito de se obter indícios que caracterizassem os registros por meio das escolhas expressas pela experiência externa e interna dos seus participantes.

- **Sintonia:** descreve as relações humanas construídas na interação em que os papéis agentivo, discursivo e social do autor é definido em relação aos papéis do leitor, além de mostrar a distância social entre o autor e leitor, bem como o posicionamento do autor em relação ao leitor (afeto e valoração). A construção dos significados interpessoais é importante para ajustar os parâmetros da variável sintonia com o que está sendo comunicado e negociado entre autor e leitor. Para verificar essa negociação, foram analisados alguns dos pronomes pessoais identificados em cada um dos tipos de texto por língua. Por meio de pronomes, é possível verificar como a mudança na sintonia gera impacto no campo e no modo, além de estabelecer como as relações entre autor e leitor se dão ao longo do texto ao se construir os significados a partir de participantes, como os pacientes, por exemplo. Em outras palavras, a sintonia foi estudada por meio do tipo de relação social entre os participantes na situação, destacando os papéis discursivo, social e agentivo. Já as relações de poder entre os participantes foram identificadas pelo gênero social, nível de expertise e classe social, além do posicionamento do autor em relação ao leitor no que toca aos parâmetros afeto e valoração. Em suma, os parâmetros da variável sintonia permitem analisar os papéis sociais desempenhados entre os participantes no contexto de situação.
- **Modo:** estabelece a forma como o texto foi construído, a partir do papel da língua na interação. Nos *corpora* analisados, partiu-se do pressuposto de que o modo é análogo entre os textos no que toca ao papel constitutivo da língua na divisão das tarefas, ao canal gráfico da comunicação, ao meio em que o texto é veiculado (escrito) e ao modo de interação (monólogo). Espera-se, todavia, que o modo retórico (*e.g.*, persuasão, argumentação, instrução e exposição) varie em cada um dos registros. Em suma, na LSF, o modo se refere à forma como a língua está sendo

usada e à sua organização específica em termos de papel, canal, modo retórico, coesão, coerência, entre outros (HALLIDAY; HASAN, 1985, p. 15). Para verificar como cada um dos tipos de texto foi realizado pela variável modo nos *corpora* desta pesquisa, foram analisadas as conjunções, responsáveis por estabelecer as relações lógicas entre as partes do texto. Halliday e Hasan (1976, p. 18) afirmam que a orientação do texto é vista na repetição das palavras que constituem os padrões do léxico, ou seja, como as palavras estabelecem relações entre si e estabelecem a coesão através do texto, por meio de conjunções. Eggins (2004) explica que a coesão pode ser de três tipos, a saber: (i) referência; (ii) conjunção; e (iii) coesão lexical. Para Eggins (2004, p. 105) os padrões coesivos de conjunções indicam a forma como o autor expressa e cria as relações lógicas entre as partes do texto. Buscaram-se os indícios que levam à caracterização dos três tipos de texto como indicadores do modo retórico – mais especificamente, como os autores dos textos usam as conjunções no intuito de fazer uma argumentação, uma exposição ou divulgação de informações acerca da doença falciforme.

A FIG. 42 abaixo ilustra especificamente o lugar dos tipos de texto desta pesquisa na tipologia dos textos:

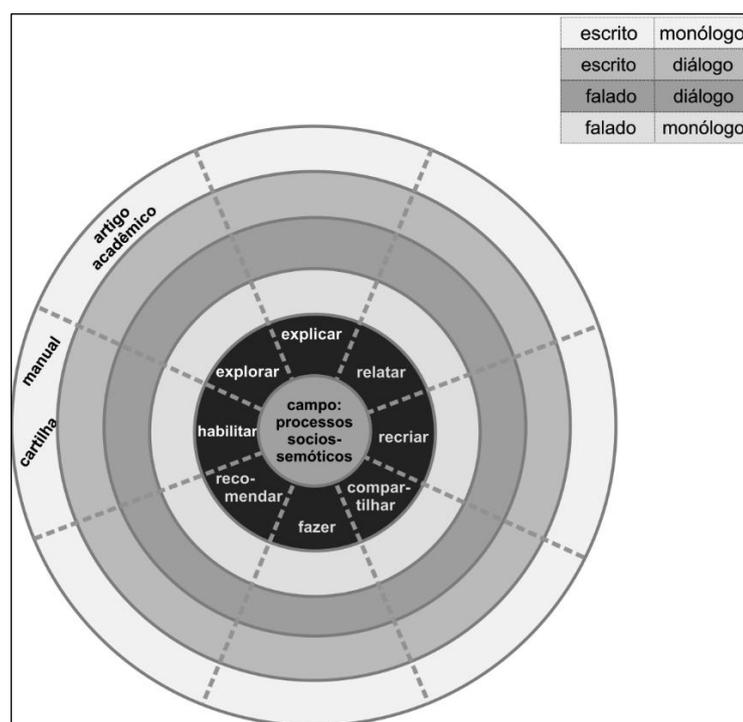


FIGURA 42. Tipologia de textos

Fonte: adaptada de Matthiessen, Teruya e Lam (2010, p. 221).

A FIG. 42 sintetiza a forma com que esses textos foram localizados dentro de cada um dos processos sociosemióticos específicos: EXPLORAR (artigo acadêmico) e HABILITAR (manual técnico e cartilha e *folder* de divulgação).

Apresentam-se a seguir os procedimentos de análise para a descrição do perfil metafuncional dos textos a partir dos sistemas de MODO, TRANSITIVIDADE, TEMA e MENSAGEM (HALLIDAY; MATTHIESSEN, 1999), por meio do programa UAM CorpusTool.

#### 2.2.2.2 Procedimento de análise do perfil metafuncional dos textos

Esta parte da pesquisa também se pauta na interface entre os Estudos da Tradução e a Linguística de Corpus para a extração de dados passíveis de análise sobre a perspectiva da Linguística Sistêmico-Funcional. Mais especificamente, procede-se à análise dos sistemas de TRANSITIVIDADE, MODO, e TEMA e MENSAGEM (HALLIDAY; MATTHIESSEN, 2004) e MENSAGEM (MARTIN; ROSE, 2007), na tentativa de entender quais escolhas metafuncionais podem ser identificadas nos diferentes tipos de textos que compuseram os *corpora* desta pesquisa. Em termos sintéticos, delinearam-se os seguintes objetivos específicos:

- Analisar as escolhas metafuncionais de textos do tipo EXPLORAR e HABILITAR e em português e em inglês (HALLIDAY; MATTHIESSEN, 2004); e
- Propor um modelo de sistematização das escolhas metafuncionais dos textos.

A abordagem empregada para a análise do perfil metafuncional é a conduzida pelo *corpus* (TOGNINI-BONELLI, 2001), uma vez que permite construir um modelo de sistematização das escolhas e apresentar padrões e distinções entre os dados constituintes de um *corpus* para o embasamento de formulações de generalizações linguísticas. Os textos foram anotados e etiquetados de forma semiautomática por meio do programa UAM CorpusTool.

Foram extraídas as telas do programa UAM CorpusTool, conforme ilustram as FIG. 43 a FIG. 46, a seguir. Essas figuras foram compiladas de forma a ilustrar como as frequências relativas dos sistemas de TRANSITIVIDADE, MODO, e TEMA e MENSAGEM, por língua e por registro, foram sintetizadas com vistas a possibilitar uma comparação inter-registros e interlíngua na descrição do perfil metafuncional dos textos.

A FIG. 43 a seguir sintetiza as frequências relativas do sistema de TRANSITIVIDADE nos textos do tipo EXPLORAR (artigo acadêmico), HABILITAR (manual técnico e cartilha e *folder* de divulgação) em português e em inglês.

Unit: processo + Show		Artigo_portug		Artigo_ingles		Manual_portug		Manual_ingles		Cartilha_port		Cartilha_ingl	
Feature	N	Percent											
Total Units	63		49		78		52		58		50		
PROCESSO-TYPE	N=63		N=49		N=78		N=52		N=58		N=50		
- material	19	30.16%	25	51.02%	17	21.79%	5	9.62%	7	12.07%	11	22.00%	
- relacional	38	60.32%	23	46.94%	60	76.92%	42	80.77%	50	86.21%	39	78.00%	
- mental	5	7.94%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	
- existencial	0	0.00%	1	2.04%	1	1.28%	5	9.62%	0	0.00%	0	0.00%	
- comportamental	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	
- verbal	1	1.59%	0	0.00%	0	0.00%	0	0.00%	1	1.72%	0	0.00%	
RELACIONAL	N=38		N=23		N=60		N=42		N=50		N=39		
- identificativo	0	0.00%	3	13.04%	5	8.33%	2	4.76%	4	8.00%	2	5.13%	
- atributivo	38	100.00%	20	86.96%	55	91.67%	40	95.24%	46	92.00%	37	94.87%	

FIGURA 43. Síntese das frequências relativas do sistema de TRANSITIVIDADE nos textos do tipo EXPLORAR e HABILITAR

Para compreender a construção do mundo das experiências (HALLIDAY, 2002), a análise da FIG. 43 a partir das frequências relativas dos termos do sistema de TRANSITIVIDADE permite identificar os processos mais significativos para todos os tipos de textos.

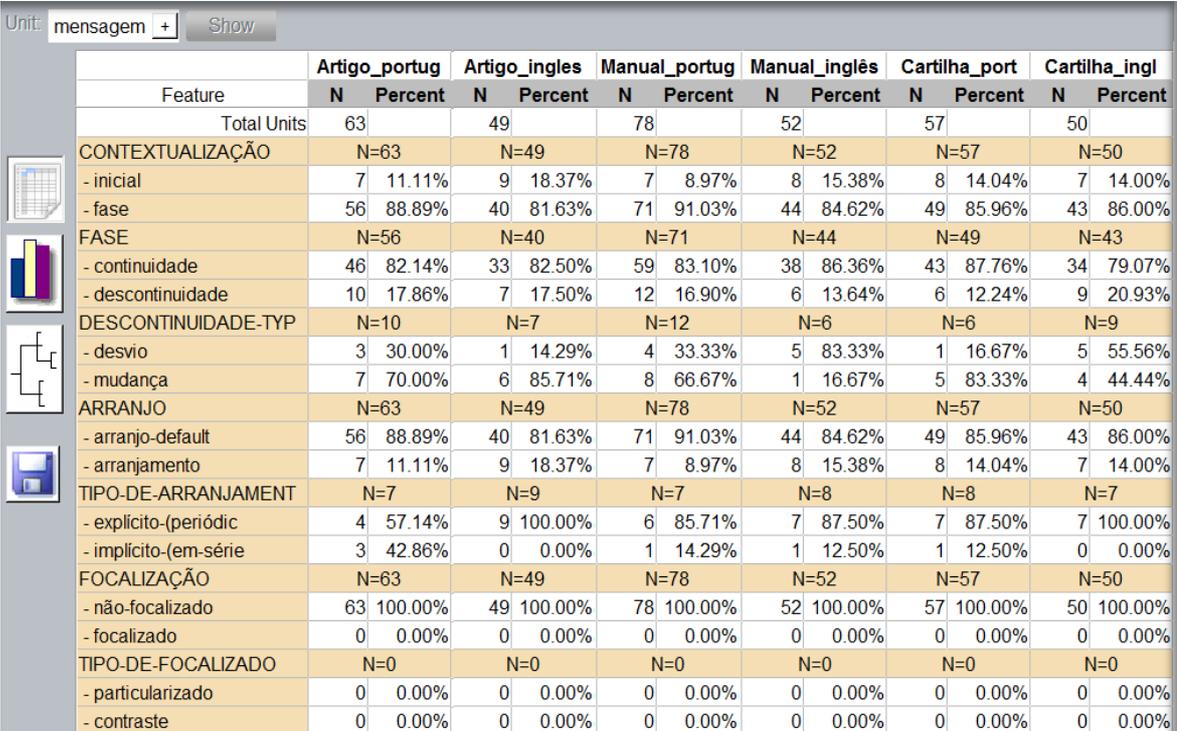
Para a construção do mundo social (HALLIDAY, 2002), as escolhas no sistema de MODO indicam como estão construídas as relações entre autor e leitor. A FIG. 44 ilustra a síntese das frequências relativas ao sistema de MODO nos textos do tipo EXPLORAR e HABILITAR.

Unit: modo + Show		Artigo_portug		Artigo_ingles		Manual_portug		Manual_ingles		Cartilha_port		Cartilha_ingl	
Feature	N	Percent											
Total Units	63		49		78		52		57		50		
MODO-TYPE	N=63		N=49		N=78		N=52		N=57		N=50		
- indicativo	63	100.00%	49	100.00%	78	100.00%	52	100.00%	57	100.00%	49	98.00%	
- imperativo	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	1	2.00%	
INDICATIVO-TYPE	N=63		N=49		N=78		N=52		N=57		N=49		
- declarativo	63	100.00%	49	100.00%	78	100.00%	52	100.00%	57	100.00%	49	100.00%	
- interrogativo	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	
TURNO	N=63		N=49		N=78		N=52		N=57		N=50		
- inicial	63	100.00%	49	100.00%	78	100.00%	52	100.00%	57	100.00%	50	100.00%	
- respondente	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	
PAPEL	N=63		N=49		N=78		N=52		N=57		N=50		
- oferta	63	100.00%	49	100.00%	78	100.00%	52	100.00%	57	100.00%	49	98.00%	
- demanda	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	1	2.00%	
MERCADORIA	N=63		N=49		N=78		N=52		N=57		N=50		
- informação	63	100.00%	49	100.00%	78	100.00%	52	100.00%	57	100.00%	49	98.00%	
- bens-e-serviços	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	1	2.00%	

FIGURA 44. Síntese das frequências relativas do sistema de MODO nos textos do tipo EXPLORAR e HABILITAR

Por fim, para analisar a realidade semiótica, foram observadas as escolhas do sistema de MENSAGEM (*cf.* FIG. 45) e TEMA em português e em inglês (*cf.* FIG. 46 e 47, respectivamente). Essas escolhas indicam a forma como os significados ideacionais e interpessoais são distribuídos ao longo do texto (HALLIDAY, 2002).

A FIG. 45 mostra a síntese das frequências relativas e os termos do sistema de MENSAGEM (MARTIN; ROSE, 2007) nos textos do tipo EXPLORAR e HABILITAR. Cumpre ressaltar que o sistema de MENSAGEM organiza as informações ao longo do texto.



Unit: mensagem + Show		Artigo_portug		Artigo_ingles		Manual_portug		Manual_ingles		Cartilha_port		Cartilha_ingl	
Feature	N	Percent											
Total Units	63		49		78		52		57		50		
CONTEXTUALIZAÇÃO	N=63		N=49		N=78		N=52		N=57		N=50		
- inicial	7	11.11%	9	18.37%	7	8.97%	8	15.38%	8	14.04%	7	14.00%	
- fase	56	88.89%	40	81.63%	71	91.03%	44	84.62%	49	85.96%	43	86.00%	
FASE	N=56		N=40		N=71		N=44		N=49		N=43		
- continuidade	46	82.14%	33	82.50%	59	83.10%	38	86.36%	43	87.76%	34	79.07%	
- descontinuidade	10	17.86%	7	17.50%	12	16.90%	6	13.64%	6	12.24%	9	20.93%	
DESCONTINUIDADE-TYP	N=10		N=7		N=12		N=6		N=6		N=9		
- desvio	3	30.00%	1	14.29%	4	33.33%	5	83.33%	1	16.67%	5	55.56%	
- mudança	7	70.00%	6	85.71%	8	66.67%	1	16.67%	5	83.33%	4	44.44%	
ARRANJO	N=63		N=49		N=78		N=52		N=57		N=50		
- arranjo-default	56	88.89%	40	81.63%	71	91.03%	44	84.62%	49	85.96%	43	86.00%	
- arranjo	7	11.11%	9	18.37%	7	8.97%	8	15.38%	8	14.04%	7	14.00%	
TIPO-DE-ARRANJAMENT	N=7		N=9		N=7		N=8		N=8		N=7		
- explícito-(periódic	4	57.14%	9	100.00%	6	85.71%	7	87.50%	7	87.50%	7	100.00%	
- implícito-(em-série	3	42.86%	0	0.00%	1	14.29%	1	12.50%	1	12.50%	0	0.00%	
FOCALIZAÇÃO	N=63		N=49		N=78		N=52		N=57		N=50		
- não-focalizado	63	100.00%	49	100.00%	78	100.00%	52	100.00%	57	100.00%	50	100.00%	
- focalizado	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	
TIPO-DE-FOCALIZADO	N=0												
- particularizado	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	
- contraste	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	

FIGURA 45. Síntese das frequências relativas do sistema de MENSAGEM nos textos do tipo EXPLORAR e HABILITAR

A partir dos dados ilustrados na FIG. 45, é possível estudar o padrão de organização textual nos diferentes tipos de textos. No que toca à contextualização, observa-se que é possível analisar as frequências relativas do sistema de MENSAGEM e quantificar as mensagens iniciais e fases (continuidade, descontinuidade:desvio e descontinuidade:mudança) em todos os textos.

Pela análise do Arranjo, é possível observar se as fases apresentam arranjo explícito (periódico) ou implícito (série). O primeiro indica que o autor optou por uma organização textual que não demandando inferência do leitor; o segundo indica que a relação entre a

mensagem e o fluxo textual não é explícita e indica que o autor optou por uma organização textual que demanda inferência do leitor. Por fim, analisa-se a focalização, a qual focalização garante proeminência textual para uma mensagem inserida na frase, podendo ser um tipo particularizado da mensagem ou um contraste entre uma mensagem e demais mensagens do texto.

O sistema semântico de MENSAGEM é realizado na gramática pelo sistema de TEMA, responsável por localizar e orientar a oração em seu contexto (HALLIDAY; MATTHIESSEN, 2004). Esse sistema deve ser analisado considerando-se a distinção em relação às línguas portuguesa e inglesa. Nesta pesquisa, seguiu-se a proposta de Figueredo (2011) para a classificação dos tipos de Tema em português e a proposta de Halliday (2002) para a classificação dos tipos de Tema em inglês, conforme ilustrado a seguir.

A FIG. 46 ilustra a síntese das frequências relativas do sistema de TEMA em português nos textos do tipo EXPLORAR e HABILITAR.

Unit: tema +	Artigo_portug		Manual_portug		Cartilha_port	
Feature	N	Percent	N	Percent	N	Percent
Total Units	63		78		57	
<b>COMPLEXIDADE</b>	N=63		N=78		N=57	
- elemental	0	0.00%	0	0.00%	0	0.00%
- ângulo	7	11.11%	0	0.00%	0	0.00%
- textual	24	38.10%	15	19.23%	12	21.05%
- interpessoal	0	0.00%	0	0.00%	1	1.75%
- tema-default	32	50.79%	63	80.77%	44	77.19%
<b>ÂNGULO-TYPE</b>	N=7		N=0		N=0	
- fonte	7	100.00%	0	0.00%	0	0.00%
- ponto-de-vista	0	0.00%	0	0.00%	0	0.00%
<b>SELEÇÃO-TEMÁTICA</b>	N=63		N=78		N=57	
- default	53	84.13%	67	85.90%	51	89.47%
- proeminente	10	15.87%	11	14.10%	6	10.53%

FIGURA 46. Síntese das frequências relativas do sistema de TEMA em português nos textos do tipo EXPLORAR e HABILITAR

Com base nos dados ilustrados na FIG. 46, é possível analisar os tipos de tema mais frequentes em relação à complexidade. A distribuição dos diferentes tipos de tema nos textos garante a organização interna dos textos.

A seleção temática (cf. FIG. 46) engloba duas opções de orientação da oração relativa à variável do contexto *campo*. A análise dessa característica permite observar a frequência de

ocorrência dos tipos *default* e proeminente. A seleção *default* garante a continuidade da oração na fase; a seleção proeminente indica a ocorrência de descontinuidade em uma fase.

A FIG. 47 mostra as frequências relativas do sistema de TEMA em inglês nos textos do tipo EXPLORAR e HABILITAR.

Feature	Artigo_inglês		Manual_inglês		Cartilha_ingl	
	N	Percent	N	Percent	N	Percent
Total Units:	49		52		50	
<b>COMPLEXIDADE</b>	N=49		N=52		N=50	
- simples	35	71.43%	41	78.85%	41	82.00%
- múltiplo	14	28.57%	11	21.15%	9	18.00%
<b>MÚLTIPLO</b>	N=14		N=11		N=9	
- textual-e-tópico	14	100.00%	11	100.00%	8	88.89%
- interpessoal-e-tópico	0	0.00%	0	0.00%	1	11.11%
- textual-interpessoa	0	0.00%	0	0.00%	0	0.00%
<b>SELEÇÃO-TEMÁTICA</b>	N=49		N=52		N=50	
- não-marcado	47	95.92%	52	100.00%	48	96.00%
- marcado	2	4.08%	0	0.00%	2	4.00%

FIGURA 47. Síntese das frequências relativas do sistema de TEMA em inglês nos textos do tipo EXPLORAR e HABILITAR

Pelos dados mostrados na FIG. 47, entende-se como os tipos de texto em inglês se apresentam em sua organização interna, no que toca à complexidade, apresenta temas simples e múltiplos. Dentre os temas múltiplos, é possível observar os mais frequentes em todos os tipos de textos. No que toca à seleção temática, é possível observar os tipos de temas mais frequentes. Os temas não marcados se caracterizam pela colocação em posição inicial de modo que não seja o mais esperado ou típico, levando-se em consideração tipológica de português e inglês por serem línguas com estrutura sujeito/verbo/objeto (SVO), correspondendo ao sujeito em uma oração declarativa; os temas marcados ocorrem em posições não esperadas na oração.

Comparando-se as FIG. 46 e 47 é possível observar a diferença existente entre as classificações dos tipos de Tema propostas inicialmente por Halliday (2002) para a língua inglesa e as classificações propostas por Figueredo (2011) para a língua portuguesa, o que sugere que o sistema de TEMA pode variar de acordo com as línguas. Por meio desse sistema foi possível mapear algumas semelhanças e diferenças que permitiram classificar os três tipos de textos analisados nesta pesquisa.

A partir dessas análises, pretendeu-se aplicar os perfis metafuncionais na descrição de tipos de textos da área da saúde, a partir dos sistemas de TRANSITIVIDADE, MODO, e TEMA e MENSAGEM, no intuito de:

1. Extrair dados dos tipos de texto a partir da descrição desses sistemas;
2. Elaborar um texto por meio do *template* (modelo), criado para a elaboração de cartilha, conforme mencionado na Seção 1.3, a partir das escolhas metafuncionais com base na gramática sistêmico-funcional hallidayana.

Esse *template* está transcrito na íntegra na FIG. 48 a seguir. Vale ressaltar que, por ser o primeiro passo para a elaboração de um texto baseado num *template* (modelo) criado com base na aplicação do perfil metafuncional de textos de popularização da ciência, são previstos ajustes em função da especificidade da cartilha.

<b>TEMPLATE (MODELO)</b>
<p><b>Pergunta inicial:</b>  <u>1 oração Indicativo:Interrogativo:Elemental/Processo Material/Inicial</u> – esta oração deve fazer uma pergunta que apresente a curiosidade sobre o fato científico do qual o texto trata.</p> <p><b>Características gerais do fato científico:</b>  <u>1 oração Indicativo:Declarativo/Processo Relacional Identificativo/Continuidade</u> – esta oração deve ter como Token o conceito principal do fato científico e como Value sua explicação.  <u>1 oração Indicativo:Declarativo/Processo Relacional Atributivo/Continuidade</u> – esta oração deve apresentar uma particularidade do conceito científico que será demonstrada de forma concreta na oração seguinte.  <u>1 oração Indicativo:Declarativo/Processo Material/Inicial</u> – esta oração deve exemplificar um fenômeno do mundo das experiências relacionado ao fato científico.  <u>1 oração Indicativo:Declarativo/Processo Relacional Atributivo/Descontinuidade:mudança</u> – esta oração deve enfatizar uma característica decorrente do fenômeno do mundo das experiências apresentado na oração anterior.  <u>1 oração Indicativo:Declarativo/Processo Existencial/Continuidade</u> – esta oração deve apresentar uma lista de participantes que ajudem a exemplificar a oração anterior.</p> <p><b>Explicação sobre o fato científico:</b>  <u>1 oração Indicativo:Declarativo/Processo Material/Inicial</u> – esta oração deve apresentar um fenômeno do mundo das experiências relacionado ao fato científico.  <u>1 oração Indicativo:Declarativo/Processo Relacional Identificativo/Descontinuidade:desvio</u> – esta oração deve focalizar o texto em um participante em particular da oração e destacar sua representatividade dentro do fenômeno apresentado.  <u>1 oração Indicativo:Declarativo/Processo Relacional Atributivo/Continuidade</u> – esta oração deve continuar a caracterizar o participante da oração anterior.  <u>1 oração Indicativo:Declarativo/Processo Material/Descontinuidade:mudança</u> – esta oração deve focalizar o texto em um evento envolvendo o participante da oração interior.  <u>1 oração Indicativo:Declarativo/Processo Relacional Atributivo/Continuidade</u> – esta oração deve seguir a oração anterior, completando seu sentido com uma explicação sobre o evento apresentado.  <u>1 oração Indicativo:Declarativo/Processo Relacional Atributivo/Continuidade</u> – esta oração deve seguir a oração anterior, completando-a com mais uma explicação sobre o evento apresentado.  <u>1 oração Indicativo:Declarativo/Processo Relacional Atributivo/ Descontinuidade:mudança</u> – esta oração deve focalizar o texto em uma citação de um estudo de um profissional da área que exemplifique o fenômeno estudado. A configuração das demais orações dependerá de como está configurada a citação escolhida.</p> <p><b>Experimento</b>  <u>1 oração Imperativo/Processo Mental/Inicial</u> – esta oração deve introduzir uma série de experimentos que podem ser feitos pelo leitor do texto para que comprove as informações apresentadas anteriormente.  <u>1 oração Imperativo/Processo Material/Continuidade</u> – esta oração deve conter um experimento que pode ser feito pelo leitor do texto para comprovar as informações apresentadas anteriormente.  <u>1 oração Indicativo:Declarativo/Processo Relacional Atributivo</u> – esta oração deve fornecer uma breve explicação sobre o resultado do experimento da oração anterior.  <u>1 oração Imperativo/Processo Material/Continuidade</u> – esta oração deve conter um experimento que pode ser feito pelo leitor do texto para comprovar as informações apresentadas anteriormente.  <u>1 oração Indicativo:Declarativo/Processo Relacional Atributivo</u> – esta oração deve fornecer uma breve explicação sobre o resultado do experimento da oração anterior.  <u>1 oração Imperativo/Processo Material/Continuidade</u> – esta oração deve conter um experimento que pode ser feito pelo leitor do texto para comprovar as informações apresentadas anteriormente.  <u>1 oração Indicativo:Declarativo/Processo Relacional Atributivo</u> – esta oração deve fornecer uma breve explicação sobre o resultado do experimento da oração anterior.</p>

FIGURA 48. *Template* (modelo) desenvolvido para a elaboração de cartilha em português

Partiu-se do pressuposto de que, com base nesse modelo de sistematização das escolhas interpessoais, experienciais e textuais, seria possível elaborar uma cartilha sobre a anemia falciforme com base no perfil de um texto de popularização da ciência (cf. Subseção 3.3.3), uma vez que este tipo de texto é novidade no contexto de cultura e, apesar de a cartilha estar

associada ao processo HABILITAR, também guarda analogia com textos do processo sociossemiótico EXPLICAR.

Vale ressaltar que, por ser o primeiro template (modelo) criado com base na aplicação do perfil metafuncional de textos de popularização da ciência, é possível que algumas das orações não sejam contempladas ou passíveis de serem seguidas na ordem em que apareçam. Vale ressaltar que, pelo seu caráter exploratório, o template (modelo) criado com base na aplicação do perfil metafuncional de textos de popularização da ciência precisa ser refinado, eliminando-se orações que dizem respeito a significados não necessários na cartilha . Além disso, a construção desse template foi importante por subsidiar a construção de significados com base nas funções dos sistemas de TRANSITIVIDADE, MODO, e TEMA e MENSAGEM. Com esse template, foi possível elaborar um texto do tipo HABILITAR sobre a anemia falciforme no ambiente multilíngue (*cf.* Subseção 3.3.3).

O próximo capítulo apresenta a análise dos dados provenientes dos corpora.

.

# **CAPÍTULO 3**

## **RESULTADOS DA ANÁLISE DOS DADOS**

### 3 ANÁLISE DOS DADOS

**E**ste capítulo apresenta a análise dos dados dos *corpora* e dos padrões de uso linguístico que geraram indícios para a categorização e descrição dos tipos de textos distribuídos nos processos sociossemióticos EXPLORAR e HABILITAR. Conforme explicado na Subseção 2.1.2 do Capítulo de Metodologia, os dados apresentados a seguir foram extraídos por meio de três ferramentas computacionais, quais sejam: (i) o ambiente de programação R, a partir da aplicação dos *scripts* desenvolvidos especificamente para atender aos objetivos desta pesquisa; (ii) o etiquetador morfossintático automático Treetagger; e (iii) o programa UAM CorpusTool. (*cf.* Subseção 2.1.2.3).

#### 3.1 Análise dos *Corpora*

Nesta seção, constam as análises dos dados que caracterizam os *corpora* em relação ao número total de *types* e *tokens*, à extração de co-ocorrências de itens lexicais e gramaticais e ao etiquetamento desses itens em classes de palavras por tipo de texto e por língua.

##### 3.1.1 Dimensão dos *corpora*

Os números apresentados na TAB. 1 descrevem a dimensão de cada *corpus* analisado no presente trabalho. Trata-se de um *corpus* de pequenas dimensões, justificado pela necessidade de validar os procedimentos metodológicos de coleta propostos no Capítulo 2, por meio da aplicação dos *scripts* no ambiente de programação R. Esses procedimentos permitiram análises baseadas na frequência e na co-ocorrência de palavras nos *corpora* compilados.

TABELA 1. Dados gerais dos *corpora* comparáveis

Tipo de texto	Língua	Total de <i>tokens</i>	Total de <i>types</i>	<i>Type/Token</i> (%) <sup>*</sup>	<i>Type/Token</i> Padronizada (%) <sup>*</sup>
Artigos acadêmicos	Português	3.035	979	0,28 (28%)	0,47 (47%)
	Inglês	3.078	868	0,32 (32%)	0,44 (44%)
Manuais técnicos	Português	3066	1.127	0,39 (39%)	0,49 (49%)
	Inglês	3.039	931	0,31 (31%)	0,41 (41%)
Cartilhas e <i>folders</i> de divulgação	Português	3.038	1.075	0,35 (35%)	0,46 (46%)
	Inglês	3.045	694	0,23 (23%)	0,33 (33%)

\* A relação *type/token* foi calculada em porcentagem para efeito de interpretação.

O total de *tokens* corresponde ao número total de palavras, incluindo as repetições desses mesmos itens. O total de *types* corresponde a cada palavra, desconsiderando-se as repetições, ou seja, informa o número de palavras diferentes dentro daquele *corpus*. A razão simples entre *type/token* (*i.e.*, número de palavras diferentes do *corpus* dividido pelo número total das palavras desse *corpus*) indica a variação lexical do texto. Quanto maior for essa razão (mais próximo de 1), mais variado será o léxico do texto; e quanto mais próximo de zero for essa razão, maior é a repetição de palavras. Em outras palavras, a razão *type/token* (variação lexical) é uma medida da produção linguística usada para verificar a diversidade lexical ou a variedade de diferentes palavras em cada um dos *corpora*. Já a relação *type/token* padronizada<sup>31</sup> mostra a razão *type/token* média: “é empregada para neutralizar a influência do tamanho do texto na computação da razão *type-token*, já que textos mais longos por natureza apresentam mais repetições e por isso tendem a possuir valores menores dessa razão em comparação aos textos curtos” (BERBER SARDINHA, 2004, p. 175).

Considerando-se a razão *type/token* padronizada, a TAB. 1 mostra que, dentre os *corpora* estudados, aquele que apresentou maior índice foi o manual técnico em português (0,49), sendo, portanto, o *corpus* que apresentou o menor número de repetições de itens lexicais, ou seja, aquele cujo léxico foi o mais variado. Para interpretar esse número, pode-se dizer que 49% dos itens lexicais desse *corpus* ocorrem apenas uma vez e que 51% repetem-se pelo menos uma vez. Já o *corpus* que apresentou o maior número de repetições em média foi o cartilhas e *folders* em inglês, isto é, o *corpus* que apresentou o léxico menos variado, ou seja,

<sup>31</sup> O cálculo da relação *type/token* padronizada está explicitado na Subseção 2.1.2.1.

33% dos itens lexicais desse *corpus* ocorrem apenas uma vez e 67% repetem-se, pelo menos uma.

A TAB. 1 revela ainda que a relação *type/token* padronizada é maior nos *corpora* em português que nos *corpora* em inglês. Esse dado indica que os textos em português apresentam um número maior de palavras diferentes, ou seja, maior variação lexical. Mais especificamente, têm-se as seguintes diferenças: 13 pontos percentuais para as cartilhas e *folders* de divulgação (0,46 em português e 0,33 em inglês); 3 pontos percentuais para o artigo acadêmico (0,47 em português e 0,44 em inglês); e 8 pontos percentuais para o manual técnico (0,49 em português e 0,41 em inglês).

Para comparar os mesmos tipos de texto, considerando-se as línguas português e inglês, foram analisadas as nuvens de palavras específicas para cada um dos tipos de texto – nuvens essas obtidas a partir de um algoritmo implementado no ambiente de programação R. A seguir, as imagens dessas nuvens estão organizadas conforme os tipos de texto em inglês e em português, observadas a partir das FIG. 49 a 54. Observe-se que, para essas imagens, foram desconsiderados os itens gramaticais, por não apresentarem relevância para o objeto de estudo.

As FIG. 49 e 50 a seguir ilustram a nuvens de palavras dos artigos acadêmicos em português e em inglês, respectivamente.



FIGURA 49. Nuvem de palavras elaborada com base no *corpus* artigo\_português

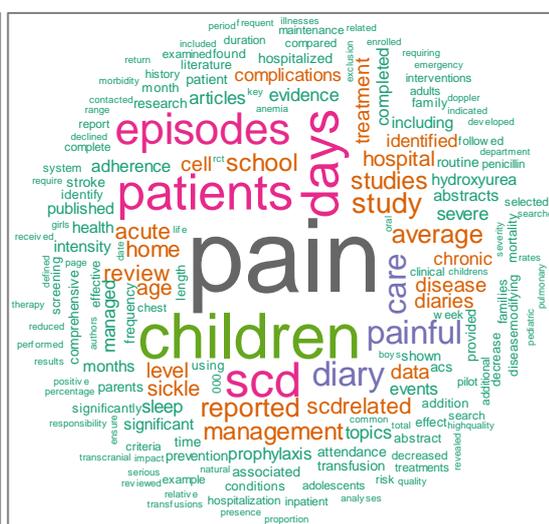


FIGURA 50. Nuvem de palavras elaborada com base no *corpus* artigo\_inglês



Comparando-se as FIG. 49, 51 e 53, nota-se que o artigo acadêmico em português apresenta três itens lexicais mais frequentes (“dor”, “doença” e “pacientes”), ao passo que os outros dois tipos de texto apresentam apenas um item lexical: “dor” se destaca no manual técnico e “falciforme” se destaca na cartilha e *folder* de divulgação. Percebe-se que os itens lexicais “falciforme”, “pacientes”, “doença” e “dor” estão presentes com uma frequência alta nos três tipos de texto em português, indicando-se, assim, que podem fazer parte dos padrões de uso linguístico mais frequentes.

Em relação aos verbos mais frequentes, destacam-se para o artigo acadêmico: “foi” e “foram”. Já em relação aos outros dois tipos de texto, os verbos que mais ocorrem são: “pode”, “deve” e “devem”. A frequência dessas palavras sugeriu que elas fossem analisadas separadamente a fim de verificar como ocorrem no texto. A partir da frequência de ocorrência desses itens lexicais e itens gramaticais, é possível identificar nódulos passíveis de serem analisados nos textos.

As FIG. 50, 52 e 54 mostram a distribuição das palavras nos tipos de texto em inglês. A palavra mais frequente no artigo acadêmico é “*pain*” e nos outros dois tipos de texto, “*sickle*”. No artigo acadêmico, destacam-se “*children*”, “*patients*”, “*SCD*”, “*episodes*” e “*days*”; no manual técnico, “*haemoglobin*”, “*disease*” e “*pain*”; nas cartilhas e *folders* de divulgação, “*disease*”, “*haemoglobin*”, “*cell*” e “*child*”. Em relação a verbos, os que se destacam no artigo acadêmico são “*reported*” e “*related*”.

A comparação entre todas as seis figuras mostra que há diferenças entre os *corpora* no que toca à frequência das palavras tanto em uma análise entre tipos de textos como em uma análise entre línguas. A frequência de ocorrência de palavras remete à forma como o domínio experiencial foi estabelecido pelo autor do texto. As palavras presentes em cada uma das nuvens remetem a indícios de como as informações sobre a anemia falciforme foram construídas em cada um dos tipos de textos. Esses indícios ensejam o estudo das variações dos registros utilizados para construir significados sobre a anemia falciforme e corroboram a afirmação de que há uma maior variação lexical nos tipos de texto em português.

A seguir, são apresentados os *types* após a realização do etiquetamento de cada um dos *corpora* desta pesquisa por meio do Treetagger.

### 3.1.2 Anotação dos corpora em classes de palavras

Como mencionado na Subseção 2.1.2.2, alguns símbolos e caracteres foram eliminados em virtude da necessidade de os *corpora* não apresentarem, por exemplo, sinais de pontuação, para que pudessem ser lidos pelo R. Houve exclusão também de algumas das categorias gramaticais presentes nos *tagsets* em inglês e em português que não se enquadravam em categorias comuns às duas línguas. Essas exclusões fizeram com que o número total de *tokens* sofresse alteração em relação àquele número apresentado na TAB. 1, referente aos dados gerais dos *corpora*.

Os dados provenientes do etiquetamento de cada um dos *corpora* estão tabulados a seguir por tipos de texto.

#### 1) Artigo acadêmico

A TAB. 2 mostra as frequências absoluta e relativa de ocorrência de cada uma das classes de palavras extraídas do artigo acadêmico.

TABELA 2. Frequência das classes de palavras no artigo acadêmico em inglês e em português

Classe de palavras	Português		Inglês	
	Frequência	Relação <i>type/token</i> (%)	Frequência	Relação <i>type/token</i> (%)
Adjetivo	348	11,6%	317	10,5%
Advérbio	87	2,9%	95	3,1%
Conjunção	164	5,5%	144	4,8%
Determinante	224	7,5%	260	8,6%
Interjeição	0	0,0%	0	0,0%
Numeral	82	2,7%	44	1,5%
Preposição	684	22,8%	540	17,8%
Pronome	48	1,6%	67	2,2%
Substantivo	950	31,6%	1093	36,1%
Verbo	413	13,8%	467	15,4%
TOTAL	3000	100,0%	3027	100,0%

Nota-se pela TAB. 2 que a frequência de ocorrência das classes de palavras apresenta uma diferença para este tipo do texto quando comparadas as línguas. À exceção da classe de

interjeição, que não apresentou ocorrências, a relação *type/token* das demais classes indica uma diferença na frequência de ocorrência de palavras de acordo com a língua. Nos textos em português, as classes de palavras numeral (2,7%), conjunção (5,5%), adjetivo (11,6%) e preposição (22,8%) apresentaram maior relação *type/token* que nos textos em inglês. Já em inglês, as classes de pronome (2,2%), advérbio (3,1%), determinante (8,6%), verbo (15,4%) e substantivo (36,1%) apresentam maior relação *type/token* que em português, indicando que há maior variabilidade dessas palavras nesse tipo de texto em inglês que em português.

Parte-se do pressuposto de que, em artigos acadêmicos, há menos orações e mais nominalizações encapsulando verbos e condensando as informações, o que implica uma alta densidade lexical. Halliday (1994, p. 57) explica que o aumento da densidade lexical se dá não devido à elevada ocorrência de itens lexicais, e sim pela diminuição da frequência de ocorrência dos verbos, que se nominalizam, o que faz com que o número de orações diminua por complexo oracional.

Dividindo-se o número de substantivos pelo número de verbos, chega-se a uma relação de 2,30 em português e 2,34 em inglês. Considerando-se o pressuposto supramencionado, espera-se que a relação substantivo por verbo seja maior no artigo acadêmico em comparação com os outros dois tipos de texto.

## 2) Manual Técnico

A TAB. 3 mostra as frequências absoluta e relativa de ocorrência de cada uma das classes de palavras extraídas do manual técnico.

TABELA 3. Frequência absoluta das classes de palavras no manual técnico em inglês e em português

Classe de palavras	Português		Inglês	
	Frequência	Relação <i>type/token</i> (%)	Frequência	Relação <i>type/token</i> (%)
Adjetivo	460	15,3%	314	10,4%
Advérbio	88	2,9%	131	4,3%
Conjunção	145	4,8%	159	5,3%
Determinante	249	8,3%	306	10,1%
Interjeição	0	0,0%	0	0,0%
Numeral	17	0,6%	38	1,3%
Preposição	593	19,7%	413	13,6%
Pronome	84	2,8%	75	2,5%
Substantivo	900	29,8%	1081	35,7%
Verbo	480	15,9%	511	16,9%
TOTAL	3016	100,0%	3028	100,0%

No que tange ao manual técnico, a análise das relações *type/token* indicam diferenças entre as línguas em relação às frequências de ocorrências em cada uma das classes de palavras analisadas. Também nesse tipo de texto não houve ocorrência de interjeições. Pelos dados da TAB. 3, nota-se que a relação *type/token* é mais elevada em português que em inglês no que diz respeito aos adjetivos (460 ou 15,3% em português contra 314 ocorrências ou 10,4% em inglês), preposições (593 ou 19,7% em português contra 413 ou 13,6% em inglês) e pronomes (84 ou 2,8% em português contra 75 ou 2,5% em inglês). Para todas as demais classes: advérbio, conjunção, determinante, interjeição, numeral, substantivo e verbo, a relação *type/token* é mais elevada no texto em inglês quando comparado ao texto em português.

Conforme afirma Halliday (1994, p. 351), quanto mais técnica a linguagem, maior densidade lexical, agrupando grande quantidade de itens lexicais por oração. Para o manual técnico, os dados da TAB. 3 apontam uma relação substantivo/verbo de 1,88 e 2,12 para o português e o inglês respectivamente. Comparando-se esses dados com os do artigo acadêmico, observa-se que o pressuposto se mantém: o artigo acadêmico tem densidade lexical maior que o manual técnico.

Espera-se, ainda, que a cartilha e *folder* de divulgação apresente uma densidade lexical ainda menor em relação aos outros dois tipos de texto analisados, em função de a língua ser menos especializada por estar mais direcionada a um público mais leigo.

### 3) Cartilhas e *folders* de divulgação

A TAB. 4 mostra as frequências absoluta e relativa de ocorrência de cada uma das classes de palavras extraídas da cartilha e *folder* de divulgação.

TABELA 4. Frequência absoluta das classes de palavras na cartilha e *folder* de divulgação em inglês e em português

Classe de palavras	Português		Inglês	
	Frequência	Relação <i>type/token</i> (%)	Frequência	Relação <i>type/token</i> (%)
Adjetivo	445	14,8%	251	8,3%
Advérbio	101	3,3%	232	7,7%
Conjunção	140	4,6%	124	4,1%
Determinante	254	8,4%	291	9,7%
Interjeição	0	0,0%	0	0,0%
Numeral	24	0,8%	27	0,9%
Preposição	554	18,4%	413	13,8%
Pronome	71	2,4%	178	5,9%
Substantivo	932	30,9%	870	28,9%
Verbo	495	16,4%	622	20,7%
TOTAL	3016	100,0%	3008	100,0%

Para este tipo de texto, também não há interjeições e se nota que a relação *type/token* apresenta diferenças interlínguas, conforme aquelas destacada para o artigo acadêmico e o manual técnico. As classes mais frequentes em português foram: adjetivo (14,8%), conjunção (4,6%), preposição (18,4%) e substantivo (30,9%). As classes de palavras mais frequentes na língua inglesa foram: substantivo (28,9%), verbo (20,7%), preposição (13,8%), determinante (9,7%) e adjetivo (8,3%).

Nota-se também uma diferença significativa em termos numéricos para a frequência de pronomes (2,4% em português e 5,9 em inglês) neste tipo de texto. Essa diferença em relação aos pronomes para esse tipo de texto pode ser significativa para a sua caracterização com base

na Linguística Sistêmico-Funcional. A análise dessa relação é um modo de descrever sistematicamente como as palavras de um texto se relacionam umas com as outras, identificando a que se referem (pessoas, lugares, qualidades e circunstâncias).

No caso da cartilha e folder de divulgação, a relação substantivo/verbo é de 1,88 em português e 1,42 em inglês. A considerar esses valores, observa-se que, para o inglês, mantém-se novamente o pressuposto de que a cartilha e folder de divulgação apresentaria uma densidade lexical ainda menor em relação aos outros dois tipos de texto analisados, em função de a língua ser menos especializada por estar mais direcionada a um público mais leigo. No entanto, o mesmo não pode ser dito em português, língua na qual a relação é idêntica àquela encontrada para o manual técnico.

A TAB. 5 a seguir mostra uma consolidação dos dados de relação *type/token* e relação substantivo/verbo.

TABELA 5. Síntese dos dados empíricos dos *corpora*

Variável	Rótulo					
	Artigos acadêmicos		Manuais técnicos		Cartilhas e <i>folders</i>	
	Português	Inglês	Português	Inglês	Português	Inglês
<i>Type/Token</i> Normal / Padronizada	0,28 / 0,47	0,32 / 0,44	0,39 / 0,49	0,31 / 0,41	0,35 / 0,46	0,23 / 0,33
Relação Substantivo/Verbo	2,30	2,34	1,88	2,12	1,88	1,42

Os dados da TAB. 5 sobre a relação *type/token* apontam que a variação lexical é maior em português que em inglês, sendo as diferenças inter-registro pouco evidentes em português, mas mais acentuada em inglês quando se compara o cartilha e *folder* de divulgação (0,33 para a relação *type/token* padronizada) com o manual técnico (0,41) e o artigo acadêmico (0,44). Em se tratando da densidade lexical, baseada na relação substantivo/verbo, observa-se que o artigo acadêmico é o mais denso (2,30 e 2,34 em português e em inglês, respectivamente), havendo diferenças interlíngua no nível de densidade dos outros dois tipos de texto: em inglês, manual técnico (2,12) em segundo lugar e cartilha (1,42) em terceiro lugar; em português, ambos em segundo lugar (1,88 para o manual técnico e 1,88 para a cartilha). Esse resultado aponta que esses dois tipos de texto, em português, guardam semelhanças em

comum, o que dificulta a caracterização distinta de cada um deles, sendo, necessário, por conseguinte, uma análise mais detalhada, com o aporte da Linguística Sistêmico-Funcional que possibilite fazer uma distinção mais clara entre o manual técnico (interação especialista – técnico da área da saúde) e a cartilha e *folder* de divulgação (interação especialista – público leigo).

Foram analisados mais detidamente os itens lexicais (substantivo, adjetivo, advérbio e verbo) e os itens gramaticais (conjunção e pronomes) por mostrarem indícios que possam ser utilizados para a categorização dos registros também por meio da abordagem da Linguística Sistêmico-Funcional. A seguir, com base nos dados tabulados nas TAB. 2 a 4, foram elaborados os GRAF. 1 a 3, que ilustram a frequência de ocorrência dos itens lexicais (substantivos, adjetivos, advérbios e verbos) e dos itens gramaticais (pronomes e conjunções) em cada um dos *corpora*. Conforme mencionado no Capítulo de Metodologia, Subseção 2.1.2.1.1, o teste do  $\chi^2$  foi realizado para investigar se a frequência total dos itens lexicais (substantivos, adjetivos, advérbios e verbos) e dos itens gramaticais (pronomes e conjunções) observada em cada tipo de texto e por língua se desviou ou não da frequência esperada em relação ao número total de *tokens*. Como os p-valores foram todos menores do que o nível de significância 0,05, rejeita-se a hipótese nula de que a configuração do tipo de texto em classe de palavras é independente da língua. Logo, pode-se concluir que há associação significativa entre a língua e a classe de palavras para todos os três tipos de textos.

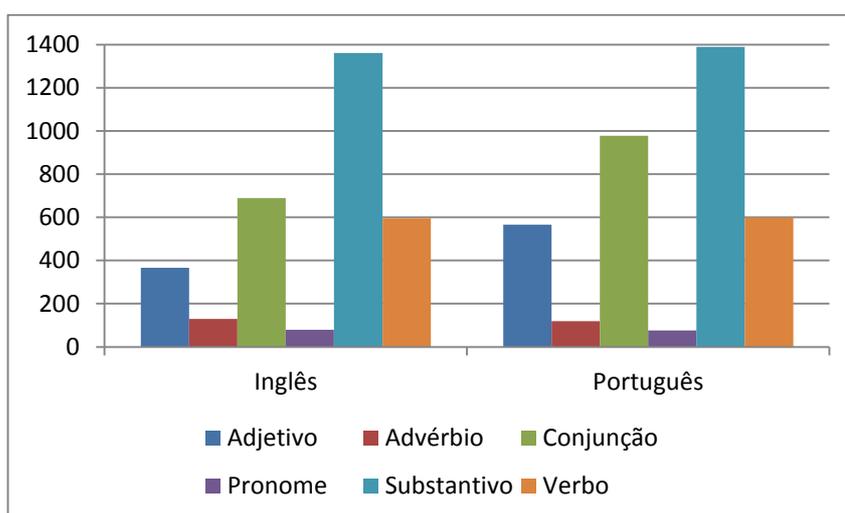


GRÁFICO 1. Frequência de ocorrência das principais classes de palavras no artigo acadêmico em inglês e em português

A análise do GRAF. 1 mostra que, no artigo acadêmico tanto em português quanto em inglês, como já era previsto, a classe de palavras mais frequente é a de substantivo, seguida pelas classes de conjunção, verbo, advérbio e pronome.

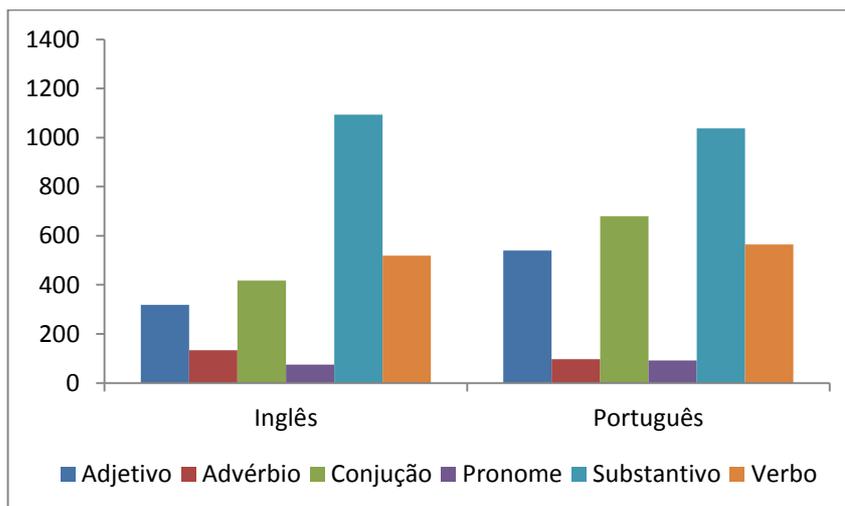


GRÁFICO 2. Frequência de ocorrência das principais classes gramaticais no manual técnico em inglês e em português

Para o manual técnico em português, o GRAF. 2 ilustra um comportamento parecido com aquele encontrado no artigo acadêmico em relação à frequência de ocorrência das palavras para o tipo de texto em português. Há uma maior ocorrência das classes de substantivos, seguida pela de conjunções, verbos, adjetivos, advérbios e pronomes. Já para o tipo de texto em inglês, observa-se que a classe de substantivos ocorre com maior frequência, seguida da classe de verbos, conjunções, adjetivos, advérbios e pronomes.

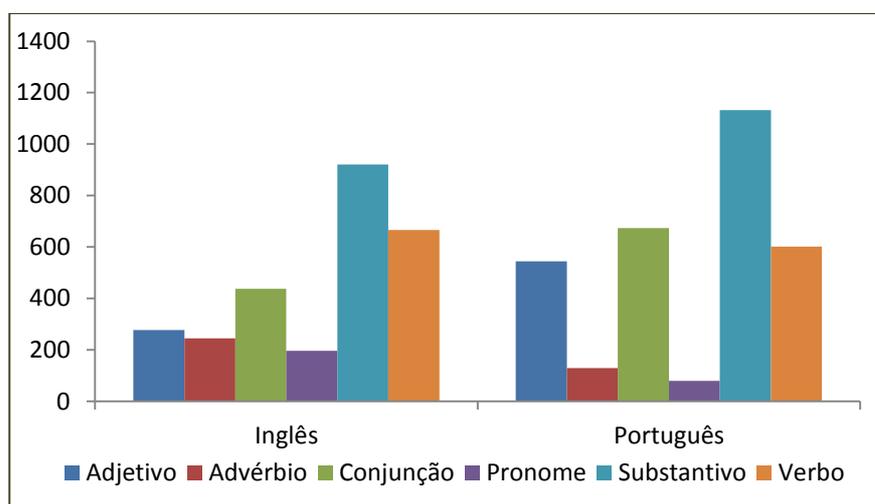


GRÁFICO 3. Frequência de ocorrência das principais classes gramaticais para a cartilha e *folder* de divulgação em inglês e em português

Para a cartilha e *folder* de divulgação, o GRAF. 3 mostra que o texto em inglês apresenta uma maior frequência de ocorrência da classe de substantivos, seguida pela de verbo, adjetivo, advérbio, pronome e conjunção. Já o texto em português, apresenta uma maior frequência de ocorrência da classe de substantivos, seguida pela de conjunção, verbo, adjetivo, advérbio e pronome.

A análise dos GRAF. 2 e GRAF. 3 para os textos em português mostra que há semelhanças entre o manual técnico e a cartilha e *folder* de divulgação em relação à frequência de ocorrência das classes de palavras, sendo a classe de substantivos mais frequente, seguida pela de conjunções, verbos, adjetivos, advérbios e pronomes.

Na cartilha e *folder* de divulgação, a frequência de ocorrência de pronomes (considerando apenas aqueles de primeira e segunda pessoas) para o inglês (5,9%) é maior quando comparada aos demais *corpora* nessa língua (2,2% nos artigos acadêmicos e 2,5% nos manuais técnicos). Essa frequência pode ser um indício da caracterização desse tipo de texto em inglês por mostrar as relações mais próximas entre os participantes da interação. Já para os textos em português, a frequência de ocorrência de pronomes na cartilha e *folder* de divulgação (2,4%) é mais elevada quando comparada aos artigos acadêmicos (1,6%) e menos elevada em relação aos manuais técnicos (2,8%). Esse resultado contribui para uma análise mais detalhada sobre esses dois tipos de texto fosse realizada com o aporte da Linguística Sistêmico-Funcional (*cf.* Subseção 3.2.3).

A FIG. 55 a seguir sintetiza os gráficos apresentados anteriormente em relação às classes de palavras por língua e por tipo de texto como forma de ilustrar os achados explicados anteriormente e a comparação inter-registros.

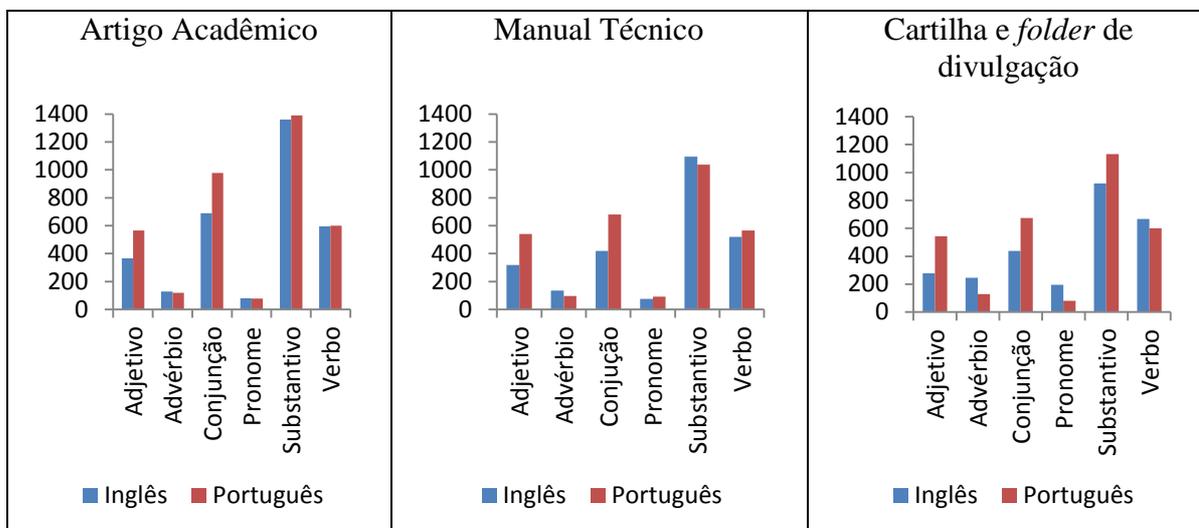


FIGURA 55. Condensado dos gráficos sobre classe de palavras por língua e por rótulos (artigos acadêmicos, manuais técnicos e cartilhas e *folders* de divulgação)

Nota-se, pela FIG. 55, que a classe de palavras que apresenta maior frequência de ocorrência é a de substantivos, seguida pela de verbos nos rótulos manual técnico e cartilha e *folder* de divulgação em inglês. Para artigo acadêmico em inglês e para os outros tipos de texto tanto em inglês quanto em português, a classe de substantivos vem seguida da classe de conjunções, verbos, adjetivos, advérbios e pronomes. Em relação às conjunções, observa-se uma maior frequência dessa classe nos tipos de texto em português. Nesta pesquisa, as conjunções foram analisadas por essa classe de palavras ser um indicador da organização da subvariável modo retórico (argumentação, persuasão, descrição, narração, exposição, instrução) de cada um dos tipos de textos por língua.

No que tange aos pronomes, observa-se que há maior frequência de ocorrência nos artigos acadêmicos e cartilhas e *folders* de divulgação em inglês. Já no manual técnico, a frequência de ocorrência de pronomes em português é mais elevada. Comparando-se a frequência de ocorrência dessa classe de palavras inter-registros, observa-se que há maior frequência de ocorrência na cartilha e *folder* de divulgação, o que pode sugerir indícios de caracterização para esse tipo de texto. É notável também na cartilha e *folder* de divulgação a frequência de ocorrência da classe de advérbios, que é maior em comparação a todos os outros tipos de texto.

Após verificar que, para os três tipos de textos (artigos acadêmicos, manuais técnicos e cartilhas e *folders* de divulgação), houve associação entre as línguas (português e inglês) e as

frequências das classes de palavras (usando-se o teste qui-quadrado), torna-se necessário identificar para quais casos a maior ocorrência de uma classe de palavras está significativamente associada com a língua em um dado tipo de texto. Para tanto, utilizou-se o teste Z de comparação de proporção, conforme explicado na Subseção 2.2.1.

As TAB. 6 a 8 ilustram as distribuições do teste Z para a frequência das classes das palavras presentes em cada um dos tipos de textos, considerando-se as línguas portuguesa e inglesa.

TABELA 6. Distribuições do teste Z para frequência das classes de palavras presentes nos artigos acadêmicos

Classe de palavras	Português		Inglês		P-valor Teste Z
	Frequência	Relação classe de palavras (%) no <i>corpus</i>	Frequência	Relação classe de palavras (%) no <i>corpus</i>	
Adjetivo	348	17,2%	317	14,5%	0,015*
Advérbio	87	4,3%	95	4,4%	1,000
Conjunção	164	8,2%	144	6,6%	0,061
Pronome	48	2,4%	67	3,7%	0,249
Substantivo	950	47,3%	1093	50,1%	0,072
Verbo	413	20,6%	467	21,39%	0,521
TOTAL	2010	100,0%	2183	100,0%	–

\* Significância: p-valor <0,05.

Pela TAB. 6, é possível verificar que apenas para a classe de palavras “adjetivo” houve diferença significativa entre as porcentagens de ocorrência considerando-se as línguas. Em outras palavras, houve maior porcentagem de adjetivos em língua portuguesa (17,2%) que em língua inglesa (14,5%), sendo essa a única diferença significativa entre as línguas no que toca ao artigo acadêmico. Esse resultado contribuiu para a análise dos processos relacionais atributivos, uma vez que a presença de muitos atributos (adjetivos) implica maior quantidade desse tipo de processo. Dentro da construção do significado da doença falciforme para cada um dos diferentes tipos de texto, os adjetivos podem indicar uma avaliação, uma opinião ou uma descrição dos participantes.

Pela TAB. 7 a seguir, referente ao manual técnico, observa-se que, para adjetivo, advérbio e substantivo, houve diferenças significativas entre as porcentagens de ocorrência nas duas línguas. Houve uma maior porcentagem de adjetivos (21,33%) nos textos em língua portuguesa em comparação aos de língua inglesa (13,83%). Já as classes de advérbio e

substantivo apresentaram maior porcentagem de ocorrência nos textos em língua inglesa (5,77% e 47,60%, respectivamente) que em língua portuguesa (4,08% e 41,72%, respectivamente).

TABELA 7. Distribuições do teste Z para frequência das classes de palavras presentes nos manuais técnicos

Classe de palavras	Português		Inglês		P-valor Teste Z
	Frequência	Relação classe de palavras (%) no <i>corpus</i>	Frequência	Relação classe de palavras(%) no <i>corpus</i>	
Adjetivo	460	21,33%	314	13,83%	<0,001*
Advérbio	88	4,09%	131	5,77%	0,012*
Conjunção	145	6,72%	159	7,00%	0,758
Pronome	84	3,89%	75	3,30%	0,329
Substantivo	900	41,72%	1081	47,60%	<0,001*
Verbo	480	22,25%	511	22,50%	0,872
TOTAL	2157	100,00%	2271	100,00%	–

\* Significância: p-valor <0,05.

Pela TAB. 8 a seguir, verifica-se que só para conjunção não houve diferença significativa entre as porcentagens de ocorrência nas duas línguas. Houve uma maior porcentagem de adjetivos e substantivos (20,38% e 42,67%) em português que em inglês (11,02% e 38,21%). Em contrapartida, houve maior porcentagem de advérbios, pronomes e verbos em inglês (10,19%, 7,82%, 27,32%, respectivamente) que em português (4,62%, 3,25%, 22,66%, respectivamente).

TABELA 8. Distribuições do teste Z para frequência das classes de palavras presentes nas cartilhas e *folders* de divulgação

Classe de palavras	Português		Inglês		P-valor Teste Z
	Frequência	Relação classe de palavras (%) no <i>corpus</i>	Frequência	Relação classe de palavras (%) no <i>corpus</i>	
Adjetivo	445	20,39%	251	11,02%	<0,001*
Advérbio	101	4,63%	232	10,19%	<0,001*
Conjunção	140	6,41%	124	5,45%	0,193
Pronome	71	3,25%	178	7,82%	<0,001*
Substantivo	932	42,67%	870	38,21%	0,003*
Verbo	495	22,66%	622	27,32%	<0,001*
TOTAL	2184	100,00%	2277	100,00%	–

\* Significância: p-valor <0,05.

Buscou-se entender como as palavras estão distribuídas ao longo dos textos na tentativa de identificar padrões de uso linguístico mais recorrentes e específicos que caracterizem cada um dos tipos de texto em português e em inglês. A tabulação dos dados gerais, que ilustram a configuração dos *corpora* desta pesquisa, possibilitou uma visão gráfica de cada um dos tipos de texto. Após analisadas as frequências de ocorrências de cada uma das classes de palavras, partiu-se para uma análise mais detalhada dos itens lexicais mais frequentes.

### 3.1.3 Análise de nódulos extraídos dos corpora em função da língua e do tipo de texto

Apresentam-se, a seguir, os resultados da análise dos nódulos “doença falciforme”, “anemia falciforme”, “*sickle cell disease*” e “*sickle cell anemia/anaemia*”. Em seguida, disponibilizam-se os dados com os nódulos com as palavras “dor”, “*pain*”, “paciente”, “*patient*”, “*you*”, “você”. Essas palavras foram selecionadas por serem frequentes. Na sequência, descrevem-se as combinações de palavras (*clusters*) mais frequentes, selecionados a partir dos nódulos analisados. Esses exemplos de *clusters* foram analisados visando mostrar como esses nódulos são distribuídos nos diferentes tipos de texto. Entende-se por *cluster* a sequência fixa de palavras, como, por exemplo, “tem por objetivo” e “sabe-se que” (SINCLAIR, 1991).

### 3.1.3.1 Análise dos nódulos “doença falciforme” e “anemia falciforme” “*sickle cell disease*” e “*sickle cell anemia/anaemia*”

#### 3.1.3.1.1 Análise dos nódulos nos *corpora* em português

Os primeiros nódulos analisados foram “doença falciforme” e “anemia falciforme” para os *corpora* em português. A partir da TAB. 9, observa-se a frequência de ocorrência desses nódulos em cada um dos tipos de texto.

TABELA 9. Frequência de ocorrência dos nódulos doença falciforme e anemia falciforme para cada um dos *corpora* em português

Rótulo	Doença falciforme	% Doença falciforme/tokens	Anemia falciforme	% Anemia falciforme/tokens
Artigos acadêmicos	5	0,11	17	0,33
Manuais técnicos	9	0,25	1	0,02
Cartilhas e <i>folders</i> de divulgação	18	0,48	5	0,13

Esses dados permitem algumas considerações para análise específica em cada um dos *corpora*:

#### 1) Artigos acadêmicos em português

Há cinco ocorrências do nódulo “doença falciforme” e 17 ocorrências do nódulo “anemia falciforme” (*cf.* TAB. 9). O nódulo “doença falciforme” representa 0,11% do número de *tokens*, ao passo que o nódulo “anemia falciforme” representa 0,33%. Manuais técnicos em português.

#### 2) Manuais técnicos em português

Há nove ocorrências do nódulo “doença falciforme” e uma ocorrência do nódulo “anemia falciforme” nesse *corpus*, o qual corresponde a 0,02% do total de *tokens*. Por conseguinte, a ocorrência desse nódulo é significativamente menor em comparação à ocorrência de “doença falciforme”, que equivale a 0,25% (*cf.* TAB. 9). Alguns *clusters* encontrados foram: “pessoas com a doença”, “ocorrer na doença”, “complicações da doença”.

### 3) Cartilhas e *folders* de divulgação em português

Nesse *corpus*, há 18 ocorrências do nódulo “doença falciforme” e cinco ocorrências do nódulo “anemia falciforme”. Esse *corpus* apresenta o maior número de ocorrência do nódulo (0,48%) em comparação aos outros dois. A ocorrência desse nódulo é significativamente maior em comparação à ocorrência de “anemia falciforme”, que equivale a 0,13% (cf. TAB. 9) nesse mesmo tipo de texto em português. Alguns *clusters* foram: “paciente com”, “portador de”, “criança com”.

### 4) Consolidação da ocorrência dos nódulos “doença falciforme” e “anemia falciforme” para cada um dos *corpora* em português

A seguir são apresentados os dados numéricos, os testes estatísticos aplicados para testar a significância dos dados obtidos em relação a cada um dos tipos de texto em português. Para verificar se há associação significativa dos nódulos “doença falciforme” e “anemia falciforme” com a língua e os tipos de texto, realizou-se o teste  $\chi^2$ . Para os *corpora* em português, houve uma associação estatisticamente significativa,  $\chi^2$  (gl=2) = 23,42,  $p < 0,001$ .

Os resultados indicam que o termo “anemia falciforme” é significativamente mais utilizado nos artigos científicos, em detrimento aos demais tipos de texto (cf. GRAF. 4). Tal situação pode ser atribuída ao conhecimento sobre o assunto por parte dos pesquisadores, pois são capazes de diferenciar anemia falciforme de doença falciforme, uma vez que, conceitualmente, esses termos não são sinônimos. A mesma afirmação pode ser feita em relação a “*sickle cell disease*” e “*sickle cell anemia/anaemia*” (cf. SERJEANT, B.; SERJEANT, G., 1991, p. 31).

No caso dos manuais técnicos, há uma maior frequência do termo “doença falciforme”. Esperava-se, contudo, encontrar uma maior ocorrência de “anemia falciforme” devido ao papel agentivo dos autores desse tipo de texto, ou seja, detentores e divulgadores do conhecimento.

O GRAF. 4 a seguir sintetiza a frequência dos nós considerando-se cada um dos tipos de texto

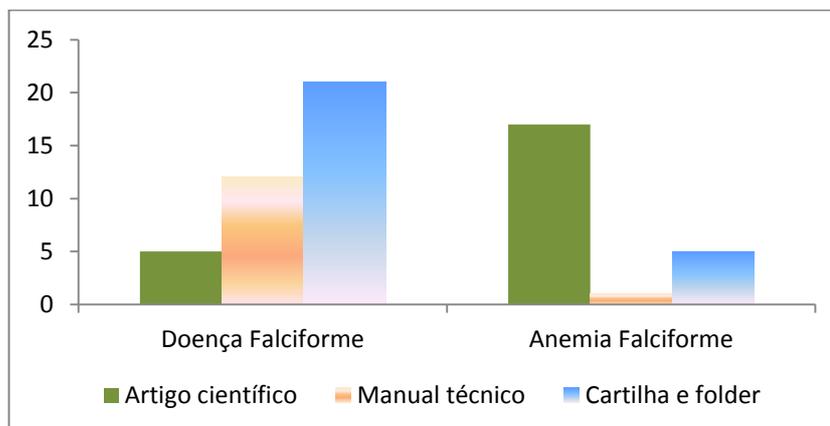


GRÁFICO 4. Associação das frequências absolutas dos nós “doença falciforme” e “anemia falciforme” em português

A análise estatística do GRAF. 4 indica que os nós “doença falciforme” e “anemia falciforme” não possuem proporções semelhantes nos respectivos tipos de textos. Essa análise sugere que a probabilidade de os nós em questão ocorrerem no texto em proporções semelhantes nos respectivos tipos de textos é menor que 0,001 ( $p < 0,05$ ).

### 3.1.3.1.2 Análise dos nós nos *corpora* em inglês

Para a análise dos *corpora* em inglês, foram identificados os nós “*sickle cell disease*” e “*sickle cell anemia/anaemia*”. A TAB. 10 exibe a frequência de ocorrência desses nós em cada um dos tipos de texto.

TABELA 10. Frequência de ocorrência dos nós “*sickle cell disease*” e “*sickle cell anemia/anaemia*” para cada um dos *corpora* em Inglês

Rótulo	<i>Sickle cell disease</i>	% <i>Sickle cell disease</i> /tokens	<i>Sickle cell anemia/anaemia</i>	% <i>Sickle cell anemia [anaemia]</i> /tokens
Artigos acadêmicos	2	0,05	2	0,05
Manuais técnicos	17	0,55	3	0,09
Cartilhas e <i>folders</i> de divulgação	26	0,80	7	0,21

Os dados da TAB. 10 possibilitam considerações para análise específica em cada um dos *corpora* em inglês:

### 1) Artigos acadêmicos em inglês

Há duas ocorrências do nóculo “*sickle cell disease*” e duas ocorrências do nóculo “*sickle cell anemia/anaemia*” nos artigos acadêmicos. Para esse *corpus*, não há, portanto, diferença entre a frequência de ocorrência dos nósculos, que representam 0,05% do número de *tokens* cada. Os *clusters* identificadas foram: “*patients with SCD*” e “*adolescents with SCD*”.

### 2) Manuais técnicos em inglês

Há 17 ocorrências do nóculo “*sickle cell disease*” e três ocorrências do nóculo “*sickle cell anemia/anaemia*” nesse *corpus*. Há uma predominância do primeiro nóculo, correspondente a 0,55% do total de *tokens* (contra 0,09% para o nóculo “*sickle cell anemia/anaemia*”). Alguns *clusters* foram encontradas foram: “*associated with*”, “*patients with*”, “*forms of*” e “*management of*”.

### 3) Cartilhas e *folders* de divulgação em inglês

Nesse *corpus*, há 26 ocorrências do nóculo “*sickle cell disease*” e sete ocorrências do nóculo “*sickle cell anemia/anaemia*”. A ocorrência do primeiro nóculo é maior, correspondendo a 0,80% do total de *tokens*, ao passo que o segundo nóculo corresponde a 0,21%. Em comparação com os outros dois *corpora* em inglês, esse apresenta o maior número de ocorrências de “*sickle cell disease*”. Algumas combinações de palavras são: “*types of*”, “*blood cells in*”, “*commonly seen*”, “*your child has*”, “*form of*”, “*child with*”.

### 4) Consolidação da ocorrência dos nósculos “*sickle cell disease*” e “*sickle cell anemia/anaemia*” para cada um dos *corpora* em inglês.

Para verificar a associação dos nósculos em inglês com o tipo de texto, foi aplicado o teste exato de Fisher. Contudo, para esses *corpora*, não foi observada uma associação estatisticamente significativa ( $p = 0,2457$ ).

Os resultados indicaram que o nóculo “*sickle cell disease*” foi mais utilizado tanto nos manuais técnicos quanto nas cartilhas e *folders* de divulgação. O GRAF. 5 a seguir ilustra a frequência dos nósculos nos tipos de textos em inglês.

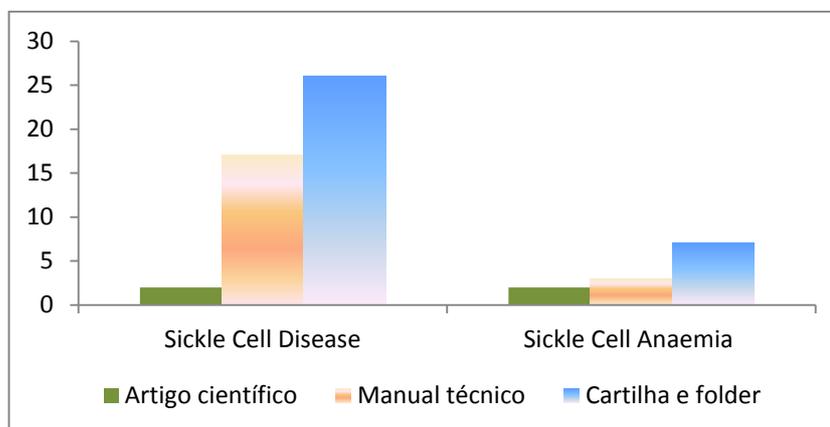


GRÁFICO 5. Associação das frequências absolutas dos nódulos em inglês

Esse resultado (cf. GRAF. 5) sugere que os nódulos “*sickle cell disease*” e “*sickle cell anemia/anaemia*” não estão distribuídos de forma semelhante nos respectivos tipos de textos. O p-valor é igual a 0,2457, o que sugere que esses nódulos, nesse caso, não sejam frequentes em proporções semelhantes nos textos em inglês.

Nota-se que, para os tipos de texto em português, há associação estatisticamente significativa dos nódulos “doença falciforme” e “anemia falciforme” com a língua e os rótulos artigo acadêmico, manual técnico e cartilha e *folder* de divulgação. Já para os nódulos “*sickle cell disease*” e “*sickle cell anemia/anaemia*”, não há associação significativa com a língua e os tipos de textos.

### 3.1.3.2 Análise dos nódulos “dor”, “pain”, “paciente”, “patient” e “you”

Os nódulos em português e inglês analisados a seguir são pertencentes à classe de palavras de substantivos (“dor”, “*pain*”, “paciente” e “*patient*”) e de pronomes (“*you*”), por serem frequentes em todos os tipos de textos em suas respectivas línguas. A análise de cada um desses nódulos buscou verificar se há associação significativa entre eles com a língua e o tipo de texto, além de identificar os principais *clusters* que ocorrem a partir desses nódulos.

A TAB. 11 fornece a frequência de ocorrência do nódulo “dor” e “*pain*” em cada um dos tipos de texto em português e inglês, respectivamente.

TABELA 11. Frequência de ocorrência dos nódulos “dor” (português) e “*pain*” (inglês) por tipos de texto

Rótulo	Português		Inglês	
	Dor	% dor/tokens	<i>Pain</i>	% <i>pain</i> /tokens
Artigos acadêmicos	24	0,56	79	2,08
Manuais técnicos	38	1,08	30	0,97
Cartilhas e <i>folders</i> de divulgação	28	0,75	12	0,37

Os dados da TAB. 11 possibilitam considerações para análise específica em cada um dos *corpora* em português e em inglês. O teste  $\chi^2$  sugere uma associação significativa dos nódulos “dor” e “*pain*” com tipo de texto. O nódulo “*pain*” ocorre com mais frequência em artigo acadêmico (79 ocorrências), ao passo que “dor”, é mais frequente em manual técnico (38 ocorrências). A frequência elevada desses dois nódulos enseja que os textos constroem significados sobre a existência da dor na doença falciforme. Os principais *clusters* foram: “dias com dor”, “da dor da doença”, “crise de dor”, “*SCD related pain*”, “*days with pain*”.

O GRAF. 6 a seguir mostra a frequência absoluta dos nódulos “*pain*” e “dor” nos tipos de texto em inglês e em português.

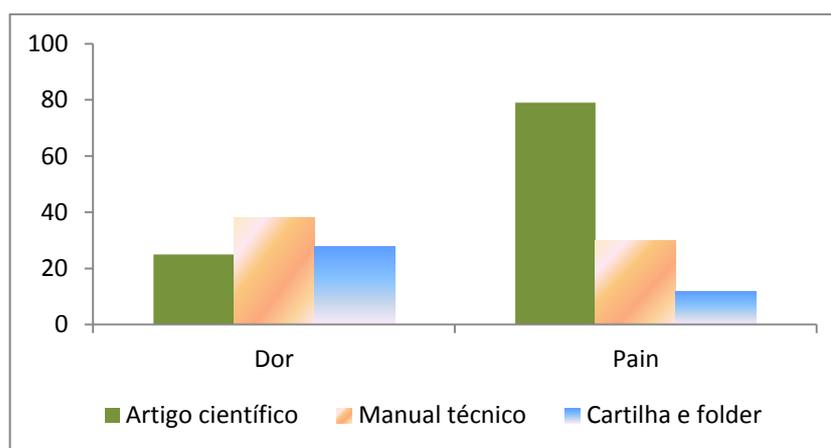


GRÁFICO 6. Associação das frequências absoluta dos nódulos “dor” e “*pain*” por tipo de texto

O GRAF. 6 mostra que os nódulos “dor” e “*pain*” não estão distribuídos nos textos de forma semelhante, pois o p-valor é  $< 0,001$ , o que sugere que os nódulos, neste caso, não guardam proporções semelhantes no artigo acadêmico, manual técnico e cartilha e *folder* de divulgação. Pagano e Figueredo (2011) realizaram um trabalho sobre as estruturas nas quais é usado o lema “dor” (em português) e “*dolor*” (em espanhol) com enfoque em *corpora*

compilados com textos sobre o assunto extraídos da Internet. Esse trabalho teve por objetivo mapear os recursos léxico-gramaticais são utilizados para se construir significado sobre a dor em cada um dessas línguas em três diferentes tipos de textos (Fórum, Informativo e Entrevista), considerando o processo sociosemiótico COMPARTILHAR. Esses autores concluíram que os textos em português tendem a usar nominalizações do lema “dor”, o que possibilita construir significados de como medir, calcular, prever ou curar essa dor nas diferentes interações entre os falantes. Para esta tese, procede-se a uma análise mais detalhada para verificar a ocorrência de nominalizações e entender como a dor na doença falciforme é construída nas interações entre os participantes, considerando-se os diferentes tipos de texto, em inglês e em português.

Os dados para a análise da frequência de ocorrência dos nódulos “pacientes” e “*patients*” estão ilustrados na TAB. 12, a seguir.

TABELA 12. Frequência de ocorrência dos nódulos “pacientes” e “*patients*” em cada um dos tipos de texto

Rótulo	Português		Inglês	
	Pacientes	% pacientes/ <i>tokens</i>	<i>Patients</i>	% <i>patients/tokens</i>
Artigos acadêmicos	43	0,96	41	1,08
Manuais técnicos	29	0,82	21	0,68
Cartilhas e <i>folders</i> de divulgação	40	1,08	12	0,37

A TAB. 12 mostra que os nódulos “pacientes” e “*patients*” ocorrem com maior frequência nos artigos científicos (43 e 41 ocorrências, respectivamente), seguidos do manual técnico (29 e 21 ocorrências, respectivamente) e, por fim, da cartilha e *folder* de divulgação (40 e 12 ocorrências, respectivamente). As combinações de palavras encontradas foram para o nódulo em inglês: “*sickle cell disease patients*” e “*patients with sickle cell disease*”.

O GRAF. 7 a seguir mostra a frequência absoluta dos nódulos “pacientes” e “*patients*” nos textos em inglês e em português.

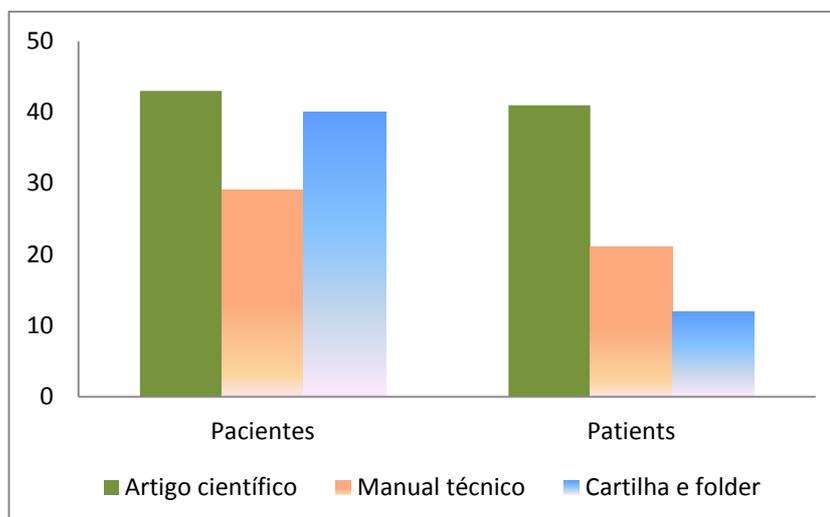


GRÁFICO 7. Associação das frequências dos nós “pacientes” e “patients” por tipo de texto

O GRAF. 7 aponta que os dados não apresentam frequências proporcionais dos nós nos respectivos tipos de texto, pois o p-valor é  $< 0,001$ . Esse resultado sugere que os nós não guardam proporções semelhantes nos artigos acadêmicos, manuais técnicos e cartilhas e *folders* de divulgação.

Por fim, como um exemplo da classe de palavras de pronomes, foram analisados como nós os pronomes pessoais “você” e “you” (cf. TAB. 13). Não são encontradas ocorrências em português, ao passo que são identificadas 13 ocorrências na cartilha e *folder* de divulgação em inglês.

TABELA 13. Frequência de ocorrência dos nós “você” e “you” em cada um dos tipos de texto

Rótulo	Português		Inglês	
	você	% você/tokens	you	% you/tokens
Artigos acadêmicos	-		-	
Manuais técnicos	-		-	
Cartilhas e <i>folders</i> de divulgação	-		13	0,40

Terminadas as discussões iniciais dos dados obtidos por meio das ferramentas computacionais, apresentam-se a seguir as análises à luz da Linguística Sistêmico-Funcional, para se verificar como se dá o processo de construção do significado dos três tipos de texto que compõem os *corpora* desta pesquisa.

Analisam-se padrões de uso linguístico para identificar indícios que permitam caracterizar os textos. A Seção 3.2 retrata a interpretação dos diferentes tipos de textos com base nas variáveis do contexto de situação (*i.e.*, campo, sintonia e modo) e seus respectivos parâmetros dos processos sociosemióticos EXPLORAR e HABILITAR.

### **3.2 Caracterização dos Textos sob a Perspectiva da Linguística Sistêmico-Funcional**

Conforme explicado na Seção 1.3, o registro compreende uma variação linguística que ocorre em função do contexto de situação, sendo modelado a partir de generalizações de tipos de texto em contexto de cultura. Com essa perspectiva, esta pesquisa se baseou na proposta de Eggins (2004) de investigar variáveis campo (relacionada ao tipo de ação social em uma dada situação), sintonia (relacionada ao tipo de relação entre os participantes na situação) e modo (relacionada ao modo de como a mensagem é realizada) para apontar as diferenças e semelhanças entre os diferentes tipos de texto.

Em todos os textos, tanto em português quanto em inglês, o autor dos textos é detentor do conhecimento ao elaborá-los com distintas configurações da sintonia, objetivando criar significados para a construção de negociação de posições e valores com participantes inseridos na comunidade ou habilitar esses participantes a fazer alguma atividade seja por meio de instruções sobre como fazer ou por regras sobre o que se deve ou não fazer. Os tipos de texto analisados neste estudo foram classificados a partir da tipologia de textos (MATTHIESSEN; TERUYA; WU, 2008), conforme sintetizado nos QUADROS 9 a 11, a seguir.

O QUADRO 9 mostra a classificação do artigo acadêmico baseada nos parâmetros das variáveis dentro do processo sociosemiótico EXPLORAR. Esse processo é responsável por instanciar significados para a construção de negociação de posições e valores com outros participantes dentro da comunidade científica por meio de pesquisas, por exemplo.

QUADRO 9. Classificação dos textos conforme a convergência do processo sociosemiótico EXPLORAR com as variáveis campo, sintonia e modo

Variáveis	Tipos de texto/registo
	Artigos acadêmicos (interação especialista – especialista)
Campo	Domínio experiencial: medicina - anemia falciforme - triagem neonatal.
	Orientação/propósito: relatos de pesquisas empíricas.
Sintonia	Atividade social: como registo associado ao processo sociosemiótico EXPLORAR, envolve a investigação, visando compartilhamento público de aspectos e valores sociais e ainda envolve comparar visões distintas e argumentar em favor de uma ou algumas delas.
	Papel agentivo do autor do texto: autor da pesquisa e detentor do conhecimento.
	Papel agentivo do leitor do texto: indivíduo da área da saúde em geral, não detentor do conhecimento da pesquisa relatada; ou pares (tendo em vista a especificidade do texto), detentor de conhecimento análogo ao da pesquisa relatada.
	Papel discursivo do autor do texto: fornecedor de informações.
	Papel discursivo do leitor do texto: receptor de informações (leitores da área da saúde em geral), crítico em relação às informações (pares).
	Papel social do autor: médico, pesquisador, professor (pessoa com autoridade e <i>expertise</i> no assunto).
	Papel social do leitor: médico, pesquisador, professor, aluno, profissional da saúde (pessoa com autoridade e <i>expertise</i> no assunto análogas ou menores que as do autor).
	Distância social entre autor e leitor: distante, porém assumida como ambos pertencendo à mesma subcomunidade acadêmica (no caso de leitores da área da saúde em geral) ou a subcomunidades próximas, no caso de pares.
Modo	Posicionamento do autor em relação ao leitor (afeto e valoração): O autor, enquanto detentor do conhecimento, elabora um artigo acadêmico para que seja apreciado pelos seus pares ou outros pesquisadores e, conseqüentemente, ser aceito na comunidade acadêmica, por meio de mecanismos de ênfase e atenuação.
	Papel na divisão de tarefas: constitutivo.
	Canal: gráfico.
	Meio: escrito.
	Modo de interação: monólogo.
	Modo retórico: argumentação, exposição.

Os textos classificados como manuais técnicos e cartilhas e *folders* de divulgação referem-se a textos associados ao processo sociosemiótico HABILITAR (*cf.* QUADRO 10), em que a variável sintonia é essencial para as trocas de informação. O autor do texto estabelece as orientações para o leitor, ao qual cabe entendê-las e executar o que está sendo instruído. Assim, o leitor espera receber a informação que o prepare para a realização da tarefa. Esse

processo é orientado para a sintonia (cf. FIG. 14), ou seja, para as relações estabelecidas entre o autor e o leitor em um dado contexto.

QUADRO 10. Classificação dos textos conforme a convergência do processo sociossemiótico HABILITAR com as variáveis campo, sintonia e modo

Variáveis	Tipo de texto/registro
	Manuais técnicos (interação especialista – técnico da área da saúde)
Campo	Domínio experiencial: medicina - anemia falciforme - pacientes diagnosticados pela triagem neonatal.
	Orientação/propósito: instruções para realização de tratamento a pacientes com doença falciforme.
	Atividade social: associado ao processo sociossemiótico habilitar, envolve desenvolver a capacidade do interlocutor de executar uma tarefa (orientar ações dos profissionais de saúde).
Sintonia	Papel agentivo do autor do texto: detentor e divulgador do conhecimento.
	Papel agentivo do leitor do texto: não detentor do conhecimento relativo às informações em tela, porém interessado em realizar procedimentos de saúde ou inserido em um meio onde esses procedimentos são necessários/indispensáveis.
	Papel discursivo do autor do texto: fornecedor de informações.
	Papel discursivo do leitor do texto: receptor de informações, executor das ações.
	Papel social do autor: médico pesquisador, professor, profissional pesquisador da área da saúde com autoridade e <i>expertise</i> maiores que o leitor (mesmo quando não tem seu nome explicitamente divulgado, estando o manual em nome de uma instituição tida como de referência e autoridade no assunto, no caso específico do Ministério da Saúde, por exemplo).
	Papel social do leitor: técnico da área da saúde com autoridade e <i>expertise</i> menores que o autor
	Distância social entre autor e leitor: distante, porém assumida enquanto uma interação especialista - técnico da área da saúde.
Posicionamento do autor em relação ao leitor (afeto e valoração): O autor, enquanto detentor do conhecimento, elabora um manual técnico para que seja utilizado como base para as ações do técnico da área da saúde no que toca aos procedimentos para a triagem neonatal e ao acompanhamento e auxílio no tratamento de pacientes diagnosticados com doença falciforme. Os procedimentos devem ser seguidos conforme proposto pelo autor.	
Modo	Papel na divisão de tarefas: constitutivo.
	Canal: gráfico.
	Meio: escrito.
	Modo de interação: monólogo.
	Modo retórico: instrução.

O QUADRO 11 a seguir apresenta informações sobre as cartilhas e *folders* de divulgação, que também estão inseridas no processo sociossemiótico HABILITAR.

QUADRO 11. Classificação dos textos conforme as variáveis campo, sintonia e modo na convergência do processo sociosemiótico HABILITAR

Variáveis	Tipos de texto/registo
	Cartilhas e <i>folders</i> de divulgação (interação especialista - leigo)
Campo	Domínio experiencial: medicina, triagem neonatal para anemia falciforme.
	Orientação/propósito: transmissão de informações na forma de orientações e instruções ao público geral.
	Atividade social: associado ao processo sociosemiótico habilitar, envolve desenvolver a capacidade do interlocutor de executar uma tarefa (neste caso, pais e familiares do doente).
Sintonia	Papel agentivo do autor do texto: detentor e divulgador do conhecimento.
	Papel agentivo do leitor do texto: não detentor do conhecimento das informações, mas por alguma razão interessado no conhecimento divulgado ( <i>e.g.</i> , pais de pacientes com anemia falciforme).
	Papel discursivo do autor do texto: fornecedor de informações.
	Papel discursivo do leitor do texto: receptor de informações.
	Papel social do autor: médico pesquisador, enfermeiro pesquisador, professor, profissional da área da saúde (assumido como detentor de <i>autoridade</i> e <i>expertise</i> maiores que as do leitor, mesmo quando seu nome não é explicitamente veiculado, estando o material divulgado em nome ou sob o nome de uma instituição reconhecida como autoridade ou referência no assunto).
	Papel social do leitor: leigo no assunto.
	Distância social entre autor e leitor: distante.
Modo	Posicionamento do autor em relação ao leitor (afeto e valoração): O autor, enquanto detentor do conhecimento, elabora cartilhas e <i>folders</i> de divulgação para que sejam utilizados como base de informação sobre a doença falciforme e a triagem neonatal para essa doença.
	Papel na divisão de tarefas: constitutivo.
	Canal: gráfico.
	Meio: escrito.
	Modo de interação: monólogo.
	Modo retórico: instrução.

As Subseções 3.2.1 a 3.2.3, a seguir, enfocam os padrões de uso linguístico dos distintos textos em português e em inglês que compuseram os *corpora* desta pesquisa (*i.e.*, artigo acadêmico; manual técnico, cartilhas e *folders* de divulgação) com base nos parâmetros das três variáveis do contexto de situação: campo, sintonia e modo. As análises se pautam nos nódulos identificados a partir das classes de palavras substantivo, pronomes e conjunções.

### 3.2.1 A variável CAMPO

O campo do discurso descreve a natureza da atividade social (HALLIDAY; HASAN, 1989, p. 56), na qual os significados experienciais são evidenciados em uma situação com vistas a um objetivo a ser alcançado por meio da linguagem. Esses significados estão inclusos em padrões de escolhas, sendo que na abordagem sistêmico-funcional são estabelecidas duas dimensões do uso da língua: (i) a identificação das escolhas lexicais e gramaticais, que gera a descrição do sistema linguístico; e (ii) a percepção da função dessa escolha, que permite descrever como a língua é usada nos contextos sociais para atingir diferentes objetivos culturais.

No caso desta pesquisa, a análise foi realizada a partir dos substantivos selecionados como nódulos para a identificação dos padrões de uso linguístico mais recorrentes. Essa identificação implica a caracterização dos tipos de textos em sua relação com o contexto de situação, observando como os participantes constroem os significados que estão sendo negociados. Neumann (2008, p. 42) sublinha que “os itens lexicais mais frequentes de um *corpus* são indicadores do domínio experiencial, pois refletem o assunto de um texto específico e, mais especificamente, do registro”<sup>32</sup>.

A TAB. 14 ilustra os 20 substantivos mais frequentes provenientes dos *corpora* compilados a partir dos textos em português. Esses substantivos são passíveis de análise para se entender como os textos retratam o assunto conforme o parâmetro “orientação/propósito” estabelecido pela variável campo.

---

<sup>32</sup> Minha tradução de: “(...) the most frequent lexical items, are indicators of the experiential domain since they reflect the subject matter of a specific text and, more generally, of the respective register.” (NEUMANN, 2008)

TABELA 14. Frequência de ocorrência de substantivos em português por tipo de texto

Qte./ Classe	Artigo acadêmico		Manual técnico		Cartilhas e <i>folders</i> de divulgação	
	Substantivo	Frequência	Substantivo	Frequência	Substantivo	Frequência
1	pacientes	28	dor	36	pacientes	28
2	estudo	27	doença	25	dor	27
3	doença	26	pacientes	15	doença	27
4	anos	24	Hb	15	Hb	17
5	dor	23	tratamento	13	hemoglobina	16
6	anemia	20	anos	11	portadores	13
7	Hb	18	transfusão	9	anemia	13
8	dias	14	idade	9	febre	12
9	causa	14	pessoa	8	crises	11
10	paciente	13	paciente	8	casos	11
11	internação	13	infecções	8	aumento	11
12	idade	13	hemácias	8	sintomas	10
13	hábitos	12	horas	7	crianças	10
14	internações	11	hidratação	7	paciente	9
15	crises	11	causa	7	vezes	8
16	crianças	11	uso	6	diagnóstico	8
17	vida	10	saúde	6	crise	8
18	portadores	10	redução	6	anos	8
19	manifestações	10	crianças	6	transfusão	7
20	média	10	complicações	6	situações	7
Total	–	318	–	216	–	261

A análise da TAB. 14 remete a itens lexicais como “pacientes”, “dor”, “doença”, “internação”, “febre”, “transfusão”, “anemia”, “hemoglobina”, “sintomas”, “crianças” e “saúde”, os quais auxiliam no reconhecimento do campo, ou seja, o assunto abordado nos três tipos de textos analisados. Como mencionado anteriormente, os parâmetros *domínio experiencial* e *atividade social* da variável campo apresentam características comuns entre os três tipos de texto e não são alvos de análise nesta pesquisa. Já o parâmetro *orientação/propósito* apresenta diferenças que merecem ser destacadas e discutidas, uma vez que cada um dos textos é construído com objetivos diferentes. Vale lembrar que essas mesmas assertivas podem ser consideradas para os *corpora* em inglês, analisados mais adiante.

No caso do artigo acadêmico em português, os autores dos textos (executores da pesquisa e detentores do conhecimento) têm o papel de apresentar e comparar visões distintas e argumentar em favor de uma ou algumas delas. A publicação, por esses autores, de relatos de pesquisas empíricas, dentre outras, é considerada uma atividade social reconhecida pelos próprios pares e por leitores da área de saúde em geral como uma forma de divulgação de conhecimentos sobre esse contexto de situação (como mostram os substantivos “pacientes”, “portadores”, “manifestações”, “causa” e “idade”). Nota-se, pelos dados tabulados na TAB. 14, que substantivos como “estudo” e “média” remetem o leitor a informações sobre possíveis produções acadêmicas, sendo esses substantivos encontrados apenas nesse tipo de texto dentre os analisados.

No que toca ao manual técnico, em que a interação se dá entre especialistas e técnicos da área da saúde, a atividade social dos autores (detentores e divulgadores do conhecimento) visa orientar as ações de atuação desses profissionais de saúde. Nesse tipo de texto, os autores produzem instruções para a realização de tratamento de pacientes com doença falciforme, como mostram os dados na TAB. 14: “dor”, “uso”, “horas” e “tratamento”.

Já no caso da cartilha e *folder* de divulgação, os substantivos podem ser atribuídos à suposta intenção dos autores (detentores e divulgadores do conhecimento) com os leitores (que não detêm o conhecimento das informações, porém estão interessados na informação divulgada nesse tipo de texto). Mais explicitamente, trata-se da intenção de orientar as ações de atuação do leitor (familiar do paciente portador de anemia falciforme). Os dados da TAB. 14 apontam que o objetivo do autor é orientar o público em geral sobre a anemia falciforme.

Em outras palavras, conforme assume Eggins (2004), as escolhas ideacionais e interpessoais apontam a forma de construção do significado referente à anemia falciforme nos diferentes tipos de texto, levando em consideração a densidade lexical, que aumenta à medida que a língua se torna mais especializada. Quanto mais lexicalmente densa a construção do significado, maior é o impacto na sintonia, pois a língua mais densa mostra um distanciamento entre os participantes. Observou-se, neste estudo, que o artigo acadêmico apresenta uma maior densidade lexical quando comparado aos outros dois textos tanto em inglês quanto em português. Esse fato faz com o artigo acadêmico seja menos acessível para o público mais leigo em comparação ao manual técnico e às cartilhas e *folders* de divulgação.

A TAB. 15 apresenta os dados referentes à frequência de ocorrência de substantivos nos textos em inglês.

TABELA 15. Frequência de ocorrência de substantivos em inglês por tipo de texto

Qte./ Classe	Tipos de texto					
	Artigo acadêmico		Manual técnico		Cartilha e <i>folder</i> de divulgação	
	Substantivo	Frequência	Substantivo	Frequência	Substantivo	Frequência
1	pain	72	sickle	61	sickle	68
2	children	44	cell	38	cell	55
3	sickle	35	haemoglobin	34	haemoglobin	52
4	days	32	pain	29	child	45
5	patients	34	disease	28	disease	39
6	episodes	30	trait	16	blood	24
7	diary	22	Hb	15	gene	21
8	study	18	beta	12	cells	19
9	care	17	care	12	children	16
10	school	14	crisis	11	thalassaemia	13
11	number	14	patient	10	anaemia	13
12	management	14	globin	10	trait	11
13	studies	13	blood	10	spleen	11
14	home	13	patients	8	doctor	10
15	use	11	complications	8	body	10
16	treatment	11	state	7	problems	8
17	review	11	nurses	7	parent	8
18	disease	11	gene	7	pain	8
19	diaries	11	example	7	health	8
20	complications	11	signs	6	crisis	7
Total		438		336		446

Os dados da TAB. 15 indicam “*pain*”, “*sickle*”, “*disease*”, “*child*”, “*patient*”, “*trait*”, “*anaemia*” e “*care*” como substantivos frequentes nos três tipos textos. Esses substantivos também remetem o leitor ao campo área da saúde, medicina, subárea triagem neonatal para doença falciforme.

Levando-se em consideração as diferenças específicas em cada um dos tipos de textos em inglês no que toca ao parâmetro *orientação/propósito* da variável campo, nota-se que, no artigo acadêmico, a forma de divulgação do conhecimento pode ser indicada pelos usos de

“patients”, “pain”, “children”, “management” e “treatment”. Os substantivos “review” e “studies” sugerem indícios da especificidade desse tipo de texto.

No que toca ao manual técnico, há interação entre especialistas e técnicos da área da saúde. O substantivo “nurses”, por exemplo, aparece sete vezes nesse tipo de texto, o que revela claramente essa interação. Nota-se que, se o autor desse tipo de texto não for um detentor do conhecimento ou pertencente à academia, usará termos que remetem às autoridades acadêmicas, como, por exemplo, o uso dos substantivos “nurses” ou “doctors”. Esse fato indica como a variável sintonia impactará no campo. Conforme já explicado, o texto, no manual técnico, designa procedimentos para a realização de acompanhamento (“care”) a pacientes com doença falciforme, sobretudo em casos de dor (“pain”) e crises (“crisis”).

Em relação à cartilha e *folder* de divulgação em inglês, o objetivo dos autores de transmitir informações sobre o contexto de situação na forma de orientação pode ser identificado pelos usos de “doctor”, por exemplo. A análise do substantivo “doctor” é importante por estabelecer um distanciamento entre o texto e o leitor. Em um artigo acadêmico, o autor escreve e dialoga com os respectivos pares, o que explica a não ocorrência dessa palavra nesse tipo de texto. No manual técnico há uma ocorrência de “doctor”, ao passo que na cartilha e *folder* de divulgação aparecem 10 ocorrências. Esse dado indica a forma como a mudança da sintonia impacta no campo, mais especificamente na forma como o autor da cartilha se posiciona em relação aos leitores do texto, sendo que esse autor possivelmente não é uma autoridade no assunto. No caso do manual técnico, o leitor a quem são escritas as instruções, é responsável pela execução das ações sugeridas por uma autoridade no assunto do texto. Em suma, as TAB. 14 e TAB. 15 remetem o leitor às listas dos 20 substantivos mais frequentes por tipo de texto e por língua. A análise das duas tabelas indica a disparidade na frequência de ocorrência desses substantivos. Essa disparidade gera indícios para a caracterização dos diferentes tipos de texto quando tais substantivos são usados como nódulos em busca por padrões recorrentes por tipo de texto e por língua.

A Subseção 3.2.2 a seguir apresenta os dados analisados sob a perspectiva da variável sintonia.

### 3.2.2 A variável SINTONIA

A variável sintonia descreve as relações humanas construídas na interação. Os papéis agentivo, discursivo e social dos autores são definidos em relação aos papéis dos leitores, além de mostrar a distância social entre esses agentes, bem como o posicionamento dos autores em relação aos leitores (afeto e valorização). Cumpre ressaltar que os tipos de texto compilados são pertencentes aos processos sociossemióticos EXPLORAR (*i.e.*, artigo acadêmico) e HABILITAR (*i.e.*, manual técnico e cartilha e *folder* de divulgação), os quais, em conjunto com os processos RECOMENDAR (*recommending*), FAZER (*doing*) e COMPARTILHAR (*sharing*) e EXPLICAR (*expounding*) fazem parte dos processos orientados para a variável sintonia. Essa variável é indicativa de trocas de mercadorias (bens de serviço e informações) negociadas no contexto de situação entre os autores e os leitores.

A forma como essas relações foram estabelecidas em cada tipo de texto foi analisada a partir da interação dos participantes, observando-se pronomes identificados em cada um dos *corpora*. A análise da classe de pronomes indica como o papel do autor é definido em relação ao seu leitor e em conformidade com o que está sendo comunicado e negociado em termos de construção de sentido. A distância entre autor/leitor é observada a partir das interações, por meio da presença ou não de pronomes pessoais ao longo dos textos.

Na TAB. 16 estão tabulados todos os pronomes pessoais do caso reto que ocorreram em cada um dos tipos de texto em português.

TABELA 16. Frequência de ocorrência de pronomes pessoais em português

N./ Classe	Tipos de texto					
	Artigo acadêmico		Manual técnico		Cartilha e <i>folder</i> de divulgação	
	Pronome	Frequência	Pronome	Frequência	Pronome	Frequência
1	ela	5	ela	3	elas	3
2	–		–		ele	1
Total	–	5	–	3	–	4

A análise da TAB. 16 mostra que não houve ocorrência do pronome “você” em nenhum dos três tipos de texto. Apesar de não ocorrer “você” em português, há outros pronomes pessoais do caso reto como “ela”, “elas” e “ele”. São cinco ocorrências do pronome “ela” no artigo acadêmico e três ocorrências desse mesmo pronome no manual técnico. Já para a cartilha e

*folder* de divulgação, há três ocorrências do pronome “elas” e uma ocorrência do pronome “ele”. Contudo, esses pronomes não foram detidamente analisados.

Já os pronomes pessoais do caso reto dos tipos de texto em inglês estão inseridos na TAB. 17.

TABELA 17. Frequência de ocorrência de pronomes pessoais em inglês

N./Classe	Tipos de texto					
	Artigo acadêmico		Manual técnico		Cartilha e <i>folder</i> de divulgação	
	Pronome	Frequência	Pronome	Frequência	Pronome	Frequência
1	we	20	it	17	it	41
2	it	8	they	18	they	24
3	they	6			you	13
4	I	2			he	6
5					we	5
Total		36		35		89

Nos tipos de tipos de textos em inglês, há uma maior frequência de ocorrência de pronomes pessoais em relação àquela observado nos *corpora* em português. Os pronomes “*you*” e “*he*” ocorrem apenas na cartilha e *folder* de divulgação em uma frequência de 13 e 6 ocorrências, respectivamente. A análise do pronome “*you*” indica que o papel do autor é pode ser definido em relação ao seu leitor e em conformidade com o que está sendo comunicado e negociado em termos de construção de sentido. Esse pronome pode ser indicativo de caracterização específica da cartilha e *folder* de divulgação em inglês, considerado um aspecto do impacto da sintonia no modo e campo.

O cartilha e *folder* de divulgação apresenta uma frequência elevada dos pronomes “*it*” (41 ocorrências) e “*they*” (24 ocorrências) quando comparados aos demais tipos de texto, pois no manual técnico o pronome “*it*” ocorre 17 vezes e o pronome “*they*”, 18 vezes. Já no artigo acadêmico, o pronome “*it*” ocorre oito vezes e o pronome “*they*” apresenta uma frequência de seis ocorrências. O pronome pessoal “*I*” apresenta uma frequência de duas ocorrências apenas no artigo acadêmico. Já o pronome pessoal “*we*” ocorre 20 vezes no artigo acadêmico, cinco vezes na cartilha e *folder* de divulgação e não ocorre no manual técnico.

Os pronomes indicam os tipos de relações dos participantes da interação. Essas relações podem ser de poder entre os participantes e são categorizadas pelo gênero social, nível de

*expertise* e classe social. No caso desta pesquisa, observou-se que essas relações variam conforme cada um dos tipos de texto (e entre línguas), mostrando a distância social entre o autor/leitor. Essa distância é observada a partir das interações, isto é, interação especialista-especialista, no caso do artigo acadêmico; especialista-técnico da área da saúde, no que toca ao manual técnico; e interação especialista-leigo no caso das cartilhas e *folders* de divulgação. Tais diferenças trazem indícios que podem ser utilizados para diferenciar os tipos de textos.

A seguir é apresentada a análise referente à variável modo.

### 3.2.3 A variável MODO

O modo é a variável do discurso responsável pela organização do texto, a partir do papel da língua (constitutivo, no caso desta pesquisa) na interação entre os participantes em uma dada atividade social. Partindo da variável modo, vê-se que o canal dos *corpora* desta pesquisa é o gráfico, o meio é o escrito, e o modo de interação é o monólogo. Conforme explicado anteriormente, somente o modo retórico (*e.g.*, persuasão, argumentação, instrução, exposição) varia em cada um dos tipos de tipos de texto que compuseram os *corpora* desta pesquisa. Assim, a forma como o autor expressa e cria as relações lógicas entre as partes do texto, por meio das conjunções, é analisada nesta pesquisa, sendo esta classe de palavras um indicador da organização da do parâmetro modo retórico.

De acordo com Halliday (2004, p. 587), a conjunção constitui um recurso para identificação de uma transição no desenvolvimento do texto, ou seja, permite que os participantes delimitem as relações retóricas utilizadas na construção do significado do texto. Os significados construídos pelas conjunções, segundo o referido autor, são semânticos, e organizam o texto como um fluxo estabelecido pelas relações coesivas entre as sentenças. Eggins (2004) corrobora essa assertiva, ao afirmar que os padrões coesivos das conjunções explicam a forma como o autor expressa e cria as relações lógicas entre as partes do texto.

Na TAB. 18 estão relacionadas todas as conjunções dos tipos de texto em português por frequência de ocorrência.

TABELA 18. Frequência de ocorrência de conjunções em português

Qte./ Classe	Tipos de Texto					
	Artigo acadêmico		Manual técnico		Cartilha e <i>folder</i> de divulgação	
	Conjunção	Frequência	Conjunção	Frequência	Conjunção	Frequência
1	e	177	e	110	e	116
2	como	25	ou	21	ou	39
3	ou	20	como	12	como	13
4	quando	5	quando	5	se	4
5	entretanto	2	pois	4	portanto	3
6	contudo	2	mas	4	quando	1
7	segundo	1	se	2	porém	1
8	portanto	1	contudo	2	embora	1
9	nem	1	quando	1	conforme	1
10	mas	1	portanto	1	–	–
11	enquanto	1	porém	1	–	–
12	embora	1	enquanto	1	–	–
13	–	–	embora	1	–	–
Total	–	237	–	165	–	179

No artigo acadêmico, foram identificados 12 tipos de conjunção, sendo as mais frequentes “e”, “como”, “ou” e “quando”; as conjunções “entretanto” e “contudo” ocorrem apenas duas vezes e as demais uma única vez. Já o manual técnico foi aquele que apresentou a maior variedade de conjunções (13 no total); as mais frequentes foram “e”, “ou”, “como”, “quando”, “pois” e “mas”. As conjunções “se” e “contudo” ocorreram duas vezes e todas as demais, uma vez. A cartilha e *folder* de divulgação foi aquele que mostrou menor variedade de conjunções quando comparado aos outros dois tipos de texto; as conjunções mais frequentes foram “e”, “ou”, “como”, “se” e “portanto”.

As conjunções listadas na TAB. 18 podem ser indicadoras da organização do modo retórico. Assim, enquanto nódulos, as conjunções podem indicar como o modo retórico é realizado, mostrando como as relações lógicas são estabelecidas ao longo de cada um dos tipos de texto.

No que toca aos tipos de textos em inglês, a TAB. 19 mostra a frequência de ocorrência das conjunções nessa língua.

TABELA 19. Frequência de ocorrência de conjunções em inglês

Qte./Classe	Tipos de texto					
	Artigo acadêmico		Manual técnico		Cartilha e <i>folder</i> de divulgação	
	Conjunção	Frequência	Conjunção	Frequência	Conjunção	Frequência
1	and	141	and	124	and	87
2	or	26	or	24	or	26
3	but	3	but	13	but	18
4	both	4	–	–	–	–
5	nevertheless	1	–	–	–	–
6	nor	1	–	–	–	–
7	either	1	–	–	–	–
Total		177		161	–	131

Para os *corpora* em inglês, o tipo de texto que apresentou maior variedade e frequência de ocorrência de conjunções foi o acadêmico, totalizando sete tipos de conjunções. Os demais tipos de textos apresentaram apenas três tipos de conjunções. Para os três tipos de textos em inglês, as conjunções que mais se destacaram foram “*and*”, “*or*” e “*but*”, sendo a frequência de “*but*” menor no artigo acadêmico (três ocorrências) quando comparado aos demais tipos de textos (13 ocorrências para o manual técnico e 18 ocorrências para a cartilha e *folder* de divulgação).

Portanto, nos tipos de textos em inglês, as conjunções são menos frequentes e menos variadas quando comparadas aos tipos de texto em português. Um ponto a ser observado é a frequência elevada da conjunção “*but*” nos três tipos de textos em inglês em comparação com a frequência do seu correspondente formal, “*mas*”, nos tipos de textos em português. Outra questão a ser destacada é que, nos tipos de textos tanto em inglês quanto em português, há pouca ocorrência de conjunções adversativas. Isso pode indicar que se trata de textos mais assertivos ou com menos espaço para expectativas contrárias. O mesmo parece válido para o artigo acadêmico em inglês, haja vista que é nítida a diferença em relação à frequência de ocorrência de todas as conjunções nesse tipo de texto em relação aos demais.

Nota-se que os autores dos textos em língua portuguesa utilizam mais tipos de conjunções para que possam estabelecer as relações lógicas entre os complexos oracionais ao longo dos textos. Essa explicitação das conjunções tem impacto na compreensão do texto, pois facilita para o leitor o entendimento de como as partes se conectam.

Até aqui foram apresentados os dados referentes à caracterização dos tipos de textos por meio dos parâmetros das variáveis do contexto (*i.e.*, campo, sintonia e modo) com base no estudo de Eggins (2004). A seguir, são apresentados os dados referentes à descrição dos perfis metafuncionais dos textos dos tipos EXPLORAR e HABILITAR em inglês e em português.

### 3.3 Descrição dos Perfis Metafuncionais dos Textos

Os dados apresentados anteriormente subsidiaram uma investigação com base na pesquisa de Eggins (2004) para se caracterizar os tipos de texto por meio das variáveis campo, sintonia e modo e também para identificar padrões de uso linguístico dos distintos tipos de texto em português e em inglês que compuseram os *corpora* desta pesquisa (*i.e.*, artigo acadêmico, manual técnico, cartilha e *folder* de divulgação). Essa investigação resultou em indícios que apontam características específicas de cada um dos tipos de texto.

Tais características podem ser observadas a partir das frequências e distribuição metafuncional em cada um dos tipos de textos, com o objetivo de propor uma descrição sistêmico-funcional (FIGUEREDO, 2011) dos sistemas de TRANSITIVIDADE (relacionado à variável campo, metafunção ideacional), MODO (relacionado à variável sintonia, metafunção interpessoal) e TEMA e MENSAGEM (relacionados à variável modo, metafunção textual), conforme ilustrado na FIG. 56. Essa descrição implica facilitar o processo de construção do conhecimento, por meio de artigos científicos, facilitar o comportamento dos participantes envolvidos na situação por meio de manuais técnicos e transmitir conhecimento de especialistas para leigos por intermédio de cartilhas ou *folders* de divulgação.

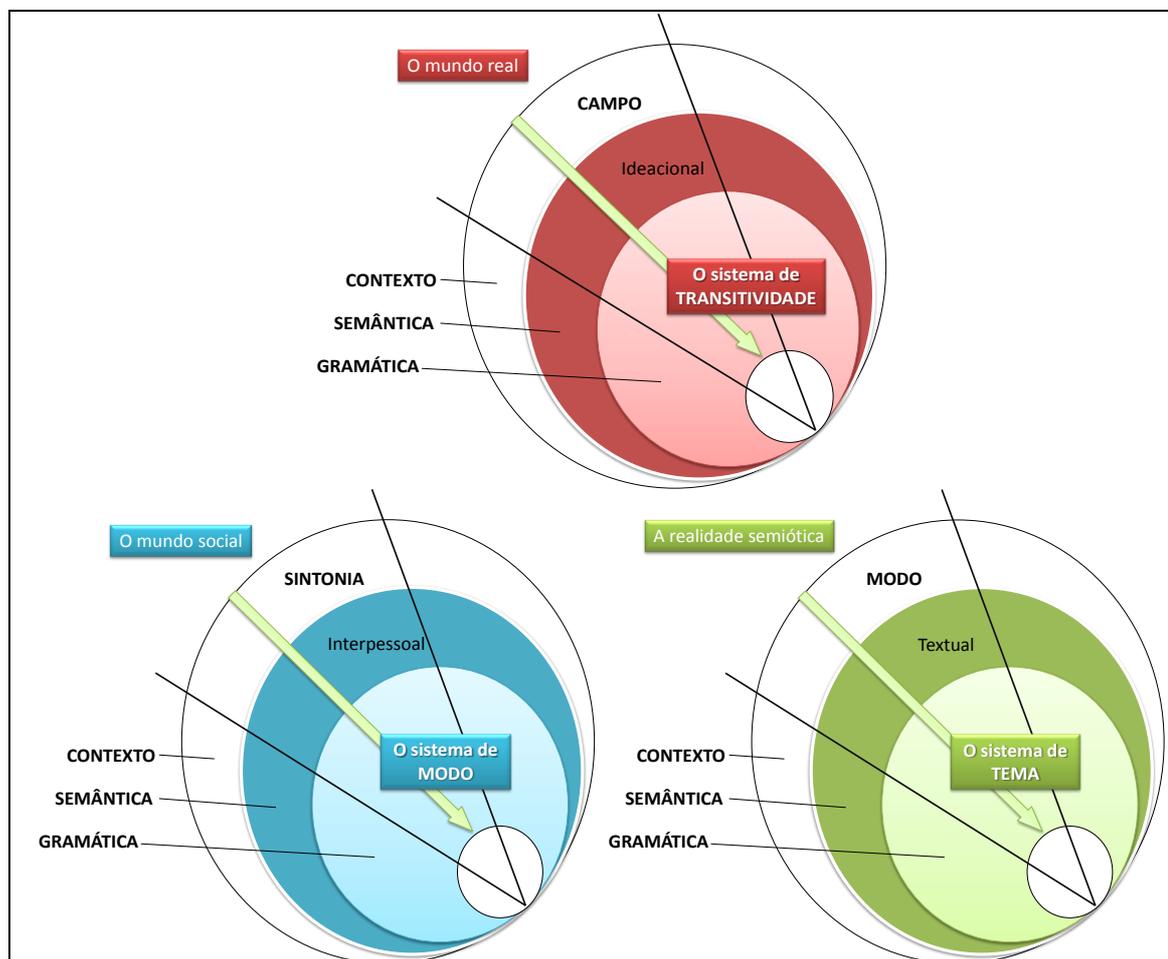


FIGURA 56. Estratificação da língua e os sistemas de TRANSITIVIDADE, MODO e TEMA relacionados às variáveis do contexto (*i.e.*, campo, sintonia e modo) e às metafunções (ideacional, interpessoal e textual)

Fonte: adaptada de Figueredo (2011, p. 75).

A FIG. 56 ilustra a estratificação da língua indicando a localização dos sistemas de TRANSITIVIDADE, MODO e TEMA. Esses sistemas estão relacionados com as variáveis do contexto (*i.e.*, campo, sintonia e modo) analisadas anteriormente (*cf.* Seção 3.2) e as metafunções (ideacional, interpessoal e textual).

Há uma forma de organizar os sistemas de TRANSITIVIDADE, MODO, e TEMA e MENSAGEM, considerando-se o nível de delicadeza (*cf.* TEICH, 1999). Sabe-se que cada sistema pertence a um estrato distinto, porém todos se relacionam por meio da realização. Neste estudo, para tecer comparações entre a língua inglesa e a língua portuguesa, a análise dos textos foi realizada a partir de um estrato mais abstrato. Apesar de os sistemas linguísticos apresentarem uma organização diferente, conservam funções passíveis de comparação em um menor nível de delicadeza (*cf.* TEICH, 1999; HALLIDAY, 2001; MATTHIESSEN, 2001; CAFFAREL; MARTIN; MATTHIESSEN, 2004; FIGUEREDO, 2007). Contudo,

dependendo do nível de delicadeza, faz-se necessário apontar as diferenças entre esses sistemas de acordo com a língua.

Nesta parte da análise, os textos classificados conforme os processos sociossemióticos EXPLORAR (artigo acadêmico) e HABILITAR (manual técnico e cartilha e *folder* de divulgação) são analisados conforme as metafunções ideacional (sistema de TRANSITIVIDADE), interpessoal (sistema de MODO), e textual (sistemas de TEMA e MENSAGEM). Esses sistemas foram analisados no intuito de obter a descrição do perfil metafuncional de cada um dos tipos de texto. A descrição desse perfil fomenta a futura construção de um *pipeline* com fases de processamento textual que visarão orientar o usuário no processo de tradução e produção monolíngue, bilíngue e/ou multilíngue de textos. Para tanto, foi criado o *template* (modelo) com base no perfil metafuncional de um texto de popularização da ciência para a elaboração da cartilha, conforme mencionado anteriormente na Seção 2.2.2.2, como um primeiro passo para se delinear um *pipeline*. A partir desse *template* foi possível elaborar um texto do tipo HABILITAR em que foram consideradas as escolhas metafuncionais com base na gramática sistêmico-funcional hallidayana (cf. Subseção 3.3.4). Vale lembrar que esse *template* foi criado a partir da necessidade de se diferenciar o manual técnico e a cartilha por guardarem semelhanças durante as análises desses dois tipos de texto.

Os dados extraídos a partir do UAM CorpusTool (cf. 2.1.2.2) foram organizados e exemplificados a partir de cada um dos tipos de texto na seguinte ordem:

1. Artigo acadêmico: a) análise ideacional; b) interpessoal; e c) textual dos textos em português e inglês.
2. Manual técnico: a) análise ideacional; b) interpessoal; e c) textual dos textos em português e inglês.
3. Cartilha e *folder* de divulgação: a) análise ideacional; b) interpessoal; e c) textual dos textos em português e inglês.

Na parte da análise ideacional, são explicitados como é representado o mundo de experiências em cada um dos tipos de textos por meio de seus principais processos. No que toca à análise interpessoal, é explicitada a relação entre autor/leitor dos textos. Por fim, a parte da análise textual explicita a distribuição dos significados ideacionais e interpessoais ao longo do texto.

### 3.3.1 *Artigo acadêmico: análise ideacional, interpessoal e textual dos textos em português e inglês*

As Subseções 3.3.1.1, 3.3.1.2 e 3.3.1.3 apresentam as análises ideacional, interpessoal e textual dos artigos acadêmicos em português e em inglês, respectivamente.

#### 3.3.1.1 Análise ideacional dos artigos acadêmicos em português e em inglês

A TAB. 20 mostra a frequência relativa das funções do sistema de TRANSITIVIDADE nos artigos acadêmicos em português e em inglês.

TABELA 20. Frequência relativa das funções do sistema de TRANSITIVIDADE nos artigos acadêmicos em português e inglês

Tipo de processo	Português		Inglês	
	N	%	N	%
<b>Material</b>	19	30,16	25	51,02
<b>Relacional</b>	Identificativo	-	3	13,04
	Atributivo	38	20	86,96
	Subtotal	38	23	49,94
<b>Mental</b>	5	7,93	-	-
<b>Existencial</b>	-	-	1	2,04
<b>Comportamental</b>	-	-	-	-
<b>Verbal</b>	1	1,59	-	-
<b>TOTAL</b>	63	100,00	49	100,00

A seguir estão exemplificados cada um dos tipos de processos (destacados em negrito) e suas principais características.

#### 1) Processos Materiais

Como é possível observar na TAB. 20, os processos materiais no artigo em português correspondem a 30,16% (19 ocorrências) do total de 63 ocorrências. Esses processos são os responsáveis por esclarecer as informações específicas ou intercorrências da anemia falciforme, como ilustrado nos exemplos (1) e (2):

- (1) No estado de Minas Gerais, o Programa Estadual de Triagem Neonatal de Minas Gerais PETN-MG **detectou** a proporção de uma criança com anemia a cada nascidos vivos.
- (2) Além disso, as hemácias falcizadas **se tornam** mais aderentes.

Já no artigo em inglês, os processos materiais correspondem a 51,02% (25 ocorrências) do total de 49 ocorrências. Os exemplos (3) e (4) ilustram como os processos materiais explicitam o que foi feito no curso da pesquisa que deu origem ao artigo. A maioria desses tipos de processos parece fazer parte da seção de metodologia.

- (3) We **identified** topics important to the care of children with SCD by using guidelines, review articles and clinical expertise.
- (4) In addition to abstracts identified electronically, we **performed** an ad hoc search of bibliographies of articles selected for this review.

## 2) Processos Relacionais

Os processos relacionais em português correspondem a 60,32% (38 ocorrências) do total de 63 ocorrências. Todas as 38 ocorrências constituem processos relacionais atributivos, os quais introduzem características da anemia falciforme, conforme ilustrado nos exemplos (5) e (6):

- (5) Por ser uma anomalia da hemoglobina beta, os sintomas comumente só **surtem** após o sexto mês de vida.
- (6) E **melhoram** a qualidade de vida.

No artigo em inglês, os processos relacionais correspondem a 49,94% (23 ocorrências). Desse total, 13,04% (três ocorrências) corresponde a processos relacionais identificativos e 86,96% (20 ocorrências) de relacionais atributivos.

Os processos relacionais identificativos do artigo em inglês apresentam informações sobre a anemia falciforme que já estão consolidadas na literatura médica, conforme mostram os exemplos (7) e (8):

- (7) Sickle cell disease (SCD) **is** a multisystem disease with hemolytic and infectious complications that contribute to high morbidity and mortality rates in children.
- (8) Sickle cell disease (SCD) **is** one of the most common genetic disorders in the United States.

Os processos relacionais atributivos do artigo em inglês justificam a necessidade de se realizar novas pesquisas sobre a anemia falciforme, conforme ilustrado nos exemplos (9) e (10):

- (9) However, a comprehensive review of the literature for the prevention and management of pediatric SCD-related complications **is** lacking.

- (10) However, research studies **are** lacking in key areas for children with SCD, including studies of disease-modifying treatments.

### 3) Processos Mentais e Verbais

Os processos mentais e verbais ocorrem apenas no artigo acadêmico em português. Os processos mentais equivalem a 7,93% (cinco ocorrências) do total de 63 ocorrências, conforme ilustram os exemplos (11) e (12). Esses processos ocorrem quando há referência aos cuidados necessários com os pacientes no que concerne à atenção psicológica.

- (11) Há necessidade de se **acreditar** nos pacientes quando dizem estar sentindo dor.

- (12) (...) e **acreditar** em suas potencialidades.

Já os processos verbais equivalem a 1,59% (cinco ocorrências) do total de 63 ocorrências. Esses processos apontam o que dizem especialistas sobre o tratamento de uma das intercorrências da anemia falciforme, conforme ilustrado no exemplo (13).

- (13) Como esclarecimento a receios quanto ao uso de analgesia potente nos episódios graves, pesquisadores **declararam** administrar narcóticos parentais.

### 4) Processos Existenciais

O único processo que ocorre somente no artigo em inglês é o existencial. Esse processo corresponde a 2,04% (uma ocorrência) do total de 49 ocorrências. A única ocorrência encontrada apresenta a amostragem da pesquisa, como ilustra o exemplo (14):

- (14) **There were** 321 (10% of abstracts screened) articles selected for full review (Appendix 2), of which 26 were rated as level I evidence.

### 5) Consolidação

A análise dos processos dos textos do tipo EXPLORAR explicitou a forma como o mundo das experiências (*cf.* FIG. 56) é construído no texto. Essa construção tem como objetivo criar significados, por meio de pesquisas empíricas sobre a anemia falciforme, para a construção de negociação de posições e valores com os outros membros da comunidade acadêmica e por leitores da área da saúde em geral.

A seguir, apresenta-se a análise interpessoal dos artigos acadêmicos em português e em inglês.

### 3.3.1.2 Análise Interpessoal dos artigos acadêmicos em português e em inglês

A análise interpessoal dos artigos acadêmicos em português e inglês levou aos dados tabulados a seguir.

A TAB. 21 mostra a frequência relativa das funções do sistema de MODO nos artigos acadêmicos em português e em inglês:

TABELA 21. Frequência relativa das funções do sistema de MODO e Funções Discursivas nos artigos acadêmicos em português e inglês

O sistema de MODO e Funções discursivas			Português		Inglês	
			N	%	N	%
Tipo de Modo	Indicativo	Declarativo	63	100	49	100
		Interrogativo	-	-	-	-
		Subtotal	63	100	49	100
	Imperativo		-	-	-	
Funções discursivas	Turno	Inicial	63	100	49	100
		Correspondente	-	-	-	-
		Subtotal	63	100	49	100
	Papel	Oferta	63	100	49	100
		Demanda	-	-	-	-
		Subtotal	63	100	49	100
	Tipo de Mercadoria	Informação	63	100	49	100
		Bens e serviços	-	-	-	-
		Subtotal	63	100	49	100

A TAB. 21 mostra que, no artigo tanto em português e inglês, o Modo Indicativo Declarativo é unânime (63 ocorrências e 49 ocorrências, respectivamente). Esse Modo tem como função semântica a oferta de informação, conforme ilustrado nos exemplos<sup>33</sup> (15), (16), (17) e (18):

- (15) Avaliação e medição criteriosas da dor falciforme – e de qualquer dor – são passos fundamentais para a programação das ações necessárias para seu alívio ou abolição.

<sup>33</sup> Cumpre observar que todos os exemplos do sistema de MODO foram sublinhados, indicando que a oração inteira corresponde a um tipo de Modo específico.

- (16) A autora e outros estudiosos consideraram adequado o uso de autorrelatos durante as crises e nos seus intervalos.
- (17) This article presents a comprehensive review of the literature for the care provided to children with SCD.
- (18) Sickle cell disease (SCD) is a multisystem disease with hemolytic and infectious complications that contribute to high morbidity and mortality rates in children.

Por se tratar de textos do tipo EXPLORAR, em que há negociação de significados com membros de uma mesma comunidade científica (no caso, especialistas da área de saúde), é esperada a grande ocorrência do Modo Indicativo Declarativo com função semântica de oferta de informação. Trata-se, em geral, de textos que apresentam informações novas que são postas em negociação.

### 3.3.1.3 Análise textual dos artigos acadêmicos em português e em inglês

Apresenta-se primeiramente a análise textual dos artigos acadêmicos em português e em seguida dos artigos acadêmicos em inglês.

#### 1) Artigo acadêmico em português

A análise textual do artigo acadêmico em português possibilitou o levantamento dos dados tabulados nas TAB. 22. Essa tabela mostra que 11,11% das mensagens do texto são iniciais (sete ocorrências), enquanto 88,89% são fases (56 ocorrências) do total de 63 ocorrências. Dessas fases, 82,14% são de continuidade (46 ocorrências) e 17,86% são de descontinuidade (10 ocorrências). Dos tipos de descontinuidade, 30% são desvios (três ocorrências) e 70% são mudanças (sete ocorrências). Todas as fases são não focalizadas (100%). Quanto ao arranjo, 77,42% das mensagens são arranjo *default* (56 ocorrências) e 11,11% são arranjos (sete ocorrências). Dos arranjos, 57,14% são explícitos (quatro ocorrências) e 42,86% são implícitos (três ocorrências).

TABELA 22. Frequência relativa das funções do sistema de MENSAGEM nos artigos acadêmicos em português

Tipos de mensagem			N	%	
Contextualização	Inicial		7	11,11	
	Fase	Continuidade	46	82,14	
		Descontinuidade	Desvio	3	30,00
			Mudança	7	70,00
			Subtotal	10	17,86
	Subtotal		56	100,00	
Subtotal			63	100,00	
Arranjo	<i>Default</i>		56	88,89	
	Arranjo	Explícito (periódico)	4	57,14	
		Implícito (em série)	3	42,86	
		Subtotal	7	100,00	
	Subtotal			63	100,00
Focalização	Não focalizada		63	100,00	
	Focalizada	Particularizada	-	-	
		Contraste	-	-	
		Subtotal	-	-	
	Subtotal			63	100,00

O fato de o artigo acadêmico em português estar organizado em sete mensagens iniciais indica que existem sete “blocos” de informação textual<sup>34</sup> em português e cada um desses “blocos” distribui, de maneira diferente, os significados ideacionais e interpessoais do texto (cf. APÊNDICES). Entende-se por bloco que cada uma das mensagens iniciais introduz um novo assunto e, portanto, configura uma organização nova de fases no texto.

O QUADRO 12 apresenta um escâner do artigo acadêmico em português classificado por TEMA e MENSAGEM para melhor visualização da distribuição desses dois sistemas no texto. Cada uma das mensagens iniciais e das descontinuidades:mudanças e descontinuidades:desvios está exemplificada em itálico. Com o objetivo de se explicar o que quer dizer a mensagem, um exemplo de cada um dos tipos de mensagem foi selecionado, com exceção da continuidade, por seguir o mesmo padrão da inicial.

<sup>34</sup> Entende-se por “blocos” o conjunto sequencial de “ondas” (HALLIDAY, 2002) com suas cristas de proeminência e não proeminência.

## QUADRO 12. Distribuição dos sistemas de TEMA e MENSAGEM no artigo acadêmico em português

<p><b>MENSAGEM 1. inicial. arranjoexplícito (periódico). não focalizado. TEMA: Tema default. default.</b>  <i>Avaliação e medição criteriosas da dor falciforme – e de qualquer dor - são passos fundamentais para a programação das ações necessárias para seu alívio ou abolição.</i></p> <p><b>MENSAGEM 2. fase. continuidade. arranjo-default. não focalizado. TEMA: Ângulo (fonte). default.</b></p> <p><b>MENSAGEM 3. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</b></p> <p><b>MENSAGEM 4. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</b></p> <p><b>MENSAGEM 5. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</b></p> <p><b>MENSAGEM 6. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</b></p> <p><b>MENSAGEM 7. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</b></p> <p><b>MENSAGEM 8. fase. descontinuidade:mudança. arranjo-default. não focalizado. TEMA: Tema default. proeminente.</b>  <i>A forma da dor ser tratada tem implicações futuras na maneira como os pacientes conviverão com ela e com a vida.</i></p> <p><b>MENSAGEM 9. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</b></p> <p><b>MENSAGEM 10. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</b></p> <p><b>MENSAGEM 11. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</b></p> <p><b>MENSAGEM 12. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</b></p> <p><b>MENSAGEM 13. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</b></p> <p><b>MENSAGEM 14. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</b></p> <p><b>MENSAGEM 15. fase. descontinuidade:mudança. arranjo-default. não focalizado. TEMA: Ângulo (fonte). proeminente.</b>  <i>Segundo o autor como quase todos os episódios são manejados em casa, a avaliação da dor e das medidas de convivência com ela precisa focalizar seu manejo satisfatório em ambiente domiciliar.</i></p> <p><b>MENSAGEM 16. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</b></p> <p><b>MENSAGEM 17. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</b></p> <p><b>MENSAGEM 18. inicial. arranjo implícito (em série). não focalizado. TEMA: Ângulo (fonte). proeminente.</b>  <i>Para a autora desde a mais tenra idade a criança e seus familiares devem aprender a lidar com a dor e tratá-la.</i></p> <p><b>MENSAGEM 19. fase. descontinuidade:desvio. arranjo-default. não focalizado. TEMA: Ângulo (fonte). proeminente.</b>  <i>Para o autor, para que se conheça e reduza o impacto real da dor sobre a função psicológica e a qualidade de vida, os episódios tratados parcial ou totalmente nos domicílios - assim como as estratégias adotadas para controle – devem ser considerados e incluídos nos estudos, cujos resultados serão inadequados se não se levar em conta essa realidade.</i></p> <p><b>MENSAGEM 20. fase. descontinuidade:mudança. arranjo-default. não focalizado. TEMA: Ângulo (fonte). proeminente.</b>  <i>No processo educativo é preciso conferir poder ao paciente e aos familiares, no sentido de deterem o conhecimento sobre a doença.</i></p> <p><b>MENSAGEM 21. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</b></p> <p><b>MENSAGEM 22. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</b></p> <p><b>MENSAGEM 23. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</b></p> <p><b>MENSAGEM 24. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</b></p> <p><b>MENSAGEM 25. inicial. arranjo explícito (periódico). não focalizado. TEMA: Ângulo (fonte). proeminente.</b>  <i>Segundo o autor, a criança com dor falciforme deve ser estimulada e treinada.</i></p> <p><b>MENSAGEM 26. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</b></p> <p><b>MENSAGEM 27. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</b></p> <p><b>MENSAGEM 28. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</b></p> <p><b>MENSAGEM 29. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</b></p> <p><b>MENSAGEM 30. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</b></p> <p><b>MENSAGEM 31. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</b></p> <p><b>MENSAGEM 32. fase. continuidade. arranjo-default. não focalizado. TEMA: Ângulo (fonte). default.</b></p> <p><b>MENSAGEM 33. fase. descontinuidade:mudança. arranjo-default. não focalizado. TEMA: Textual. proeminente.</b>  <i>Apesar de variada gravidade clínica, um contingente significativo de pacientes sofre as consequências de uma doença crônica e grave.</i></p> <p><b>MENSAGEM 34. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</b></p> <p><b>MENSAGEM 35. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</b></p> <p><b>MENSAGEM 36. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</b></p> <p><b>MENSAGEM 37. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</b></p> <p><b>MENSAGEM 38. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</b></p> <p><b>MENSAGEM 39. fase. descontinuidade:mudança. arranjo-default. não focalizado. TEMA: Tema default. proeminente.</b>  <i>No estado de Minas Gerais, o Programa Estadual de Triagem Neonatal de Minas Gerais PETN-MG detectou a proporção de uma criança com anemia falciforme a cada nascidos vivos.</i></p> <p><b>MENSAGEM 40. inicial. arranjo explícito (periódico). não focalizado. TEMA: Tema default. default.</b>  <i>Em determinadas situações, como hipóxia, desidratação e acídose, as moléculas de Hb S formam polímeros alongados que modificam o citoesqueleto das hemácias, dando origem à forma característica de foice (falcização).</i></p> <p><b>MENSAGEM 41. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</b></p> <p><b>MENSAGEM 42. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</b></p> <p><b>MENSAGEM 43. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</b></p> <p><b>MENSAGEM 44. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</b></p> <p><b>MENSAGEM 45. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</b></p> <p><b>MENSAGEM 46. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</b></p> <p><b>MENSAGEM 47. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</b></p> <p><b>MENSAGEM 48. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</b></p> <p><b>MENSAGEM 49. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</b></p> <p><b>MENSAGEM 50. fase. descontinuidade:mudança. arranjo-default. não focalizado. TEMA: Textual. proeminente.</b>  <i>Apesar da alteração principal estar restrita às hemácias, trata-se de doença sistêmica cujos efeitos podem incidir sobre qualquer órgão.</i></p> <p><b>MENSAGEM 51. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</b></p> <p><b>MENSAGEM 52. inicial. arranjo explícito (periódico). não focalizado. TEMA: Tema default. default.</b>  <i>Por ser uma anomalia da globina beta, os sintomas comumente só surgem após o sexto mês de vida, quando a produção da cadeia já é claramente predominante em relação à da cadeia gama da HbF.</i></p> <p><b>MENSAGEM 53. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</b></p> <p><b>MENSAGEM 54. inicial. arranjo explícito (periódico). não focalizado. TEMA: Tema default. default.</b>  <i>O diagnóstico laboratorial da anemia falciforme é baseado na mobilidade eletroforética mais lenta da Hb S em relação à Hb A normal.</i></p> <p><b>MENSAGEM 55. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</b></p> <p><b>MENSAGEM 56. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</b></p> <p><b>MENSAGEM 57. inicial. arranjo explícito (periódico). não focalizado. TEMA: Tema default. default.</b>  <i>Fatores ambientais, como o local onde vive o paciente, prevalência de doenças infectocontagiosas, condições socioeconômicas e acesso e qualidade da assistência médica, podem influenciar no fenótipo da doença.</i></p> <p><b>MENSAGEM 58. fase. descontinuidade:desvio. arranjo-default. não focalizado. TEMA: Tema default. proeminente.</b>  <i>No Brasil, a maioria dos indivíduos acometidos possui baixa renda e necessita dos cuidados médicos fornecidos pelo Sistema Único de Saúde.</i></p> <p><b>MENSAGEM 59. fase. descontinuidade:mudança. arranjo-default. não focalizado. TEMA: Textual. proeminente.</b>  <i>Apesar de ser causada pela substituição de um único nucleotídeo, o fenótipo da anemia falciforme é o produto de vários genes.</i></p> <p><b>MENSAGEM 60. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</b></p> <p><b>MENSAGEM 61. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</b></p> <p><b>MENSAGEM 62. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</b></p> <p><b>MENSAGEM 63. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</b></p>
---

Conforme observado no QUADRO 12, a Mensagem inicial 1 organiza o texto de forma que as informações sobre o manejo da dor em pacientes com anemia falciforme sejam distribuídas juntas. Após as mensagens iniciais existem as mensagens de continuidade, descontinuidade:mudança e descontinuidade:desvio. As de continuidade seguem o mesmo padrão de informações das mensagens iniciais. Portanto, o conjunto de orações que vai desde a Mensagem inicial 1 até a próxima mensagem inicial (*i.e.*, Mensagem 18) refere-se a um “bloco” de informação textual.

Quando há necessidade, dentro de uma mesma fase, de focar algum aspecto específico do texto que necessita ser explicado com mais clareza, ou até mesmo exemplificado, são utilizadas as mensagens de descontinuidade:mudança. No caso do artigo acadêmico, especificamente, as mensagens de descontinuidade:mudança se iniciam majoritariamente com temas textuais (conjunções adversativas). Esse dado corrobora aquele achado na Seção 3.2.3, ou seja, em que observou-se no texto em português maior utilização de tipos de conjunção para que as relações lógicas entre os complexos oracionais ao longo dos textos pudesse ser estabelecida. Portanto, a explicitação das conjunções como as adversativas, por exemplo, podem indicar que uma das ideias do texto é contraditória com algum pressuposto que precisa ser explicitado. Esse fato tem impacto na compreensão do texto, pois facilita para o leitor o entendimento de como as partes se conectam.

A Mensagem 33, por exemplo, focaliza o texto considerando-se o quadro da severidade da anemia falciforme; apesar de esse quadro apresentar variações, todas as intercorrências são consideradas graves, na maior parte dos casos.

Já as mensagens de descontinuidade:desvio são utilizadas quando há a necessidade de se focar algum participante específico. A Mensagem 19, por exemplo, focaliza o texto em um dos autores que foram citados na pesquisa.

No que toca ao sistema de TEMA dos artigos acadêmicos em português, todos os tipos de temas ilustrados estão identificados na TAB. 23.

TABELA 23. Frequência relativa das funções do sistema de TEMA nos artigos acadêmicos em português

Tipos de temas		N	%
Complexidade	elemental	-	-
	ângulo	7	11,11
	textual	24	38,10
	interpessoal	-	-
	tema- <i>default</i>	32	50,79
	Subtotal	63	100,00
Ângulo	fonte	7	100,00
	ponto de vista	-	-
	Subtotal	7	100,00
Seleção Temática	<i>default</i>	53	84,13
	proeminente	10	15,87
	Subtotal	63	100,00

A TAB. 23 mostra a frequência relativa das funções do sistema de TEMA no artigo acadêmico em português, distribuído quanto à complexidade e seleção temática. Em relação à complexidade, observa-se que 11,11% são temas ângulo-fonte (sete ocorrências), 38,10% são temas textuais (24 ocorrências) e 50,79% são temas-*default* (32 ocorrências). Quanto à seleção temática, 84,13% são *default* (53 ocorrências) e 15,87% são proeminentes (10 ocorrências).

## 2) Artigo acadêmico em inglês

No que toca à análise textual do artigo acadêmico em inglês, os dados estão tabulados conforme se observa na TAB. 24.

TABELA 24. Frequência relativa das funções do sistema de MENSAGEM no artigo acadêmico em inglês

Tipos de mensagem			N	%	
Contextualização	Inicial		9	18,37	
	Fase	Continuidade	33	82,50	
		Descontinuidade	Desvio	1	14,29
			Mudança	6	85,71
			Subtotal	7	100,00
	Subtotal		40	100,00	
Subtotal		49	100,00		
Arranjo	<i>Default</i>		40	81,63	
	Arranjo	Explícito (periódico)	9	100,00	
		Implícito (em série)	-	-	
		Subtotal	9	18,37	
Subtotal		49	100,00		
Focalização	Não focalizada		49	100,00	
	Focalizada	Particularizada	-	-	
		Contraste	-	-	
		Subtotal	-	-	
Subtotal		49	100,00		

Os dados ilustrados na TAB. 24 mostram que 18,37% das mensagens do texto são iniciais (nove ocorrências), enquanto 81,63% são fases (40 ocorrências). As nove mensagens iniciais indicam que o texto é constituído de nove “blocos” de informação textual, cada qual com um tipo de informação diferente sobre a anemia falciforme.

Dentre as fases, 82,50% são de continuidade (33 ocorrências) e 17,50% são de descontinuidade (sete ocorrências). Dos tipos de descontinuidade, 14,29% são desvio (uma ocorrência) e 85,71% são mudança (seis ocorrências). Todas as fases são não focalizadas (100%). Quanto ao arranjo, 81,63% das mensagens são arranjo-*default* (40 ocorrências) e 18,37% são arranjos (9 ocorrências). Dos arranjos, 100% da 49 ocorrências são do tipo explícito (periódico).

O QUADRO 13 apresenta o artigo acadêmico em inglês classificado por TEMA e MENSAGEM para melhor visualização da distribuição desses dois sistemas no texto. Cada

uma das mensagens iniciais, das descontinuidades:mudanças e das descontinuidades:desvios está exemplificada em *itálico*. No intuito de explicar o que quer dizer a mensagem, foi selecionado um exemplo de cada um dos tipos de mensagem.

QUADRO 13. Distribuição dos sistemas de TEMA e MENSAGEM no artigo acadêmico em inglês

<p>MENSAGEM 1. inicial. arranjo:explícito (periódico). não focalizado. TEMA: simples. não marcado.  <i>Sickle cell disease (SCD) is a multisystem disease with hemolytic and infectious complications that contribute to high morbidity and mortality rates in children.</i>  <b>MENSAGEM 2. fase. descontinuidade:mudança. arranjo-default. não focalizado. TEMA: múltiplo (textual e tópico). não marcado.</b>  <i>However, a comprehensive review of the literature for the prevention and management of pediatric SCD-related complications is lacking.</i>  <b>MENSAGEM 3. inicial. arranjo:explícito (periódico). não focalizado. TEMA: simples. não marcado.</b>  <i>This article presents a comprehensive review of the literature for the care provided to children with SCD.</i>  <b>MENSAGEM 4. fase. continuidade. arranjo-default. não focalizado. TEMA: simples. não marcado.</b>  <b>MENSAGEM 5. inicial. arranjo:explícito (periódico). não focalizado. TEMA: simples. não marcado.</b>  <i>Sickle cell disease (SCD) is one of the most common genetic disorders in the United States.</i>  <b>MENSAGEM 6. fase. continuidade. arranjo-default. não focalizado. TEMA: simples. não marcado.</b>  <b>MENSAGEM 7. fase. continuidade. arranjo-default. não focalizado. TEMA: simples. não marcado.</b>  <b>MENSAGEM 8. fase. descontinuidade:desvio. arranjo-default. não focalizado. TEMA: simples. não marcado.</b>  <i>Universal newborn screening and advances in routine care, including penicillin prophylaxis and immunization against serious bacterial infections, have significantly decreased mortality in young children with SCD.</i>  <b>MENSAGEM 9. fase. continuidade. arranjo-default. não focalizado. TEMA: múltiplo (textual e tópico). não marcado.</b>  <b>MENSAGEM 10. fase. descontinuidade:mudança. TEMA: múltiplo (textual e tópico). não marcado.</b>  <i>However, research studies are lacking in key areas for children with SCD, including studies of disease-modifying treatments.</i>  <b>MENSAGEM 11. fase. continuidade. arranjo-default. não focalizado. TEMA: múltiplo (textual e tópico). não marcado.</b>  <b>MENSAGEM 12. fase. continuidade. arranjo-default. não focalizado. TEMA: múltiplo (textual e tópico). não marcado.</b>  <b>MENSAGEM 13. fase. continuidade. arranjo-default. não focalizado. TEMA: múltiplo (textual e tópico). não marcado.</b>  <b>MENSAGEM 14. fase. descontinuidade:mudança. arranjo default. não focalizado. TEMA: múltiplo (textual e tópico). não marcado.</b>  <i>However, only 1 small randomized controlled trial (RCT) has been published on the use of hydroxyurea in children to date.</i>  <b>MENSAGEM 15. fase. continuidade. arranjo-default. não focalizado. TEMA: simples. default.</b>  <b>MENSAGEM 16. fase. continuidade. arranjo-default. não focalizado. TEMA: simples. não marcado.</b>  <b>MENSAGEM 17. inicial. arranjo:explícito (periódico). não focalizado. TEMA: simples. não marcado.</b>  <i>We identified topics important to the care of children with SCD by using guidelines, review articles and clinical expertise.</i>  <b>MENSAGEM 18. fase. continuidade. arranjo-default. não focalizado. TEMA: múltiplo (textual e tópico). não marcado.</b>  <b>MENSAGEM 19. fase. continuidade. arranjo-default. não focalizado. TEMA: simples. não marcado.</b>  <b>MENSAGEM 20. fase. continuidade. arranjo-default. não focalizado. TEMA: simples. não marcado.</b>  <b>MENSAGEM 21. fase. continuidade. arranjo-default. não focalizado. TEMA: simples. não marcado.</b>  <b>MENSAGEM 22. fase. continuidade. arranjo-default. não focalizado. TEMA: simples. não marcado.</b>  <b>MENSAGEM 23. fase. continuidade. arranjo-default. não focalizado. TEMA: múltiplo (textual e tópico). não marcado.</b>  <b>MENSAGEM 24. inicial. arranjo:explícito (periódico). não focalizado. TEMA: simples. não marcado.</b>  <i>We searched for these topics in Ovid Medline for Evidence Level of Articles Selected for Full Review For each topic, we used several key words to identify relevant articles.</i>  <b>MENSAGEM 25. fase. continuidade. arranjo-default. não focalizado. TEMA: simples. não marcado.</b>  <b>MENSAGEM 26. fase. continuidade. arranjo-default. não focalizado. TEMA: simples. não marcado.</b>  <b>MENSAGEM 27. fase. continuidade. arranjo-default. não focalizado. TEMA: simples. não marcado.</b>  <b>MENSAGEM 28. fase. continuidade. arranjo-default. não focalizado. TEMA: múltiplo (textual e tópico). marcado.</b>  <b>MENSAGEM 29. inicial. arranjo:explícito (periódico). não focalizado. TEMA: simples. não marcado.</b>  <i>Each abstract was independently reviewed by 2 authors (Drs Kavanagh, Wang, and Sprinz).</i>  <b>MENSAGEM 30. fase. continuidade. arranjo-default. não focalizado. TEMA: simples. não marcado.</b>  <b>MENSAGEM 31. fase. descontinuidade:mudança. não focalizado. TEMA: múltiplo (textual e tópico). marcado.</b>  <i>If the research methods were unclear in the abstract, the article was reviewed in detail.</i>  <b>MENSAGEM 32. inicial. arranjo:explícito (periódico). não focalizado. TEMA: simples. não marcado.</b>  <i>Studies were excluded if they enrolled adults only, were in vitro or animal studies, gene or genome-wide association studies, non-English-language abstracts or articles, commentaries or editorials, or case reports.</i>  <b>MENSAGEM 33. fase. continuidade. arranjo-default. não focalizado. TEMA: simples. não marcado.</b>  <b>MENSAGEM 34. fase. descontinuidade:mudança. não focalizado. TEMA: múltiplo (textual e tópico). não marcado.</b>  <i>Although these studies could provide important information on the historical natural history of SCD, they might not reflect outcomes that could be achieved under current standards of care.</i>  <b>MENSAGEM 35. fase. continuidade. arranjo-default. não focalizado. TEMA: simples. não marcado.</b>  <b>MENSAGEM 36. fase. continuidade. arranjo-default. não focalizado. TEMA: simples. não marcado.</b>  <b>MENSAGEM 37. fase. continuidade. arranjo-default. não focalizado. TEMA: simples. não marcado.</b>  <b>MENSAGEM 38. fase. continuidade. arranjo-default. não focalizado. TEMA: simples. não marcado.</b>  <b>MENSAGEM 39. fase. continuidade. arranjo-default. não focalizado. TEMA: simples. não marcado.</b>  <b>MENSAGEM 40. inicial. arranjo explícito (default). não focalizado. TEMA: simples. não marcado.</b>  <i>The hierarchy of research design was assigned following the US Preventive Services Task Force ratings.</i>  <b>MENSAGEM 41. fase. continuidade. arranjo-default. não focalizado. TEMA: simples. não marcado.</b>  <b>MENSAGEM 42. fase. continuidade. arranjo-default. não focalizado. TEMA: simples. não marcado.</b>  <b>MENSAGEM 43. fase. continuidade. arranjo-default. não focalizado. TEMA: simples. não marcado.</b>  <b>MENSAGEM 44. inicial. arranjo explícito (periódico). não focalizado. TEMA: simples. não marcado.</b>  <i>We identified 3188 abstracts for screening.</i>  <b>MENSAGEM 45. fase. continuidade. arranjo-default. não focalizado. TEMA: múltiplo (textual e tópico). não marcado.</b>  <b>MENSAGEM 46. fase. continuidade. arranjo-default. não focalizado. TEMA: simples. não marcado.</b>  <b>MENSAGEM 47. fase. continuidade. arranjo-default. não focalizado. TEMA: simples. não marcado.</b>  <b>MENSAGEM 48. fase. continuidade. arranjo-default. não focalizado. TEMA: simples. não marcado.</b>  <b>MENSAGEM 49. fase. continuidade. arranjo-default. não focalizado. TEMA: múltiplo (textual e tópico). não marcado.</b></p>
--

A Mensagem inicial 1 e suas continuidades organizam o texto apresentando as informações sobre a anemia falciforme já consolidadas na literatura médica (*cf.* QUADRO 13). Esse “bloco” de informações inicia-se na Mensagem 1 e termina na Mensagem 3.

No caso do artigo acadêmico, as mensagens de descontinuidade:mudança se iniciam em sua maior parte com temas textuais (conjunções adversativas). Esses temas textuais indicam que uma das ideias do texto é contraditória com algum pressuposto que precisa ser explicitado. Em inglês, a Mensagem 14 indica que, apesar de haver vários estudos sobre a anemia falciforme, em algumas de suas especificidades, o número de estudos ainda é reduzido. No caso do artigo em inglês, não se observa a repetição da conjunção adversativa com temas textuais.

Nos artigos acadêmicos em inglês, a Mensagem 8 enfoca as ações e os avanços da triagem neonatal que auxiliam no tratamento das crianças com anemia falciforme.

Todos os tipos de temas ilustrados acima em relação ao sistema de TEMA dos artigos acadêmicos em inglês estão identificados na TAB. 25. Esses temas não foram analisados detalhadamente por não constituírem foco desta pesquisa.

TABELA 25. Frequência relativa das funções do sistema de TEMA no artigo acadêmico em inglês

Tipos de temas		N	%
Complexidade	Simple	35	71,43
	Múltiplo	14	28,57
	Subtotal	49	100,00
Múltiplo	Textual e tópico	14	100,00
	Interpessoal e tópico	-	-
	Textual interpessoal e tópico	-	-
	Subtotal	14	100,00
Seleção temática	Não marcado	47	95,92
	Marcado	2	4,08
	Subtotal	49	100,00

A TAB. 25 ilustra a frequência relativa das funções do sistema de TEMA no artigo acadêmico em inglês, distribuído quanto à complexidade e seleção temática. Em relação à complexidade, observa-se que 71,43% são temas simples (35 ocorrências) e 28,57% são temas múltiplos (14

ocorrências). Dos múltiplos, 100% são temas do tipo textual e tópico. No que toca à seleção temática, 95,92% dos temas são não marcados (47 ocorrências) e 4,8% são marcados (2 ocorrências).

### 3.3.2 *Manuais técnicos: análise ideacional, interpessoal e textual dos textos em português e inglês*

As Subseções 3.3.2.1, 3.3.2.2 e 3.3.2.3 mostram as análises ideacional, interpessoal e textual dos manuais técnicos em português e em inglês, respectivamente.

#### 3.3.2.1 Análise ideacional dos manuais técnicos em português e em inglês

A TAB. 26 mostra a frequência relativa das funções do sistema de TRANSITIVIDADE nos manuais técnicos em português e em inglês.

TABELA 26. Frequência relativa das funções do sistema de TRANSITIVIDADE nos manuais em português e inglês

Tipo de processo		Português		Inglês	
		N	%	N	%
Material		17	21,79	5	9,62
Relacional	Identificativo	5	8,33	2	4,76
	Atributivo	55	91,67	40	95,24
	Subtotal	60	100	42	100,00
Mental		-	-	-	-
Existencial		1	1,28	5	9,62
Comportamental		-	-	-	-
Verbal		-	-	-	-
TOTAL		63	100	49	100,00

A seguir estão exemplificados cada um dos tipos de processos (destacados em negrito) e suas principais características.

#### 1) Processos Materiais

Como é possível observar na TAB. 26, os processos materiais no manual em português correspondem a 21,79% (17 ocorrências) do total de 78 ocorrências. No manual em

português, os processos materiais apresentam as ações que devem ser tomadas no tratamento dos pacientes, como mostram os exemplos (19) e (20).

(19) Em casos graves, **realiza-se** a exsanguíneotransfusão.

(20) A maior parte do sangue sequestrado **retorna** para a circulação.

Já no manual em inglês, os processos materiais correspondem a 9,62% (cinco ocorrências) do total de 52 ocorrências. Esses processos apresentam as ações dos procedimentos do tratamento, conforme se nota nos exemplos (21) e (22).

(21) TCDs **are done** routinely.

(22) (...) and the approved local policy pain medication **is administered** following which the patient will be assessed by the medical team.

## 2) Processos relacionais

Os processos relacionais no manual em português correspondem a 79,92% (60 ocorrências) do total de 78 ocorrências. Desse total, 8,33% (cinco ocorrências) corresponde a processos relacionais identificativos e 91,67% (55 ocorrências) consiste em processos relacionais atributivos.

No manual em português, os processos relacionais identificativos apresentam características tanto dos sintomas quanto dos tratamentos da anemia falciforme, conforme ilustrado nos exemplos (23) e (24).

(23) O tratamento **pode ser** ambulatorial com reavaliação em 48 horas.

(24) Pacientes com dor torácica, tosse, dispneia, hipoxemia, prostração e alteração radiológica recente do tórax provavelmente **têm** síndrome torácica aguda.

Nos manuais técnicos em inglês, os processos relacionais correspondem a 80,77% (42 ocorrências) do total de 52 ocorrências. Desse total, 4,76% (duas ocorrências) corresponde a processos relacionais identificativos e 95,24% (40 ocorrências), a processos relacionais atributivos.

Assim como nos manuais técnicos em português, naqueles em inglês os processos relacionais identificativos apresentam características tanto dos sintomas quando dos tratamentos da anemia falciforme, como pode ser observado nos exemplos (25) e (26):

- (25) The patient **is** often managed quicker when attending the day care unit compared to attending A&E.
- (26) Leg ulcers **are** probably caused when minor cuts and abrasions fail to heal properly because of poor circulation to the lower limbs.

Quanto aos processos relacionais atributivos, nos manuais técnicos em português eles mostram características de complicações decorrentes da anemia falciforme, conforme mostram os exemplos (27) e (28):

- (27) O acidente vascular cerebral (AVC) **é** uma das mais graves complicações da doença falciforme.
- (28) **É** a segunda causa mais comum de morte em crianças.

Já nos manuais técnicos em inglês, os processos relacionais atributivos atribuem identidade às complicações decorrentes da anemia falciforme, de acordo com o observado nos exemplos (29) e (30):

- (29) Leg ulceration **is** one of the less well understood complications of SCD.
- (30) Priapism **is** a persistent and abnormal erection of the penis accompanied by pain, tenderness and swelling.

### 3) Processos existenciais

Nos manuais técnicos em português, há uma única ocorrência de processo existencial (1,28% da amostra), conforme apontado no exemplo (31):

- (31) **Existem** três formas clínicas.

No manual em inglês, os processos existenciais correspondem a 4,76% (duas ocorrências) do total de 52 ocorrências. Esses processos mostram que há várias complicações associadas à anemia falciforme, para as quais não há tratamento específico, como observado nos exemplos (32) e (33):

- (32) **There are** a number of short and long term complications associated with sickle cell disease.
- (33) **There is** no specific treatment for individuals with sickle cell disease.

### 3) Consolidação

A análise dos processos dos textos do tipo HABILITAR explicitou a forma como o mundo das experiências (*cf.* FIG. 56) é construído no texto com o intuito de facilitar o comportamento dos participantes em relação às várias hemoglobinopatias, mais especificamente a anemia falciforme. O autor, detentor do conhecimento, elabora o manual técnico para que seja utilizado como base para as ações do técnico da área da saúde no que toca aos procedimentos para a realização da triagem neonatal para anemia falciforme e no que toca ao auxílio ao tratamento dos pacientes diagnosticados com essa doença. Tais procedimentos devem ser seguidos, espera-se, conforme a sugestão do autor.

A seguir é apresentada a análise interpessoal dos manuais técnicos em português e em inglês.

#### 3.3.2.2 Análise Interpessoal dos manuais técnicos em português e em inglês

A TAB. 27 mostra a frequência relativa das funções do sistema de MODO nos manuais técnicos em português e em inglês:

TABELA 27. Frequência relativa das funções do sistema de MODO e Funções Discursivas nos manuais técnicos em português e inglês

Tipo de Modo e Funções discursivas			Português		Inglês	
			N	%	N	%
Tipo de Modo	Indicativo	Declarativo	78	100	52	100
		Interrogativo	-	-	-	-
		Subtotal	78	100	52	100
	Imperativo		-	-	-	
Funções discursivas	Turno	Inicial	78	100	52	100
		Correspondente	-	-	-	-
		Subtotal	78	100	52	100
	Papel	Oferta	78	100	52	100
		Demanda	-	-	-	-
		Subtotal	78	100	52	100
	Tipo de Mercadoria	Informação	78	100	52	100
		Bens e serviços	-	-	-	-
		Subtotal	78	100	52	100

Tanto no manual em português como no manual em inglês, o Modo Indicativo Declarativo foi unânime (*i.e.*, 100% do total de 78 e 52 ocorrências, respectivamente). Os exemplos (34), (35), (36) e (37) referem-se aos tipos de Modo nos manuais em português:

- (34) Qualquer tipo de pneumonia exige hospitalização para antibioticoterapia venosa, suportes clínico e ventilatório.
- (35) Caso esteja indicada a antibioticoterapia empírica (febre sem foco), pode-se utilizar ampicilina.
- (36) As part of routine care it is also recommended that children have Transcranial Doppler (TCD) scanning of the brain.
- (37) On arrival in the day-care unit an initial vital signs, general observations and pain assessment is conducted by the nurse.

Por se tratar de textos do tipo HABILITAR, observa-se que há oferta de informações, visando orientar ações de atuação dos profissionais de saúde. Pressupõe-se, no caso, que os especialistas precisam de conhecimento sobre as intercorrências, urgências/emergências e medidas profiláticas da doença falciforme.

### 3.3.2.3 Análise Textual do manual técnico em português e em inglês

A análise textual dos manuais técnicos é apresentada a seguir, de forma isolada para o português e para o inglês.

#### 1) Manual técnico em português

A TAB. 28 exhibe a frequência relativa das funções do sistema de MENSAGEM no manual técnico em português

TABELA 28. Frequência relativa das funções do sistema de MENSAGEM no manual técnico em português

Tipos de mensagem			N	%	
Contextualização	Inicial		7	8,97	
	Fase	Continuidade		59	83,10
		Descontinuidade	Desvio	4	33,33
			Mudança	8	66,67
			Subtotal	12	16,90
	Subtotal		71	100,00	
Subtotal			78	100,00	
Arranjo	<i>Default</i>		71	91,03	
	Arranjo	Explícito (periódico)	6	85,71	
		Implícito (em série)	1	14,29	
		Subtotal	7	100,00	
	Subtotal			78	100,00
Focalização	Não focalizada		78	100,00	
	Focalizada	Particularizada	-	-	
		Contraste	-	-	
		Subtotal	-	-	
	Subtotal			78	100,00

Os dados tabulados nessa tabela mostram que 8,97% das mensagens do texto são iniciais (sete ocorrências), ao passo que 91,03% são fases (71 ocorrências). Dessas fases, 83,10% são de continuidade (59 ocorrências) e 16,90% são de descontinuidade (12 ocorrências). Dos tipos de descontinuidade, 33,33% são desvios (quatro ocorrências) e 66,67% são mudanças (oito ocorrências). Todas as fases são não focalizadas (100%). Quanto ao arranjo, 91,3% das mensagens são arranjo *default* (71 ocorrências) e 8,97% são arranjos (sete ocorrências). Dos arranjos, 85,71% são explícitos-periódicos (seis ocorrências) e 14,29% são implícitos em série (uma ocorrência).

O QUADRO 14 sintetiza um escâner do manual técnico em português classificado por TEMA e MENSAGEM para melhor entendimento da distribuição desses dois sistemas no texto. Cada uma das mensagens iniciais e das descontinuidade:mudança e descontinuidade:desvio está

exemplificada em *itálico*. Foi selecionado um exemplo de cada um dos tipos de mensagem, à exceção da continuidade, por seguir o mesmo padrão da inicial.

QUADRO 14. Distribuição dos sistemas de TEMA e MENSAGEM no manual técnico em português

<p><b>MENSAGEM 1. inicial. arranjo explícito (periódico). não focalizado. TEMA: Tema <i>default. default.</i></b>  <i>Qualquer tipo de pneumonia exige hospitalização para antibioticoterapia venosa, suportes clínico e ventilatório.</i></p> <p><b>MENSAGEM 2. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. default.</i></b></p> <p><b>MENSAGEM 3. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Textual. <i>default.</i></b></p> <p><b>MENSAGEM 4. fase. descontinuidade:mudança. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. proeminente.</i></b>  <i>Caso esteja indicada a antibioticoterapia empírica (febre sem foco), pode-se utilizar ampicilina.</i></p> <p><b>MENSAGEM 5. fase. descontinuidade:desvio. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. default.</i></b>  <i>É a segunda causa mais comum de morte em crianças menores de anos de idade.</i></p> <p><b>MENSAGEM 6. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. default.</i></b></p> <p><b>MENSAGEM 7. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Textual. <i>default.</i></b></p> <p><b>MENSAGEM 8. Inicial. arranjo explícito (periódico). não focalizado. TEMA: Tema <i>default. default.</i></b>  <i>Na anemia falciforme, o SEA pode ocorrer nos primeiros meses de idade, sendo menos frequente após os anos.</i></p> <p><b>MENSAGEM 9.fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. default.</i></b></p> <p><b>MENSAGEM 10. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. default.</i></b></p> <p><b>MENSAGEM 11. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. default.</i></b></p> <p><b>MENSAGEM 12. fase. descontinuidade:mudança. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. proeminente.</i></b>  <i>A correção da hipovolemia com soluções cristaloides pode ser feita, enquanto se aguarda a transfusão de hemácias.</i></p> <p><b>MENSAGEM 13. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Textual. <i>default.</i></b></p> <p><b>MENSAGEM 14. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Textual. <i>default.</i></b></p> <p><b>MENSAGEM 15. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. default.</i></b></p> <p><b>MENSAGEM 16. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. default.</i></b></p> <p><b>MENSAGEM 17. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. default.</i></b></p> <p><b>MENSAGEM 18. inicial. arranjo explícito (periódico). não focalizado. TEMA: Tema <i>default. proeminente.</i></b>  <i>Pacientes com dor torácica, tosse, dispneia, hipoxemia, prostração e alteração radiológica recente do tórax provavelmente têm síndrome torácica aguda.</i></p> <p><b>MENSAGEM 19. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. default.</i></b></p> <p><b>MENSAGEM 20. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. default.</i></b></p> <p><b>MENSAGEM 21. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. default.</i></b></p> <p><b>MENSAGEM 22. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. default.</i></b></p> <p><b>MENSAGEM 23. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. default.</i></b></p> <p><b>MENSAGEM 24. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. default.</i></b></p> <p><b>MENSAGEM 25. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Textual. <i>default.</i></b></p> <p><b>MENSAGEM 26. fase. descontinuidade:desvio. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. proeminente.</i></b>  <i>A saturimetria, com oximetria de pulso, deve ser instituída.</i></p> <p><b>MENSAGEM 27. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. default.</i></b></p> <p><b>MENSAGEM 28. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. default.</i></b></p> <p><b>MENSAGEM 29. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. default.</i></b></p> <p><b>MENSAGEM 30. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. default.</i></b></p> <p><b>MENSAGEM 31. inicial. arranjo explícito (periódico). não focalizado. TEMA: Tema <i>default. proeminente.</i></b>  <i>O acidente vascular cerebral (AVC) é uma das mais graves complicações da doença falciforme.</i></p> <p><b>MENSAGEM 32. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Textual. <i>default.</i></b></p> <p><b>MENSAGEM 33. fase. descontinuidade:desvio. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. proeminente.</i></b>  <i>O AVC hemorrágico, mais comum em adultos, responde por dos casos e apresenta mais morbimortalidade.</i></p> <p><b>MENSAGEM 34. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Textual. <i>default.</i></b></p> <p><b>MENSAGEM 35. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. default.</i></b></p> <p><b>MENSAGEM 36. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. default.</i></b></p> <p><b>MENSAGEM 37. fase. descontinuidade:desvio. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. proeminente.</i></b>  <i>As crianças a partir dos anos de idade são mais afetadas, com incidência de até os anos.</i></p> <p><b>MENSAGEM 38. inicial. arranjo implícito (em-série). não focalizado. TEMA: Tema <i>default. proeminente.</i></b>  <i>Alterações de consciência, déficits neurológicos focais, convulsões, parestias, afasia, confusão mental e cefaleia de grande intensidade ou duração são sinais e sintomas sugestivos de acidente vascular cerebral.</i></p> <p><b>MENSAGEM 39. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. default.</i></b></p> <p><b>MENSAGEM 40. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. default.</i></b></p> <p><b>MENSAGEM 41. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. default.</i></b></p> <p><b>MENSAGEM 42. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Textual. <i>default.</i></b></p> <p><b>MENSAGEM 43. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. default.</i></b></p> <p><b>MENSAGEM 44. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Textual. <i>default.</i></b></p> <p><b>MENSAGEM 45. fase. continuidade. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. default.</i></b></p> <p><b>MENSAGEM 46. fase. descontinuidade:desvio. arranjo-<i>default.</i> não focalizado. TEMA: Tema <i>default. proeminente.</i></b>  <i>O rastreamento para doença cérebro-vascular pelo doppler transcraniano deve ser iniciado aos anos de idade e repetido anualmente.</i></p>
--

QUADRO 14. Distribuição dos sistemas de TEMA e MENSAGEM no manual técnico em português (cont.)

MENSAGEM 47. fase. continuidade. arranjo- <i>default</i> . não focalizado. TEMA: Tema <i>default</i> . <i>default</i> .
MENSAGEM 48. fase. descontinuidade:mudança. arranjo- <i>default</i> . não focalizado. TEMA: Tema <i>default</i> . proeminente. <i>A exsanguíneo transfusão parcial deve ser iniciada imediatamente após estabilização do paciente, para redução do nível de Hb S para menos de.</i>
MENSAGEM 49. fase. continuidade. arranjo- <i>default</i> . não focalizado. TEMA: Tema <i>default</i> . <i>default</i> .
MENSAGEM 50. inicial. arranjo explícito (periódico). não focalizado. TEMA: Tema <i>default</i> . <i>default</i> . <i>A principal causa de aplasia medular eritrocítica na doença falciforme é a infecção pelo parvovirus.</i>
MENSAGEM 51. fase. continuidade. arranjo- <i>default</i> . não focalizado. TEMA: Textual. <i>default</i> .
MENSAGEM 52. fase. continuidade. arranjo- <i>default</i> . não focalizado. TEMA: Tema <i>default</i> . <i>default</i> .
MENSAGEM 53. fase. continuidade. arranjo- <i>default</i> . não focalizado. TEMA: Tema <i>default</i> . <i>default</i> .
MENSAGEM 54. fase. continuidade. arranjo- <i>default</i> . não focalizado. TEMA: Tema <i>default</i> . <i>default</i> .
MENSAGEM 55. fase. descontinuidade:mudança. arranjo- <i>default</i> . não focalizado. TEMA: Tema <i>default</i> . proeminente. <i>O tratamento consiste na estabilização hemodinâmica pela transfusão de hemácias com monitorização até a elevação dos reticulócitos.</i>
MENSAGEM 56. inicial. arranjo explícito (periódico). não focalizado. TEMA: Tema <i>default</i> . <i>default</i> . <i>O priapismo é a ereção persistente e dolorosa do pênis.</i>
MENSAGEM 57. fase. continuidade. arranjo- <i>default</i> . não focalizado. TEMA: Tema <i>default</i> . <i>default</i> .
MENSAGEM 58. fase. continuidade. arranjo- <i>default</i> . não focalizado. TEMA: Tema <i>default</i> . <i>default</i> .
MENSAGEM 59. fase. continuidade. arranjo- <i>default</i> . não focalizado. TEMA: Tema <i>default</i> . <i>default</i> .
MENSAGEM 60. fase. continuidade. arranjo- <i>default</i> . não focalizado. TEMA: Textual. <i>default</i> .
MENSAGEM 61. fase. descontinuidade:mudança. arranjo- <i>default</i> . não focalizado. TEMA: Tema <i>default</i> . proeminente. <i>Priapismo repetitivo: o tratamento é, frequentemente, domiciliar.</i>
MENSAGEM 62. fase. continuidade. arranjo- <i>default</i> . não focalizado. TEMA: Tema <i>default</i> . <i>default</i> .
MENSAGEM 63. fase. continuidade. arranjo- <i>default</i> . não focalizado. TEMA: Textual. <i>default</i> .
MENSAGEM 64. fase. descontinuidade:mudança. arranjo- <i>default</i> . não focalizado. TEMA: Tema <i>default</i> . proeminente. <i>Priapismo agudo prolongado: as possíveis etiologias são infecções particularmente da próstata, traumas, medicações com efeitos no sistema nervoso central, abuso de álcool, uso de maconha ou após atividade sexual.</i>
MENSAGEM 65. fase. continuidade. arranjo- <i>default</i> . não focalizado. TEMA: Tema <i>default</i> . <i>default</i> .
MENSAGEM 66. fase. continuidade. arranjo- <i>default</i> . não focalizado. TEMA: Textual. <i>default</i> .
MENSAGEM 67. fase. continuidade. arranjo- <i>default</i> . não focalizado. TEMA: Tema <i>default</i> . <i>default</i> .
MENSAGEM 68. fase. continuidade. arranjo- <i>default</i> . não focalizado. TEMA: Tema <i>default</i> . <i>default</i> .
MENSAGEM 69. fase. continuidade. arranjo- <i>default</i> . não focalizado. TEMA: Tema <i>default</i> . <i>default</i> .
MENSAGEM 70. fase. continuidade. arranjo- <i>default</i> . não focalizado. TEMA: Textual. <i>default</i> .
MENSAGEM 71. fase. continuidade. arranjo- <i>default</i> . não focalizado. TEMA: Tema <i>default</i> . <i>default</i> .
MENSAGEM 72. fase. descontinuidade:mudança. arranjo- <i>default</i> . não focalizado. TEMA: Tema. proeminente. <i>Existem três formas clínicas.</i>
MENSAGEM 73. fase. continuidade. arranjo- <i>default</i> . não focalizado. TEMA: Tema <i>default</i> . <i>default</i> .
MENSAGEM 74. fase. continuidade. arranjo- <i>default</i> . não focalizado. TEMA: Textual. <i>default</i> .
MENSAGEM 75. fase. continuidade. arranjo- <i>default</i> . não focalizado. TEMA: Tema <i>default</i> . <i>default</i> .
MENSAGEM 76. fase. continuidade. arranjo- <i>default</i> . não focalizado. TEMA: Tema <i>default</i> . <i>default</i> .
MENSAGEM 77. fase. continuidade. arranjo- <i>default</i> . não focalizado. TEMA: Tema <i>default</i> . <i>default</i> .
MENSAGEM 78. fase. continuidade. arranjo- <i>default</i> . não focalizado. TEMA: Tema <i>default</i> . <i>default</i> .

No texto em português, a Mensagem inicial 1 e suas continuidades são responsáveis por abordar o tratamento de uma das intercorrências da anemia falciforme (a pneumonia). Esse “bloco” de informações inicia-se na Mensagem 1 e termina na Mensagem 8.

Da mesma forma que ocorre com os demais tipos de texto, após as mensagens iniciais, existem as mensagens de continuidade, descontinuidade:mudança e descontinuidade:desvio. A Mensagem 4, por exemplo, que é de descontinuidade:mudança destaca um dos aspectos do tratamento da anemia falciforme. A Mensagem 33, por exemplo, que é de descontinuidade:desvio, enfoca o acidente vascular cerebral (AVC) hemorrágico.

No que toca ao sistema de TEMA dos manuais técnicos em português, todos os tipos de temas ilustrados acima estão tabulados na TAB. 29.

TABELA 29. Frequência relativa das funções do sistema de TEMA no manual técnico em português

Tipos de temas		N	%
Complexidade	elemental	-	-
	ângulo	-	-
	textual	15	19,23
	interpessoal	-	-
	tema- <i>default</i>	63	80,77
	Subtotal	78	100,00
Ângulo	fonte	-	-
	ponto de vista	-	-
	Subtotal	-	-
Seleção Temática	<i>default</i>	67	85,90
	proeminente	11	14,10
	Subtotal	78	100,00

A TAB. 29 mostra a frequência relativa das funções do sistema de TEMA do manual técnico em português, distribuído quanto à complexidade, ângulo e seleção temática. Em relação à complexidade, observa-se que 19,23% são temas textuais (15 ocorrências) e 80,77% são temas-*default* (63 ocorrências). Quanto à seleção temática, 85,90% são *default* (67 ocorrências) e 14,10% são proeminentes (11 ocorrências).

## 2) Manual técnico em inglês

Os dados tabulados na TAB. 30 foram extraídos a partir da análise textual do manual técnico em inglês.

TABELA 30. Frequência relativa das funções do sistema de MENSAGEM no manual técnico em inglês

Tipos de mensagem			N	%	
Contextualização	Inicial		8	15,38	
	Fase	Continuidade	38	86,36	
		Descontinuidade	Desvio	5	83,33
			Mudança	1	16,67
		Subtotal		6	100,00
	Subtotal		44	100,00	
Subtotal			52	100	
Arranjo	<i>Default</i>		44	84,62	
	Arranjo	Explícito (periódico)	7	87,50	
		Implícito (em série)	1	12,50	
		Subtotal		8	100,00
	Subtotal		52	100,00	
Focalização	Não focalizada		52	100,00	
	Focalizada	Particularizada	-	-	
		Contraste	-	-	
		Subtotal		-	-
	Subtotal		52	100,00	

A TAB. 30 ilustra a frequência relativa das funções do sistema de MENSAGEM no manual técnico em inglês. Os dados ilustrados nessa tabela mostram que 15,38% das mensagens do texto são iniciais (oito ocorrências), enquanto 84,62% são fases (44 ocorrências). Dessas fases, 86,36% são de continuidade (38 ocorrências) e 13,64% são de descontinuidade (seis ocorrências). Dos tipos de descontinuidade, 83,33% são desvio (cinco ocorrências) e 16,67% são mudança (uma ocorrência). Todas as fases são não focalizadas (100%). Quanto ao arranjo, 84,62% das mensagens são arranjo *default* (44 ocorrências) e 15,38% são arranjos (8 ocorrências). Dos arranjos, 87,50% são explícitos-periódicos (sete ocorrências) e 12,50% são implícitos-em série (uma ocorrência).

O QUADRO 15, a seguir, ilustra o manual técnico em inglês classificado por TEMA e MENSAGEM para melhor identificação da distribuição desses dois sistemas no texto. Cada uma das mensagens iniciais e das descontinuidades:mudanças e descontinuidades:desvios está

exemplificada em *itálico*. No intuito de explicar o que quer dizer a mensagem, o QUADRO 15 reproduz um exemplo de cada um dos tipos de mensagem.

QUADRO 15. Distribuição dos sistemas de TEMA e MENSAGEM no manual técnico em inglês

<p>MENSAGEM 1. inicial. arranjo explícito (periódico). não focalizado. TEMA: simples. não marcado. <i>The majority of mild to moderate pain episodes abate spontaneously with bed rest.</i></p> <p>MENSAGEM 2. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: múltiplo (textual e tópico). não marcado.</p> <p>MENSAGEM 3. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p>MENSAGEM 4. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p>MENSAGEM 5. inicial. arranjo explícito (periódico). não focalizado. TEMA: simples. não marcado. <i>A number of factors may precipitate a sickle cell crisis, these include: Hypoxia. Acidosis. Dehydration. Infection. Extreme Fatigue. Trauma. Temperature Changes (sudden). Stress / Anxiety. Increase physical / physiological demand (Pregnancy, physical exercise).</i></p> <p>MENSAGEM 6. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: múltiplo (textual e tópico). não marcado.</p> <p>MENSAGEM 7. fase. descontinuidade:mudança. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado. <i>There are a number of short and long term complications associated with sickle cell disease.</i></p> <p>MENSAGEM 8. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: múltiplo (textual e tópico). não marcado.</p> <p>MENSAGEM 9. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: múltiplo (textual e tópico). não marcado.</p> <p>MENSAGEM 10. inicial. arranjo explícito (periódico). não focalizado. TEMA: simples. não marcado. <i>It is recommended that children with sickle cell disease have the full course of childhood immunisations.</i></p> <p>MENSAGEM 11. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p>MENSAGEM 12. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p>MENSAGEM 13. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p>MENSAGEM 14. fase. descontinuidade:desvio. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado. <i>As part of routine care it is also recommended that children have Transcranial Doppler (TCD) scanning of the brain from the age of three years and every year thereafter.</i></p> <p>MENSAGEM 15. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p>MENSAGEM 16. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p>MENSAGEM 17. inicial. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado. <i>In adults the commonest cause of hospital admission is acute vaso-occlusive sickle cell crisis.</i></p> <p>MENSAGEM 18. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: múltiplo (textual e tópico). não marcado.</p> <p>MENSAGEM 19. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p>MENSAGEM 20. fase. descontinuidade:desvio. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado. <i>Laboured or rapid respiration may be the only indication that is observed by a vigilant family member or nurse.</i></p> <p>MENSAGEM 21. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p>MENSAGEM 22. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p>MENSAGEM 23. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p>MENSAGEM 24. inicial. arranjo explícito (periódico). não focalizado. TEMA: simples. não marcado. <i>Leg ulceration is one of the less well understood complications of SCD.</i></p> <p>MENSAGEM 25. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p>MENSAGEM 26. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p>MENSAGEM 27. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p>MENSAGEM 28. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p>MENSAGEM 29. fase. descontinuidade:desvio. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado. <i>Treatment consists of careful cleaning and dressing over a period of weeks or even months, but there remains a high risk of recurrence.</i></p> <p>MENSAGEM 30. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p>MENSAGEM 31. inicial. arranjo explícito (periódico). não focalizado. TEMA: simples. não marcado. <i>Ulcers are one of the few visible and unsightly complications of sickle cell disease.</i></p> <p>MENSAGEM 32. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: múltiplo (textual e tópico). não marcado.</p> <p>MENSAGEM 33. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado. TEMA: múltiplo (textual e tópico). não marcado.</p> <p>MENSAGEM 34. inicial. arranjo explícito (periódico). não focalizado. TEMA: simples. não marcado. <i>Priapism is a persistent and abnormal erection of the penis accompanied by pain, tenderness and swelling.</i></p> <p>MENSAGEM 35. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p>MENSAGEM 36. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: múltiplo (textual e tópico). não marcado.</p> <p>MENSAGEM 37. fase. descontinuidade:desvio. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado. <i>Sickling causes priapism by occluding the circulatory spaces and small blood vessels of the penis.</i></p> <p>MENSAGEM 38. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p>MENSAGEM 39. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p>MENSAGEM 40. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p>MENSAGEM 41. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p>MENSAGEM 42. inicial. arranjo explícito (periódico). arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado. <i>There is no specific treatment for individuals with sickle cell disease.</i></p> <p>MENSAGEM 43. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: múltiplo (textual e tópico). não marcado.</p> <p>MENSAGEM 44. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p>MENSAGEM 45. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: múltiplo (textual e tópico). não marcado.</p> <p>MENSAGEM 46. fase. descontinuidade:desvio. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p>MENSAGEM 47. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p>MENSAGEM 48. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: múltiplo (textual e tópico). não marcado.</p> <p>MENSAGEM 49. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p>MENSAGEM 50. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: múltiplo (textual e tópico). não marcado.</p> <p>MENSAGEM 51. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p>MENSAGEM 52. fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p>
--

A Mensagem inicial 5 e suas continuidades são as responsáveis por apresentar os fatores que desencadeiam uma crise de dor na anemia falciforme. Esse “bloco” de informações inicia-se na Mensagem 1 e termina na Mensagem 9. Após as mensagens iniciais, há as mensagens de continuidade, descontinuidade:mudança e descontinuidade:desvio. As mensagens de descontinuidade:mudança e descontinuidade:desvio estão exemplificadas pelas Mensagens 7 e 37, respectivamente. A Mensagem 7 destaca as complicações associadas à anemia falciforme, e a Mensagem 37 aborda as intercorrências da falcização das hemácias em relação ao priapismo.

No que toca ao sistema de TEMA dos manuais técnicos em inglês, todos os tipos de temas estão tabulados na TAB. 31.

TABELA 31. Frequência relativa das funções do sistema de TEMA no manual técnico em inglês

Tipos de temas		N	%
Complexidade	Simple	41	78,85
	Múltiplo	11	21,15
	Subtotal	52	100,00
Múltiplo	Textual e tópico	11	100,00
	Interpessoal e tópico	-	-
	Textual interpessoal e tópico	-	-
	Subtotal	11	100,00
Seleção temática	Não marcado	52	100,00
	Marcado	-	-
	Subtotal	52	100,00

Em relação à complexidade, observa-se que 78,85% dos temas são simples (41 ocorrências) e 21,15% são múltiplos (11 ocorrências). Dos múltiplos, 100% são temas do tipo textual e tópico. No que toca à seleção temática, 100% dos temas são não marcados.

### 3.3.3 Cartilhas e folders de divulgação: análise ideacional, interpessoal e textual dos textos em português e inglês

A seguir estão apresentadas as análises ideacional, interpessoal e textual das cartilhas e folders de divulgação em português e em inglês (cf. Subseções 3.3.3.1, 3.3.3.2 e 3.3.3.3)

#### 3.3.3.1 Análise ideacional das cartilhas e folders de divulgação em português e em inglês

A TAB. 32 mostra a frequência relativa das funções do sistema de TRANSITIVIDADE nas cartilhas e folders de divulgação em português e em inglês:

TABELA 32. Frequência relativa das funções do sistema de TEMA nos manuais em português e inglês

Tipo de processo	Português		Inglês	
	N	%	N	%
Material	7	12,07	11	22
Relacional	Identificativo	4	8	5,13
	Atributivo	46	92	94,87
	Subtotal	50	100	39
Mental	-	-	-	-
Existencial	-	-	-	-
Comportamental	-	-	-	-
Verbal	1	1,72	-	-
<b>TOTAL</b>	<b>63</b>	<b>100</b>	<b>49</b>	<b>100,00</b>

Exemplifica-se a seguir cada um dos tipos de processos (com destaque em **negrito**) e suas principais características.

#### 1) Processos Materiais

Os dados tabulados na TAB. 32 mostram que os processos materiais na cartilha e folder de divulgação em português correspondem a 12,07% (sete ocorrências) do total de 58 ocorrências. Nesse tipo de texto, os processos materiais apresentam as ações que devem ser tomadas no tratamento dos pacientes, como mostram os exemplos (38) e (39):

(38) (...) e de se **instalar** imediatamente analgesia adequada

(39) (...) e **procurar** atendimento médico quando necessário

Já nas cartilhas e *folders* de divulgação em inglês, os processos materiais correspondem a 22% (11 ocorrências) do total de 50 ocorrências. Esses processos são responsáveis por indicar o que acontece com cada célula ou órgão, conforme ilustram os exemplos (40) e (41):

(40) When the red blood cells come to the end of their useful life they **are broken down** in the body.

(41) The liver **clears** the bilirubin from the body.

## 2) Processos Relacionais

Os processos relacionais nas cartilhas e *folders* de divulgação em português correspondem a 86,21% (46 ocorrências) do total de 50 ocorrências. Desse total, 8% (quatro ocorrências) são processos relacionais identificativos e 92% (46 ocorrências) são processos relacionais atributivos.

Na cartilha e *folder* de divulgação em português, os processos relacionais identificativos apresentam características da anemia falciforme e seus sintomas, conforme ilustrado nos exemplos (42) e (43).

(42) A doença falciforme é uma alteração genética.

(43) A primeira manifestação de dor na maioria das crianças é a dactilite (ou síndrome mão-pé).

No que toca aos processos relacionais atributivos nos textos em português, os exemplos (44) e (45) mostram as características dos sintomas das intercorrências da anemia falciforme:

(44) O portador assintomático de falciforme, também conhecido por portador do traço de Hb S ou heterozigoto para a Hb S, não é anêmico.

(45) Elas **são** causadas pelo dano tissular isquêmico secundário à obstrução do fluxo sanguíneo pelas hemácias falcizadas.

Já na cartilha e *folder* de divulgação em inglês, os processos relacionais correspondem a 78% (39 ocorrências) do total de 50 ocorrências. Desse total, 5,13% (duas ocorrências) são processos relacionais identificativos e 94,87% (37 ocorrências) são processos relacionais atributivos.

Na cartilha e *folder* de divulgação em inglês, os processos relacionais identificativos apresentam características dos órgãos e funções do corpo humano, conforme ilustrado nos exemplos (46) e (47).

(46) One of the substances that are produced **is** a yellow pigment called bilirubin.

(47) The spleen **is** an organ that lies on the left side of the tummy under the rib cage.

Em relação aos processos relacionais atributivos, observa-se pelos exemplos (48) e (49) que eles indicam características da anemia falciforme:

(48) SBeta Thal occurs if your child **has** inherited a haemoglobin gene from one parent and a persisting haemoglobin gene from the other parent.

(49) Sickle cell trait means that a person **has** inherited one normal haemoglobin A gene and one sickle haemoglobin gene from their parents.

### 3) Processos verbais

Há apenas uma ocorrência de processo verbal, a qual se encontra na cartilha e *folder* de divulgação em português, conforme pode ser observado na TAB. 32, o que corresponde a 1,72% (1 ocorrência) do total de 58 ocorrências. Esse processo representa o que dizem pacientes sobre os seus próprios sintomas, conforme consta no exemplo (50).

(50) Os pacientes mais velhos **citam** que a depressão e exaustão física podem ser fatores precipitantes das crises.

### 4) Consolidação

A análise dos processos dos textos do tipo HABILITAR possibilitou entender a maneira como o mundo das experiências (*cf.* FIG. 56) é construído nas cartilhas e *folders* de divulgação com o intuito de orientar a conduta do leitor. O autor, detentor do conhecimento, elabora uma cartilha ou *folder* de divulgação para que seja utilizado como base de informação para a triagem neonatal para a doença falciforme, mais especificamente com esclarecimentos em relação à anemia falciforme que orientem as intercorrências da doença.

#### 3.3.3.2 Análise interpessoal das cartilhas e *folders* de divulgação em português e em inglês

A TAB. 33 mostra a frequência relativa das funções do sistema de MODO nas cartilhas e *folders* de divulgação em português e em inglês:

TABELA 33. Frequência relativa das funções do sistema de MODO e Funções Discursivas nas cartilhas e *folders* de divulgação em português e inglês

Tipo de Modo e Funções discursivas			Português		Inglês	
			N	%	N	%
Tipo de Modo	Indicativo	Declarativo	57	100	50	100
		Interrogativo	-	-	-	-
		Subtotal	57	100	49	98
	Imperativo	-	-	1	2	
Funções discursivas	Turno	Inicial	57	100	50	100
		Correspondente	-	-	-	-
		Subtotal	57	100	50	100
	Papel	Oferta	57	100	49	98
		Demanda	-	-	1	2
		Subtotal	57	100	50	100
	Tipo de Mercadoria	Informação	57	100	49	98
		Bens e serviços	-	-	1	2
		Subtotal	57	100	50	100

Na cartilha e *folder* de divulgação em português, 100% das ocorrências de Modo (57 ao todo) correspondem ao Indicativo Declarativo. Esse Modo está explicitado nos exemplos (50) e (51):

- (50) O termo doença falciforme define as hemoglobinopatias nas quais pelo menos uma das hemoglobinas mutantes é a HbS.
- (51) O diagnóstico laboratorial da doença falciforme é realizado pela detecção da Hb S e da sua associação com outras frações.

Já na cartilha e *folder* de divulgação em inglês, o Modo Indicativo Declarativo é encontrado em 98% dos casos (49 do total de 50 ocorrências). Esse Modo está ilustrado nos exemplos (52) e (53):

- (52) In this book we will mainly write about sickle cell disease as if it is one condition although we recognize that the different types are not all the same in the way they affect a person's health.
- (53) Having a trait is commonly known as being a carrier and does not affect the person in any way.

Ainda na cartilha e *folder* de divulgação em inglês, encontra-se uma ocorrência do Modo Imperativo. Essa ocorrência tem como função semântica a demanda de bens e serviços,

ilustrando uma das ações que podem ser tomadas pelo leitor para obter mais informações sobre um tipo específico de doença. Veja o exemplo (54).

(54) For more information about beta thalassaemia contact the Thalassaemia Society or one of the specialist Centres.

Por se tratar de textos do tipo HABILITAR, observa-se que há oferta de informações, visando instruir o público leigo. Essa instrução é passada pelos especialistas, que são detentores do conhecimento sobre as intercorrências, urgências/emergências e medidas profiláticas da doença falciforme.

### 3.3.3.3 Análise textual das cartilha e *folder* de divulgação em português e em inglês

Como para os outros tipos de texto, a análise textual das cartilhas e *folders* de divulgação é descrita a seguir de forma separada entre as línguas, uma vez que os sistemas de TEMA em português e em inglês apresentam diferenças na caracterização.

#### 1) Cartilha e *folder* de divulgação em português

Os dados tabulados na TAB. 34 foram extraídos a partir da análise textual da cartilha e *folder* de divulgação em português.

TABELA 34. Frequência relativa das funções do sistema de MENSAGEM na cartilha e *folder* de divulgação em português

Tipos de mensagem			N	%	
Contextualização	Inicial		8	14,04	
	Fase	Continuidade	43	87,76	
		Descontinuidade	Desvio	1	16,67
			Mudança	5	83,33
		Subtotal		6	100,00
	Subtotal		43	87,76	
Subtotal			57	100,00	
Arranjo	<i>Default</i>		49	85,96	
	Arranjo	Explícito (periódico)	7	87,50	
		Implícito (em série)	1	12,50	
		Subtotal		8	100,00
	Subtotal			57	100,00
Focalização	Não focalizada		57	100,00	
	Focalizada	Particularizada	-	-	
		Contraste	-	-	
		Subtotal		-	-
	Subtotal			57	100,00

Os dados ilustrados nessa tabela mostram que 14,04% das mensagens do texto são iniciais (oito ocorrências), enquanto 85,96% são fases (49 ocorrências). Dessas fases, 87,76% são de continuidade (43 ocorrências) e 12,24% são de descontinuidade (6 ocorrências). Dos tipos de descontinuidade, 16,67% são desvio (uma ocorrência) e 83,33% são mudança (cinco ocorrências). Todas as fases são não focalizadas (100%). Quanto ao arranjo, 85,96% das mensagens são *default* (49 ocorrências) e 14,04% são arranjos (oito ocorrências). Dos arranjos, 87,50% são explícitos (sete ocorrências) e 12,50% são implícitos (uma ocorrência).

O QUADRO 16 a seguir mostra a descrição da cartilha e *folder* de divulgação em português classificado por TEMA e MENSAGEM para melhor identificação da distribuição desses dois sistemas no texto. Cada uma das mensagens iniciais e das descontinuidades:mudanças e descontinuidades:desvios está exemplificada em itálico.

QUADRO 16. Distribuição dos sistemas de TEMA e MENSAGEM  
na cartilha e *folder* de divulgação em português

<p>MENSAGEM 1. inicial. arranjo explícito (periódico). não focalizado. TEMA: Tema default. default. <i>A doença falciforme é uma alteração genética.</i></p> <p>MENSAGEM 2. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 3. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</p> <p>MENSAGEM 4. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 5. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 6. fase. descontinuidade:mudança. arranjo-default. não focalizado. Tema default. proeminente. <i>As doenças falciformes mais frequentes são a anemia falciforme, a S talassemia ou microdrepanocitose e as duplas heterozigoses Hb SC e Hb SD.</i></p> <p>MENSAGEM 7. inicial. arranjo explícito (periódico). não focalizado. TEMA: Textual. default. <i>Para o diagnóstico seguro de uma das três situações acima é de fundamental importância conhecer a forma de herança da doença falciforme.</i></p> <p>MENSAGEM 8. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 9. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 10. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 11. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 12. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 13. inicial. arranjo explícito (periódico). não focalizado. TEMA: Tema default. default. <i>O portador assintomático de falciforme, também conhecido por portador do traço de Hb S ou heterozigoto para a Hb S, não é anêmico.</i></p> <p>MENSAGEM 14. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 15. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</p> <p>MENSAGEM 16. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 17. inicial. arranjo explícito (periódico). não focalizado. TEMA: Tema default. default. <i>A Hb S tem uma característica química especial que em situações de ausência ou diminuição da tensão de oxigênio provoca a sua polimerização.</i></p> <p>MENSAGEM 18. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 19. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 20. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 21. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 22. fase. descontinuidade:mudança. arranjo-default. não focalizado. TEMA: Tema default. proeminente. <i>É importante destacar que a frequência de portadores assintomáticos de falciforme no Brasil é de conforme a intensidade da população negra inserida na região.</i></p> <p>MENSAGEM 23. inicial. arranjo explícito (periódico). não focalizado. TEMA: Interpessoal. default. <i>Como se sabe, a HbS teve origem no continente africano, e sua introdução no Brasil ocorreu notadamente durante o período da escravidão.</i></p> <p>MENSAGEM 24. fase. descontinuidade:mudança. arranjo-default. não focalizado. TEMA: Tema default. proeminente. <i>Dados da Triagem Nacional sinalizam nascidos por ano com a doença falciforme e com o traço.</i></p> <p>MENSAGEM 25. inicial. arranjo explícito (periódico). não focalizado. TEMA: Tema default. default. <i>O diagnóstico laboratorial da doença falciforme é realizado pela detecção da Hb S e da sua associação com outras frações.</i></p> <p>MENSAGEM 26. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 27. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 28. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 29. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 30. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 31. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</p> <p>MENSAGEM 32. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</p> <p>MENSAGEM 33. inicial. arranjo explícito (periódico). não focalizado. TEMA: Tema default. default. <i>Em o Ministro da Saúde lançou o Programa Nacional de Triagem Neonatal por meio da Portaria Ministerial n.º 822.</i></p> <p>MENSAGEM 34. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 35. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 36. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</p> <p>MENSAGEM 37. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 38. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</p> <p>MENSAGEM 39. inicial. arranjo explícito (periódico). não focalizado. TEMA: Tema default. default. TEMA: Tema default. default.</p> <p><i>A seguir, apresentamos um quadro sobre as principais características laboratoriais dos diferentes tipos de doença falciforme.</i></p> <p>MENSAGEM 40. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 41. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</p> <p>MENSAGEM 42. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 43. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 44. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 45. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 46. fase. descontinuidade:mudança. arranjo-default. não focalizado. TEMA: Tema default. proeminente. <i>Os pacientes mais velhos citam que a depressão e exaustão física podem ser fatores precipitantes das crises.</i></p> <p>MENSAGEM 47. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 48. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 49. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 50. fase. descontinuidade:mudança. arranjo-default. não focalizado. TEMA: Tema default. proeminente. <i>A dor abdominal pode simular abdômen agudo cirúrgico ou infeccioso, ou processos ginecológicos.</i></p> <p>MENSAGEM 51. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 52. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 53. fase. descontinuidade:mudança. arranjo-default. não focalizado. TEMA: Tema default. proeminente. <i>O tratamento consiste em eliminar os fatores precipitantes.</i></p> <p>MENSAGEM 54. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 55. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 56. fase. continuidade. arranjo-default. não focalizado. TEMA: Tema default. default.</p> <p>MENSAGEM 57. fase. continuidade. arranjo-default. não focalizado. TEMA: Textual. default.</p>
--

Nesse caso, a Mensagem Inicial 13 e suas continuidades são as responsáveis por apresentar os sintomas da anemia falciforme e demais hemoglobinopatias. Esse “bloco” de informações inicia-se na Mensagem 13 e termina na Mensagem 16.

Dando sequência às mensagens iniciais, há as mensagens de continuidade, descontinuidade:mudança e descontinuidade:desvio. A Mensagem 22 é um exemplo de descontinuidade:mudança e destaca dados sobre a frequência de portadores falcêmicos assintomáticos no Brasil. Já as mensagens de descontinuidade:desvio podem ser exemplificadas pela Mensagem 46, que destaca os pacientes falcêmicos em relação às suas crises dolorosas.

Em relação ao sistema de TEMA das cartilhas e *folders* de divulgação, todos os tipos de temas mostrados anteriormente estão tabulados na TAB. 35.

TABELA 35. Frequência relativa das funções do sistema de TEMA nas cartilhas e *folders* de divulgação em português

Tipos de temas		N	%
Complexidade	Elemental	-	-
	Ângulo	-	-
	Textual	12	21,06
	Interpessoal	1	1,75
	Tema- <i>default</i>	44	77,19
	Subtotal	57	100,00
Ângulo	Fonte	-	-
	Ponto de vista	-	-
	Subtotal	-	-
Seleção Temática	<i>Default</i>	51	89,47
	Proeminente	6	10,53
	Subtotal	57	100,00

A TAB. 35 mostra a frequência relativa das funções do sistema de TEMA na cartilha e *folder* de divulgação em português, distribuído quanto à complexidade, ângulo e seleção temática. Em relação à complexidade, 21,06% são temas textuais (12 ocorrências), 1,75% é tema interpessoal (uma ocorrências) e 77,19% são temas-*default* (44 ocorrências). Quanto à seleção temática, 89,47% dos temas são *default* (51 ocorrências) e 10,53% são proeminentes (seis ocorrências).

3) Cartilha e *folder* de divulgação em inglês

A análise textual da cartilha e *folder* de divulgação possibilitou o levantamento dos dados tabulados nas TAB. 36.

TABELA 36. Frequência relativa das funções do sistema de MENSAGEM na cartilha e *folder* de divulgação em inglês

Tipos de mensagem			N	%	
Contextualização	Inicial		7	14,00	
	Fase	Continuidade	34	79,07	
		Descontinuidade	Desvio	5	55,56
			Mudança	4	44,44
		Subtotal		9	100,00
	Subtotal		43	100,00	
Subtotal			50	100,00	
Arranjo	<i>Default</i>		43	86	
	Arranjo	Explícito (periódico)	7	100,00	
		Implícito (em série)	-	-	
		Subtotal	7	100,00	
Subtotal			50	100,00	
Focalização	Não focalizada		50	100,00	
	Focalizada	Particularizada	-	-	
		Contraste	-	-	
		Subtotal	-	-	
Subtotal			50	100,00	

A TAB. 36 ilustra a frequência relativa das funções do sistema de MENSAGEM na cartilha e *folder* de divulgação em inglês. Os dados mostram que 14% das mensagens do texto são iniciais (sete ocorrências), enquanto 86% são fases (43 ocorrências). Dessas fases, 79,07% são de continuidade (34 ocorrências) e 20,93% são de descontinuidade (9 ocorrências). Dos tipos de descontinuidade, 55,56% são desvios (cinco ocorrências) e 44,44% são mudanças (quatro ocorrências). Todas as fases são não focalizadas (100%). Quanto ao arranjo, 86% das mensagens são arranjos *default* (43 ocorrências) e 14% são arranjos (sete ocorrências). Dos arranjos, 100% do total de 7 ocorrências são explícitos (periódicos).

O QUADRO 17 a seguir mostra o “retrato” da cartilha e *folder* de divulgação em inglês, classificado por TEMA e MENSAGEM para melhor identificação da distribuição desses dois sistemas no texto. Cada uma das mensagens iniciais e das descontinuidades:mudanças e descontinuidades:desvios está exemplificada em itálico. Um exemplo de cada um dos tipos de mensagem foi selecionado para fins de ilustração a seguir.

A Mensagem Inicial 18 do texto em inglês, bem como suas continuidades, é responsável por apresentar as características genéticas da anemia falciforme. Esse “bloco” de informações inicia-se na Mensagem 18 e termina na Mensagem 27.

As mensagens de descontinuidade:mudança e descontinuidade:desvio são ilustradas por meio das Mensagens 32 e 8, respectivamente. A Mensagem 32 indica uma circunstância sobre a produção maior de glóbulos vermelhos em crianças com anemia falciforme e sua relação com a circulação sanguínea; a Mensagem 8 destaca o assunto que o especialista deseja informar sobre a anemia falciforme na cartilha.

QUADRO 17. Distribuição dos sistemas de TEMA e MENSAGEM na cartilha e *folder* de divulgação em inglês

<p><b>MENSAGEM 1.</b> inicial. arranjo explícito (periódico). não focalizado. TEMA: múltiplo (textual e tópico) não marcado. <i>When no haemoglobin A is produced the condition is known as sickle beta zero thalassaemia.</i></p> <p><b>MENSAGEM 2.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p><b>MENSAGEM 3.</b> fase. descontinuidade:desvio. arranjo-<i>default</i>. não focalizado. TEMA: simples. marcado. <i>For more information about beta thalassaemia contact the Thalassaemia Society or one of the specialist Centres.</i></p> <p><b>MENSAGEM 4.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p><b>MENSAGEM 5.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p><b>MENSAGEM 6.</b> inicial. arranjo explícito (periódico). não focalizado. TEMA: simples. não marcado. <i>Sickle haemoglobin D disease: is a more unusual form of sickle cell disease.</i></p> <p><b>MENSAGEM 7.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p><b>MENSAGEM 8.</b> fase. descontinuidade:desvio. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado. <i>In this book we will mainly write about sickle cell disease as if it is one condition although we recognize that the different types are not all the same in the way they affect a person's health.</i></p> <p><b>MENSAGEM 9.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p><b>MENSAGEM 10.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. marcado.</p> <p><b>MENSAGEM 11.</b> inicial. arranjo explícito (periódico). não focalizado. TEMA: simples. não marcado. <i>Having a trait is commonly known as being a carrier and does not affect the person in any way.</i></p> <p><b>MENSAGEM 12.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p><b>MENSAGEM 13.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. marcado.</p> <p><b>MENSAGEM 14.</b> fase. descontinuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado. <i>Sickle cell trait means that a person has inherited one normal haemoglobin_A gene and one sickle haemoglobin gene from their parents.</i></p> <p><b>MENSAGEM 15.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p><b>MENSAGEM 16.</b> fase. descontinuidade:desvio. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado. <i>The importance of knowing if one has sickle cell trait or any other unusual haemoglobin trait is because the unusual haemoglobin gene can be passed on to children through genetic inheritance.</i></p> <p><b>MENSAGEM 17.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p><b>MENSAGEM 18.</b> inicial. arranjo explícito (periódico). não focalizado. TEMA: simples. não marcado. <i>It is thought that the sickle cell haemoglobin gene was a small change (mutation) in the haemoglobin gene many thousands of years ago in countries where malaria was common.</i></p> <p><b>MENSAGEM 19.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p><b>MENSAGEM 20.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p><b>MENSAGEM 21.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p><b>MENSAGEM 22.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p><b>MENSAGEM 23.</b> fase. descontinuidade:desvio. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado. <i>In areas of the world where malaria is a danger people who have sickle cell trait were more likely to survive malaria than those who have the usual hemoglobin S.</i></p> <p><b>MENSAGEM 24.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p><b>MENSAGEM 25.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: múltiplo (interpessoal e tópico). não marcado.</p> <p><b>MENSAGEM 26.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p><b>MENSAGEM 27.</b> inicial. arranjo explícito (periódico). não focalizado. TEMA: simples. marcado. <i>During the first three to six months of life, your child may not show signs of having sickle cell disease.</i></p> <p><b>MENSAGEM 28.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p><b>MENSAGEM 29.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p><b>MENSAGEM 30.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p><b>MENSAGEM 31.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p><b>MENSAGEM 32.</b> fase. descontinuidade:mudança. arranjo-<i>default</i>. não focalizado. TEMA: simples. marcado. <i>When a child is making a lot more Haemoglobin their red blood cells will not live in the circulation as long as cells that contain the usual Haemoglobin S.</i></p> <p><b>MENSAGEM 33.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p><b>MENSAGEM 34.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p><b>MENSAGEM 35.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p><b>MENSAGEM 36.</b> fase. descontinuidade:desvio. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado. <i>This sort of anaemia is known as a haemolytic anaemia.</i></p> <p><b>MENSAGEM 37.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p><b>MENSAGEM 38.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p><b>MENSAGEM 39.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: múltiplo (textual e tópico). não marcado.</p> <p><b>MENSAGEM 40.</b> fase. descontinuidade:mudança. arranjo-<i>default</i>. não focalizado. TEMA: simples. marcado. <i>In the United Kingdom, most children get enough folic acid from their normal diet.</i></p> <p><b>MENSAGEM 41.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: múltiplo (textual e tópico). não marcado.</p> <p><b>MENSAGEM 42.</b> inicial. arranjo explícito (periódico). não focalizado. TEMA: simples. não marcado. <i>When the red blood cells come to the end of their useful life they are broken down in the body.</i></p> <p><b>MENSAGEM 43.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p><b>MENSAGEM 44.</b> fase. descontinuidade:desvio. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado. <i>The liver clears the bilirubin from the body.</i></p> <p><b>MENSAGEM 45.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: múltiplo (textual e tópico). não marcado.</p> <p><b>MENSAGEM 46.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p><b>MENSAGEM 47.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p><b>MENSAGEM 48.</b> inicial. arranjo explícito (periódico). não focalizado. TEMA: simples. não marcado. <i>The spleen is an organ that lies on the left side of the tummy under the rib cage.</i></p> <p><b>MENSAGEM 49.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p> <p><b>MENSAGEM 50.</b> fase. continuidade. arranjo-<i>default</i>. não focalizado. TEMA: simples. não marcado.</p>
---

Considerando-se o sistema de TEMA da cartilha e *folder* de divulgação, a TAB. 37 mostra uma consolidação de todos os tipos de temas mencionados anteriormente.

TABELA 37. Frequência relativa das funções do sistema de TEMA na cartilha e *folder* de divulgação em inglês

Tipos de temas		N	%
Complexidade	Simple	41	82,00
	Múltiplo	9	18,00
	Subtotal	50	100,00
Múltiplo	Textual e tópico	8	88,89
	Interpessoal e tópico	1	11,11
	Textual interpessoal e tópico	-	-
	Subtotal	9	100,00
Seleção temática	Não marcado	48	96,00
	Marcado	2	4,00
	Subtotal	50	100,00

A TAB. 37 mostra a frequência relativa das funções do sistema de TEMA na cartilha e *folder* de divulgação, distribuído quanto à complexidade e seleção temática. Em relação à complexidade, 82% dos temas são simples (41 ocorrências) e 18% são múltiplos (nove ocorrências). Dos múltiplos, 88,89% são temas do tipo textual e tópico (oito ocorrências) e 11,11% são interpessoal e tópico (uma ocorrência). No que toca à seleção temática, 96% dos temas são não marcados (48 ocorrências) e 4% são marcados (duas ocorrências).

Após a apresentação tabular de todos os dados a partir do UAM CorpusTool®, objetivou-se ilustrar graficamente o perfil metafuncional dos três diferentes tipos de textos em português e em inglês. Para tanto, os dados tabulados foram transformados em gráficos tridimensionais (*cf.* 2.1.2.3), segregados por tipo de texto e língua, conforme ilustrado na FIG. 57, a seguir.

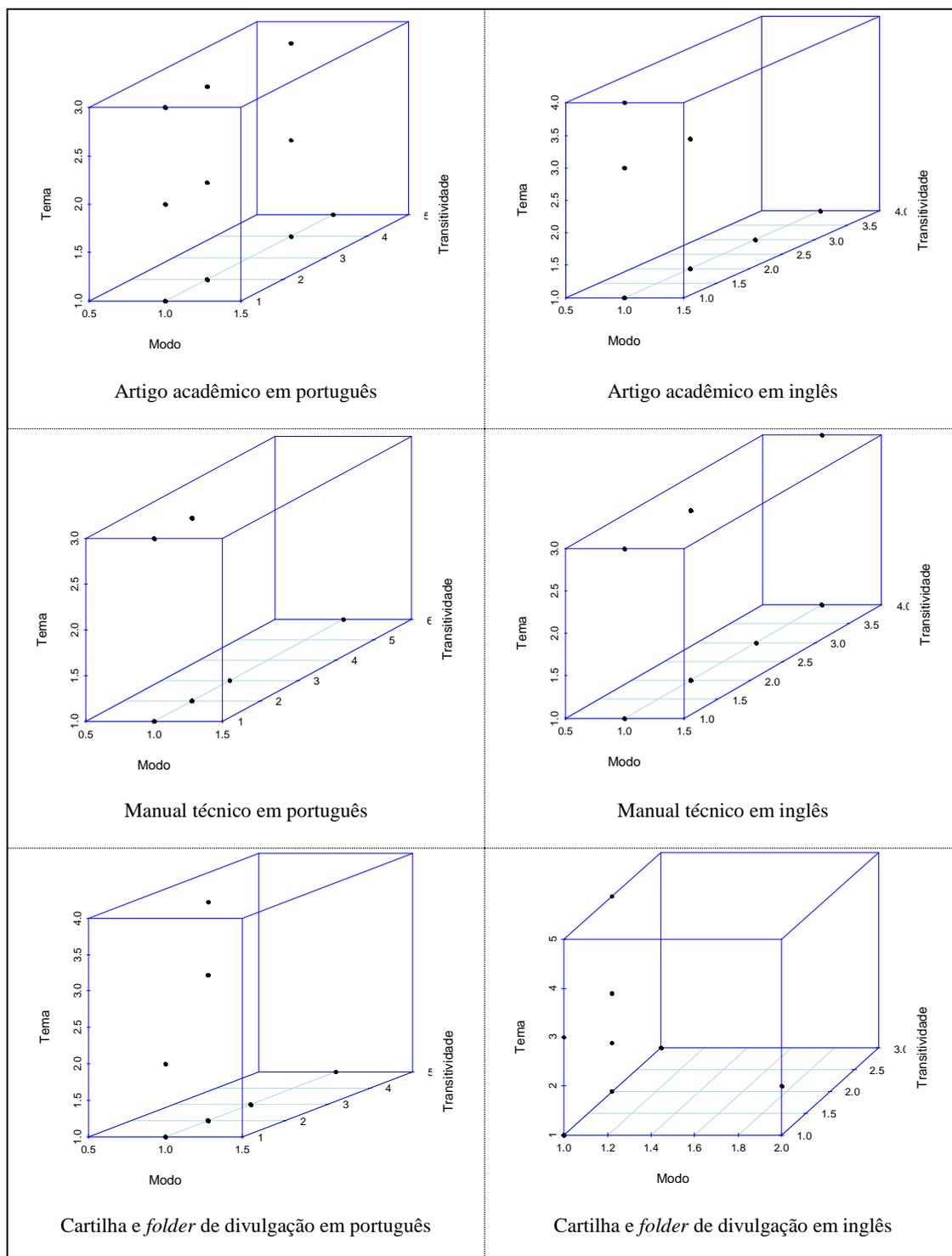


FIGURA 57. Representação gráfica tridimensional dos perfis metafuncionais dos tipos de textos EXPLORAR e HABILITAR

O diagrama tridimensional permite organizar espacialmente o local de interseção entre os pontos no espaço cartesiano XYZ (MODO = X, TEMA = Y e TRANSITIVIDADE = Z). A análise desses pontos permite visualizar a classificação para cada uma das orações em cada tipo de texto, considerando-se os sistemas de TRANSITIVIDADE, MODO e TEMA (cf. Subseção 2.1.2.3). Esses pontos são usados para representar simultaneamente os valores

categóricos das três variáveis quantitativas medidas em cada elemento do conjunto de dados de forma estática (*cf.* FIG. 39, Seção 2.2), ou seja, esses valores categóricos estão representados de acordo com os tipos de processos (material, relacional, mental, existencial, comportamental e verbal), os tipos de Modo (*i.e.* Indicativo Declarativo, Indicativo Interrogativo e Imperativo) e os tipos de Tema (*i.e.* Português: **complexidade:** elemental, ângulo – fonte, ponto de vista –, textual, interpessoal, tema-*default*; **seleção temática:** *default* e proeminente / Inglês: **complexidade:** simples, múltiplo – textual e tópico, interpessoal e tópico, textual interpessoal e tópico –; **seleção temática:** marcado e não marcado), conforme explicado no Capítulo 2 (*cf.* Subseção 2.1.2.3).

Na tentativa de uma melhor visualização da distribuição das orações no espaço geográfico, os dados categóricos foram organizados em dendrogramas, os quais permitem identificar aglomerações dos dados obtidos por meio da análise das funções dos sistemas de MODO, TRANSITIVIDADE e TEMA – mais especificamente, com base nos dados das tabelas construídas com base nos dados do UAM CorpusTool (*cf.* Anexo 2). Cada ponto ilustrado no gráfico tridimensional (*cf.* FIG. 57) refere-se a um agrupamento das orações que foi formado pelo dendrograma no momento da distância zero. Isso significa que, nesse momento, a distância Euclidiana ainda não foi calculada e as orações foram organizadas de acordo com as semelhanças existentes com base na classificação dos sistemas de TRANSITIVIDADE, MODO e TEMA. Esses dendrogramas apresentam a forma como as orações foram distribuídas, ilustrando graficamente a localização das orações em cada um dos pontos no gráfico.

As FIG. 58 a 60 ilustram os dendrogramas construídos a partir dos dados tabulados no UAM CorpusTool para os artigos acadêmicos, manuais técnicos e cartilhas e *folders* de divulgação em português e em inglês, respectivamente. Os dados são diferenciados ou pelo sistema de TRANSITIVIDADE, MODO ou TEMA. No caso desta pesquisa, notou-se que há o predomínio do Modo Indicativo Declarativo: oferta de informação. Há uma ocorrência do Modo Imperativo na cartilha e *folder* de divulgação em inglês. Para uma melhor organização do texto, as tabelas geradas a partir dos dados extraídos por meio desses sistemas, bem como as legendas referentes a elas, não foram incluídas no corpo do texto. Para a análise dos dendrogramas, faz-se necessário consultá-las (*cf.* Anexo 2).

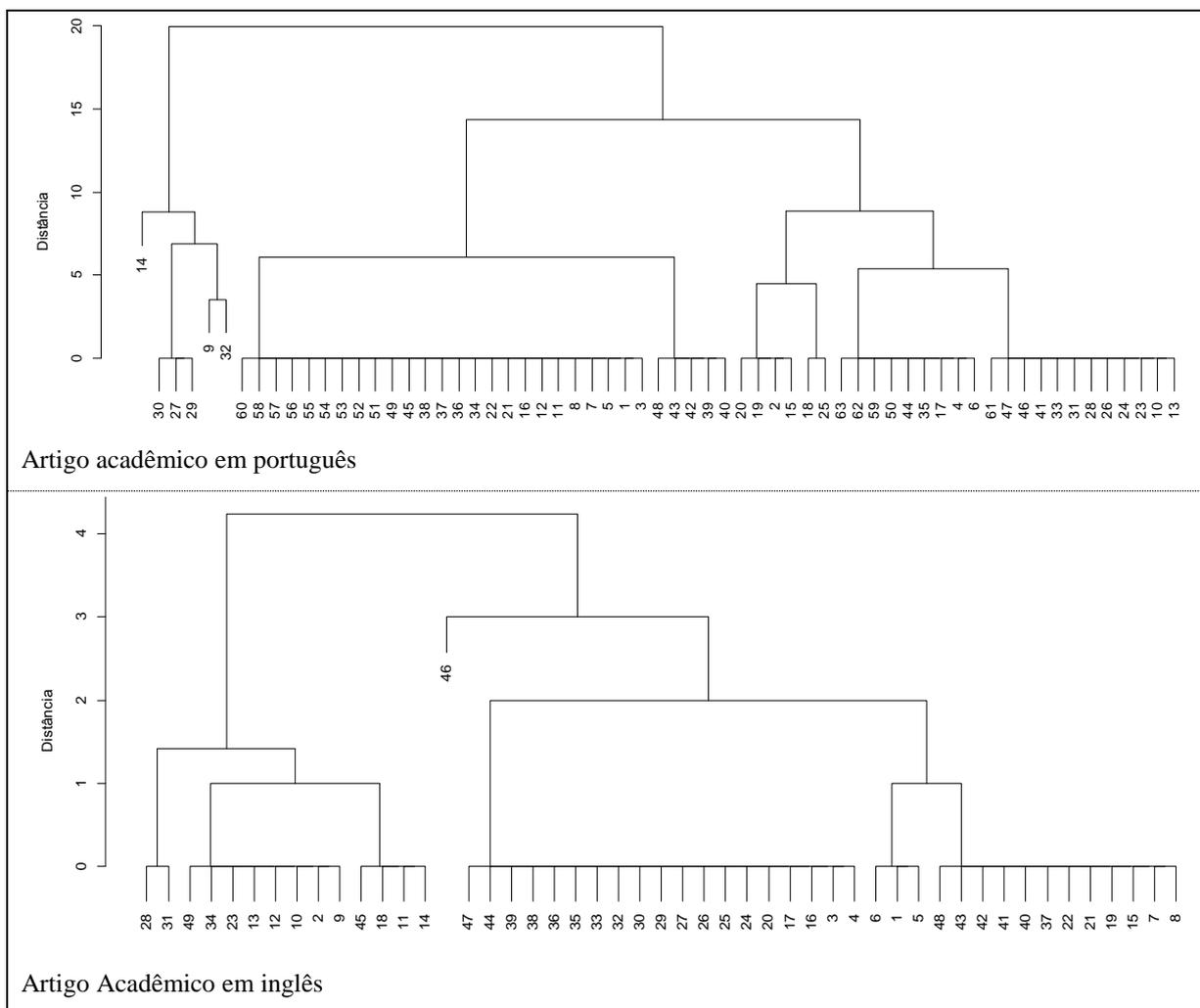


FIGURA 58. Dendrograma dos artigos científicos em português e em inglês

A FIG. 58 mostra como as orações foram agrupadas de acordo com as semelhanças entre elas a partir dos dados categóricos atribuídos aos sistemas de TRANSITIVIDADE, MODO e TEMA. O texto em português apresenta 63 orações agrupadas de acordo com as semelhanças existentes entre elas.

A oração 14 do artigo científico em português, por exemplo, destaca-se das demais em virtude de sua classificação: o MODO( $X$ ) = 1, a TRANSITIVIDADE ( $Y$ ) = 5 (processo verbal) e o TEMA( $Z$ ) = 1 (*default*). O distanciamento dessa oração em relação às demais ocorre em função do tipo de processo. Apesar dessa distância, essa oração guarda proporções de semelhança com as orações 9 ( $X = 1$ ;  $Y = 4$ ;  $Z = 1$ ) e 32 ( $X = 1$ ;  $Y = 4$ ;  $Z = 2$ ). Observa-se que a diferença entre as orações 9 e 32 concentra-se no tipo de processo. A oração 9 se distancia da oração 14 em função da transitividade, e a 32 distancia-se da oração 14 em função da transitividade e tema.

Já o texto em inglês apresenta 57 orações, das quais uma se destaca: a oração 46 ( $X = 1$ ;  $Y = 5$ ;  $Z = 1$ ). As orações guardam proporções de semelhança entre si dentro de cada agrupamento, dependendo da classificação em relação à TRANSITIVIDADE, MODO e TEMA; por isso, o dendrograma se divide em pequenos grupos. Percebe-se, portanto, a diferença gráfica entre os artigos científicos em português e em inglês.

A FIG. 59 apresenta os dendrogramas dos manuais técnicos em português e em inglês.

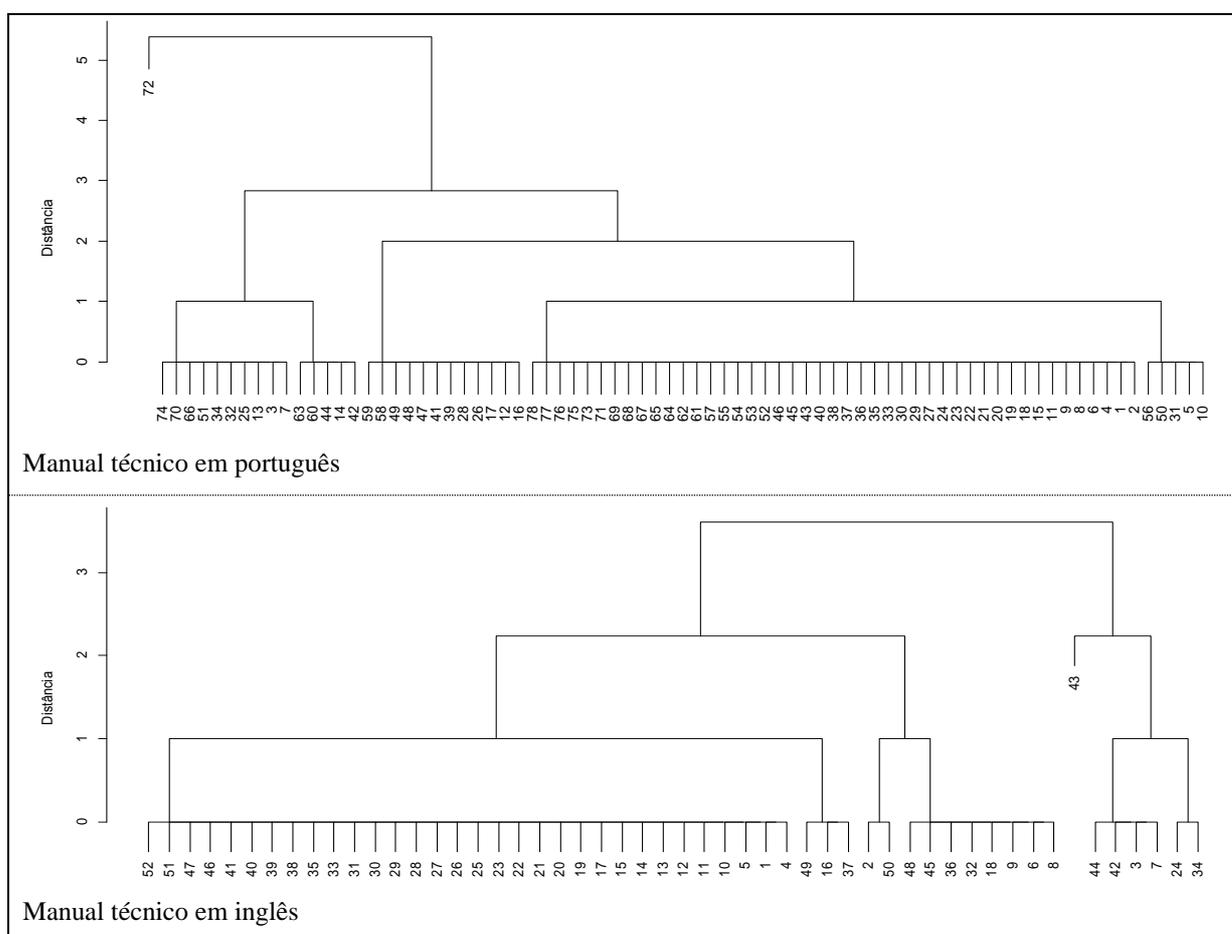


FIGURA 59. Dendrograma dos manuais técnicos em inglês e em português

Como mostra a FIG. 59, os manuais técnicos em português e em inglês apresentam 78 e 52 orações, respectivamente. No manual técnico em português, a oração 72 distancia-se das demais ( $X=1$ ,  $Y=6$ ,  $Z=1$ ), sendo a única oração que apresenta processo existencial. Já em inglês, a oração 43 ( $X=1$ ,  $Y=4$ ,  $Z=3$ ) distancia-se das demais por apresentar processo mental e tema textual *default*.

Por fim, a FIG. 60 apresenta os dendrogramas das cartilhas e *folders* de divulgação.

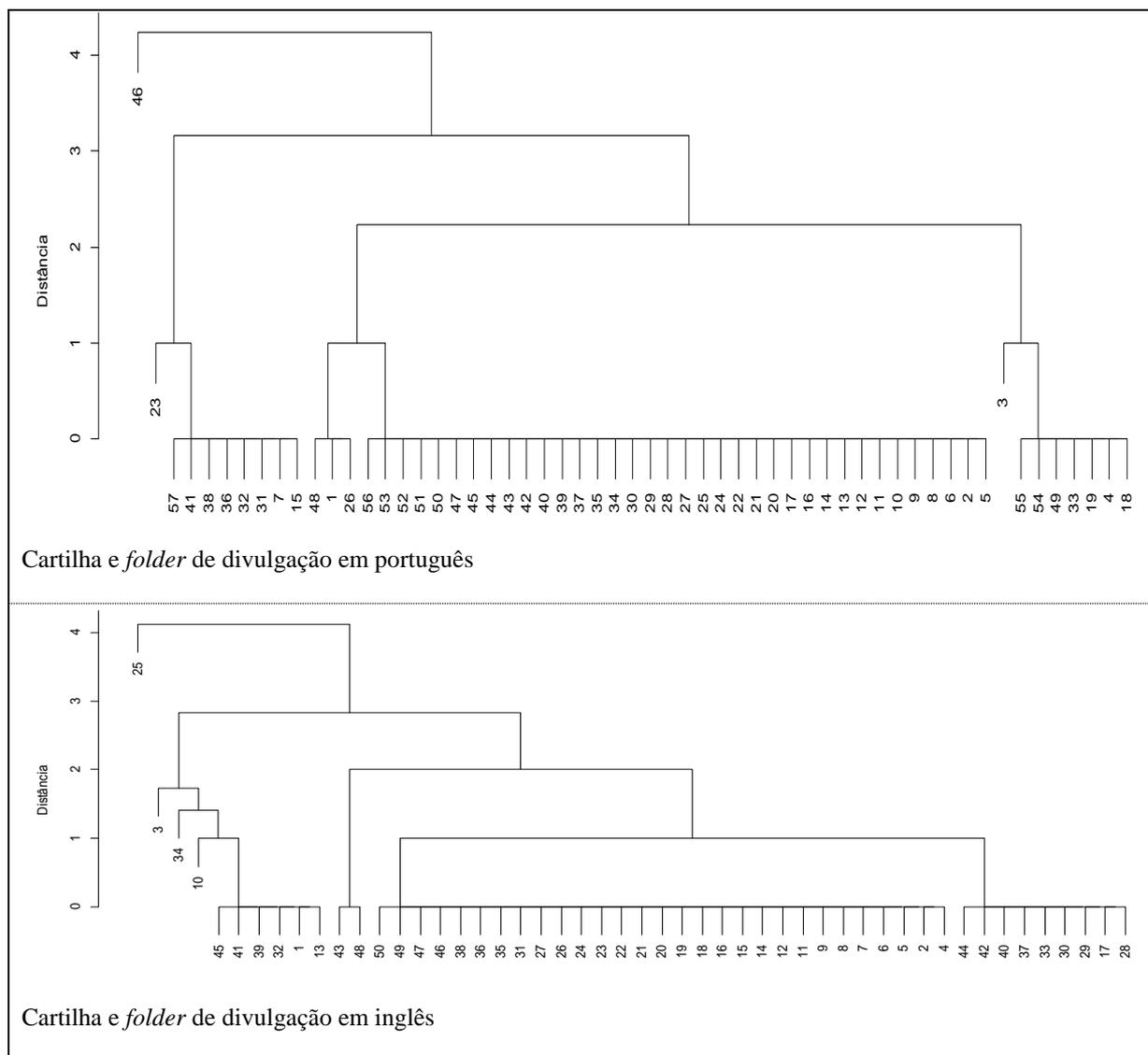


FIGURA 60. Dendrograma das cartilhas e *folders* de divulgação em inglês e em português

A FIG. 60 evidencia que a disposição gráfica das orações guarda alguma semelhança em relação ao número de ramificações. Esses dois textos são os que apresentaram dendrogramas mais parecidos. O texto em português apresenta 57 orações e o texto em inglês apresenta 50 orações. No texto em português, há três orações que se distanciam em comparação às demais e também entre si em tanto em relação à transitividade como em relação ao tema, quais sejam: 46 ( $X=1, Y=5, Z=1$ ), 23 ( $X=1, Y=2, Z=4$ ) e 3 ( $X=1, Y=1, Z=3$ ). Já no texto em inglês, há quatro orações que se distanciam, a saber: 25 ( $X=1, Y=2, Z=5$ ), 3 ( $X=2, Y=1, Z=2$ ), 34 ( $X=1, Y=1, Z=3$ ) e 10 ( $X=1, Y=2, Z=2$ ). Vale sublinhar que a oração 3 é a única em que ocorre Modo imperativo ( $X=2$ ), que corresponde a demanda de bens de serviço.

De modo geral, o agrupamento dos textos conforme ilustrado nas FIG. 58 a 60 é alterado de acordo com os sistemas de TRANSITIVIDADE, mais especificamente por meio dos

diferentes tipos de processos encontrados em cada um dos tipos de textos e por meio do sistema de TEMA, que aponta a distribuição dos significados ideacionais e interpessoais nos textos, permitindo a categorização dos diferentes tipos de textos.

Vale lembrar que durante as extrações de dados gerais dos *corpora*, bem como a extração dos dados dos perfis metafuncionais dos textos desta pesquisa, observou-se que aqueles classificados como pertencentes aos processos sociossemióticos EXPLORAR e HABILITAR guardavam características comuns, o que dificultava a caracterização de cada um deles. No intuito de se tornar as escolhas ideacionais, interpessoais e textuais presentes na cartilha mais acessíveis ao público leigo, foi elaborado um texto com base no *template* criado para a elaboração de cartilha a partir das orações de um texto de popularização da ciência (cf. BRAGA, 2013), conforme ilustrado na Subseção 3.3.4, a seguir.

### **3.3.4 Elaboração do texto do tipo HABILITAR a partir do template (modelo)**

No caso desta pesquisa, conforme mencionado anteriormente, foi criado um *template* que sugeriu a construção de um texto consciente do tipo EXPLICAR, no intuito de propor uma solução para a necessidade de se entender o posicionamento do autor em relação ao leitor para a construção de significados quando da elaboração de uma cartilha. Vale ressaltar que a escolha por um texto de popularização da ciência como base para a elaboração do *template* ocorreu por este tipo de texto estar associado aos processos sociossemióticos EXPLICAR e HABILITAR, considerou-se, portanto, que tais processos seriam uma base para a elaboração da cartilha. Apesar de a cartilha estar mais claramente associada ao processo HABILITAR, também guarda analogia com textos do processo sociossemiótico EXPLICAR. Além disso, parece não haver, em português, cartilha e *folder* de divulgação específica para a anemia falciforme (considerando-se o domínio experiencial da variável campo: medicina – triagem neonatal para anemia falciforme), entendendo-se, por conseguinte, que este é um novo registro no contexto de cultura, não havendo, portanto, um *corpus* paralelo do qual possam se extrair dados passíveis de serem contrastados. Consequentemente, nesta pesquisa, foi proposta a criação do *template* (modelo) para a elaboração da cartilha tendo como base o perfil metafuncional de um texto de popularização da ciência. Contudo, há partes em um texto de popularização que não guardam as mesmas características de uma cartilha. Mais especificamente, em um texto de popularização da ciência há uma parte em que uma

autoridade científica é citada, podendo ser um pesquisador ou até mesmo um artigo. Esta parte no *template* está ilustrada como “Fato Científico” e ela não ocorre nas cartilhas.

Para a criação do texto sobre a anemia falciforme, seguiram-se passo a passo as orientações do *template*, considerando-se as orações identificadas e classificadas com base nos sistemas de TRANSITIVIDADE, MODO, e TEMA e MENSAGEM provenientes de um texto de popularização da ciência. A partir do *template*, foi possível redigir o texto a seguir:

**Pergunta inicial: O que é anemia falciforme?**

1 oração Indicativo:Interrogativo:Elemental/Processo Material/Inicial – esta oração deve fazer uma pergunta que apresente a curiosidade sobre o fato científico do qual o texto trata.

**Características gerais do fato científico:** A anemia falciforme é uma doença do sangue. O sangue da pessoa com essa doença apresenta alteração nas células vermelhas, denominadas hemácias. As hemácias são transportadas pelo sangue e têm um formato arredondado.

**Características gerais do fato científico:**

1 oração Indicativo:Declarativo/Processo Relacional Identificativo/Continuidade – esta oração deve ter como Token o conceito principal do fato científico e como Value sua explicação.

1 oração Indicativo:Declarativo/Processo Relacional Atributivo/Continuidade – esta oração deve apresentar uma particularidade do conceito científico que será demonstrada de forma concreta na oração seguinte.

1 oração Indicativo:Declarativo/Processo Material/Inicial – esta oração deve exemplificar um fenômeno do mundo das experiências relacionado ao fato científico.

1 oração Indicativo:Declarativo/Processo Relacional Atributivo/Descontinuidade:Mudança – esta oração deve enfatizar uma característica decorrente do fenômeno do mundo das experiências apresentado na oração anterior.

1 oração Indicativo:Declarativo/Processo Existencial/Continuidade – esta oração deve apresentar uma lista de participantes que ajudem a exemplificar a oração anterior.

**Explicação sobre o fato científico:** Quando uma pessoa tem anemia falciforme, as hemácias, em vez de redondas são em formas de meia-lua ou de foice. Existem consequências da anemia falciforme, como, por exemplo:

- Crise de dor
- Acúmulo de sangue no baço (barriga)
- Cor amarela nos olhos (icterícia)

A crise de dor acontece porque as hemácias bloqueiam a circulação do sangue. A dor é um dos principais sintomas da anemia falciforme e deve ser controlada. Essa dor pode ser controlada por meio de analgésicos, dependendo de sua intensidade. Quando há dor forte, a pessoa é encaminhada para o posto de saúde mais próximo de sua residência. É importante levar a pessoa para o posto de saúde. A dor varia de acordo com a idade, como pode ser visto na frase abaixo:

1 oração Indicativo:Declarativo/Processo Material/Inicial – esta oração deve apresentar um fenômeno do mundo das experiências relacionado ao fato científico.

1 oração Indicativo:Declarativo/Processo Relacional Identificativo/Descontinuidade:Desvio – esta oração deve focalizar o texto em um participante em particular da oração e destacar sua representatividade dentro do fenômeno apresentado.

1 oração Indicativo:Declarativo/Processo Relacional Atributivo/Continuidade – esta oração deve continuar a caracterizar o participante da oração anterior.

1 oração Indicativo:Declarativo/Processo Material/Descontinuidade:Mudança – esta oração deve focalizar o texto em um evento envolvendo o participante da oração anterior.

1 oração Indicativo:Declarativo/Processo Relacional Atributivo/Continuidade – esta oração deve seguir a oração anterior, completando seu sentido com uma explicação sobre o evento apresentado.

1 oração Indicativo:Declarativo/Processo Relacional Atributivo/Continuidade – esta oração deve seguir a oração anterior, completando-a com mais uma explicação sobre o evento apresentado.

1 oração Indicativo:Declarativo/Processo Relacional Atributivo/Descontinuidade:Mudança – esta oração deve focalizar o texto em uma citação de um estudo de um profissional da área que exemplifique o fenômeno estudado.

A configuração das demais orações dependerá de como está configurada a citação escolhida.

**Fato científico**

“Os bebês têm mais infecções e dores com inchaço nas mãos e nos pés. Nas crianças maiores, as dores estão mais localizadas nas pernas, nos braços e na barriga.”

Observe a pessoa doente. Se ela tiver febre, dê um remédio para baixar a temperatura. Se a temperatura não baixar, pode ser que a pessoa esteja com alguma infecção. Ligue para o Centro de Educação e Apoio para Hemoglobinopatias – CEHMOB-MG (CEHMOB atende, 0800 7226500) para mais informações. Se não for possível ligar, procure orientação médica no posto de saúde mais próximo de sua casa.

1 oração Imperativo/Processo Mental/Inicial – esta oração deve introduzir uma série de testes que podem ser feitos pelo leitor do texto para que comprove as informações apresentadas anteriormente.

1 oração Imperativo/Processo Material/Continuidade – esta oração deve conter um teste que pode ser feito pelo leitor do texto para comprovar as informações apresentadas anteriormente.

1 oração Indicativo:Declarativo/Processo Relacional Atributivo – esta oração deve fornecer uma breve explicação sobre o resultado do teste da oração anterior.

1 oração Imperativo/Processo Material/Continuidade – esta oração deve conter um teste que pode ser feito pelo leitor do texto para comprovar as informações apresentadas anteriormente.

1 oração Indicativo:Declarativo/Processo Relacional Atributivo – esta oração deve fornecer uma breve explicação sobre o resultado do teste da oração anterior.

1 oração Imperativo/Processo Material/Continuidade – esta oração deve conter um teste que pode ser feito pelo leitor do texto para comprovar as informações apresentadas anteriormente.

1 oração Indicativo:Declarativo/Processo Relacional Atributivo – esta oração deve fornecer uma breve explicação sobre o resultado do teste da oração anterior.

Como pode ser observado, esse texto do tipo HABILITAR foi gerado a partir de escolhas metafuncionais com base na gramática sistêmico-funcional hallidayana, no intuito de criar um modelo de produção textual – ou textualização, conforme sugere Coulthard (1992) – que possa ser mais bem identificado e distanciado dos manuais técnicos. A partir das orações de textos do tipo HABILITAR foi possível se chegar à produção de um texto mais próximo daquele considerado como texto de popularização da ciência com base na descrição do perfil metafuncional no ambiente multilíngue (FIGUEREDO, 2011; BRAGA, 2013).

Conforme mencionado anteriormente, as cartilhas, em geral, na área da Saúde no país, apesar de serem direcionadas ao público leigo e pertencerem ao processo sociossemiótico HABILITAR, não são muito acessíveis a esse público, uma vez que apresentam uma linguagem mais densa e menos acessível quando comparada aos textos caracterizados como textos de popularização da ciência (pertencente ao processo sociossemiótico EXPLICAR) em outras áreas, aproximando-se mais dos tipos de texto HABILITAR. Nesse sentido, a análise linguística deste estudo mostrou ser possível, por meio do *template* de base sistêmico-funcional, elaborar um texto de sobre a anemia falciforme que pode se tornar acessível ao público leigo na área da Saúde. Contudo é necessário testar a acessibilidade deste *template* a partir da veiculação de uma cartilha elaborada por meio dele.

Cumprе ressaltar que a linguagem nesse tipo de texto HABILITAR é estruturada para se entender a mensagem a partir dos significados ideacionais (a partir do mundo das experiências), interpessoais (o papel de cada um dos participantes, proximidade entre autor-leitor, relações de poder) e textuais (escrita) realizados no nível léxico-gramatical pelos sistemas de TRANSITIVIDADE, MODO e TEMA. Todos esses significados atuam simultaneamente, considerando-se o impacto das diferentes variáveis do contexto de situação. Essa estrutura não foi observada durante as análises da cartilha e *folder* de divulgação, uma vez que os dados apontaram a proximidade deste tipo de texto com aquele analisado no tipo de texto HABILITAR. Do ponto de vista da variável campo, está explícita no texto gerado a partir do *template* a natureza da atividade social (HALLIDAY; HASAN, 198), na qual os significados experienciais sobre a anemia falciforme foram evidenciados em uma situação com vistas a elaborar elaboração de teoria para a construção do conhecimento do público leigo sobre o que é a anemia falciforme. No que toca à variável sintonia, fica clara a distância social entre o autor/leitor, que é observada a partir da interação especialista, detentor do conhecimento sobre a anemia falciforme e o público leigo, que precisa ser informado acerca

desse assunto. Por fim, partindo da variável modo vê-se que ela foi a responsável por organizar texto, a partir do papel construtivo da língua na interação entre os participantes em função do contexto de situação. Nesse sentido, o canal do texto com base no *template* é o gráfico, o meio é o escrito, o modo de interação é o monólogo e o modo retórico é a exposição. Em suma, a proposta de *template* buscou mostrar como a descrição do perfil metafuncional com base na LSF pode auxiliar na construção de textos do tipo HABILITAR no ambiente multilíngue.

No capítulo seguinte, são discutidos os resultados ora apresentados.

# **CAPÍTULO 4**

## **DISCUSSÃO DOS RESULTADOS**

## 4 DISCUSSÃO DOS RESULTADOS

Neste capítulo, são retomados os resultados obtidos apresentados nos capítulos anteriores. Em seguida, esses resultados são discutidos à luz da literatura, considerando-se os objetivos gerais e específicos delineados para esta tese, e, por fim, responde-se às perguntas que conduziram a pesquisa realizada.

### 4.1 Revisão dos Achados

A Introdução e o Capítulo 1 constituem a fundamentação desta tese. A introdução explicitou:

1. A inserção desta pesquisa no escopo de pesquisas com vistas a contribuir para a elaboração de um modelo que possibilite a produção e tradução de textos no ambiente multilíngue, estando este estudo pautado na interface entre a Linguística de Corpus (MCENERY; XIAO, 2007; GRANGER, 2003) e os Estudos da Tradução, com o aporte da Linguística Sistêmico-Funcional (HALLIDAY; MATTHIESSEN, 2004; MATTHIESSEN; TERUYA; WU, 2008; FIGUEREDO, 2011);
2. A localização desta pesquisa nos estudos puros, teóricos, parciais restritos ao tipo de texto de acordo com a proposta programática de Holmes (1972) para os Estudos da Tradução. Seguindo essa proposta, a pesquisa também pode contribuir para os estudos aplicados, com enfoque no uso de ferramentas de auxílio à tradução e aplicação da tecnologia da informação; e
3. O caráter interdisciplinar e interinstitucional da pesquisa em tela, em função da sua inserção no âmbito de um projeto de cooperação entre o LETRA e o NUPAD, considerando-se a necessidade de orientar a produção textual e a tradução de diferentes tipos de textos, dada a crescente internacionalização do Núcleo ao firmar parcerias com os EUA e diversos países da África e da Europa.

No Capítulo 1, foram apresentadas as principais características e perspectivas da Linguística de Corpus (MCENERY; XIAO, 2007; GRANGER, 2008), considerando-se a busca por padrões de uso linguístico (SINCLAIR, 1991; BERBER SARDINHA, 2004) em textos sobre triagem neonatal para anemia falciforme compilados em *corpora* comparáveis classificados

de acordo com três tipos de texto específicos: (i) artigos acadêmicos (interação especialista – especialista); (ii) manuais técnicos (interação especialista – técnico da área da saúde); e (iii) cartilhas e *folders* de divulgação (interação especialista – leigo). Em seguida, deu-se ênfase à interface entre a Linguística de Corpus e os Estudos da Tradução (MUNDAY, 2001), com o aporte teórico-metodológico da Linguística Sistêmico-Funcional (EGGINS, 2004; HALLIDAY; MATTHIESSEN, 2004; FIGUEREDO, 2011). Essa interface contribuiu para a relação entre os Estudos da Tradução e os estudos linguísticos sistêmico-funcionais, possibilitando tanto a descrição quanto a explicação da língua em uso nos contextos sociais, além de sua proposta de incorporação da dimensão semântica, mais especificamente no ambiente multilíngue.

Após essa fundamentação, apresentaram-se, no Capítulo 2, a metodologia da pesquisa, iniciando-se com os procedimentos de seleção dos textos que compuseram os *corpora* comparáveis (MCENERY; XIAO, 2007; GRANGER, 2008) e a metodologia de coleta de dados por meio da utilização do R e do UAM CorpusTool, a partir dos quais foi possível extrair os dados passíveis de análise semiautomática, além de da utilização do etiquetador morfossintático Treetagger, que possibilitou anotar os *types* em classes de palavras. Em seguida, explicou-se o desenho dos *corpora* comparáveis (BIBER, 1990) e a proposta de classificação dos tipos de textos pertencentes aos processos sociossemióticos EXPLORAR e HABILITAR (MATTHIESSEN *et. al.*, 2008) para a categorização dos registros baseada na pesquisa de Eggins (2004). Por fim, apresentaram-se os procedimentos para a descrição metafuncional dos textos produzidos no ambiente multilíngue (FIGUEREDO, 2011). Portanto, neste capítulo, foram apresentadas os procedimentos de coleta dos dados (provenientes dos três diferentes tipos de textos) e os procedimentos de análise desses dados, considerando-se os procedimentos de análise com base na Linguística de Corpus com o aporte da Linguística Sistêmico-Funcional.

No intuito de identificar as especificidades de cada tipo de texto que são passíveis de análises semiautomáticas, foi realizada, no Capítulo 3, uma análise dos dados extraídos a partir do R e do UAM CorpusTool® e anotados pelo etiquetador morfossintático Treetagger. Foram analisados, em primeiro lugar, os dados que caracterizaram o desenho dos *corpora* no que toca à dimensão (relação *type/token*) e ao etiquetamento em classes de palavras por tipo de texto e por língua. A esses dados foram aplicados três testes estatísticos, quais sejam: teste qui-quadrado ( $\chi^2$ ), teste exato de Fisher e teste Z de comparação de proporção. Terminadas as

apresentações dos resultados das análises dos dados quantitativos obtidos por meio das ferramentas computacionais e com o aporte da Linguística de Corpus, foram analisados os dados à luz da Linguística Sistêmico-Funcional. A análise com base na LSF retratou a interpretação dos diferentes tipos de texto com base nas variáveis do contexto de situação (*i.e.*, campo, sintonia e modo) e seus respectivos parâmetros considerando os processos sociosemióticos EXPLORAR e HABILITAR (*i.e.*, artigo acadêmico; manual técnico, cartilhas e *folders* de divulgação) e enfocando a língua dentro do contexto de cultura (EGGINS, 2004). Essa análise resultou em indícios que apontam características específicas para cada um dos tipos de texto na busca por padrões de uso linguístico. Esses padrões foram identificados por meio dos sistemas responsáveis pela realização léxico-gramatical de significados vinculados às metafunções ideacional, interpessoal e textual. Por fim, foram apresentados os dados referentes às frequências e distribuição metafuncional em cada um dos tipos de textos, com o objetivo de propor uma descrição sistêmico-funcional (FIGUEREDO, 2011) dos sistemas de TRANSITIVIDADE (relacionado à variável campo, metafunção ideacional), MODO (relacionado à variável sintonia, metafunção interpessoal) e TEMA e MENSAGEM (relacionados à variável modo, metafunção textual). A descrição metafuncional facilitou o entendimento do processo de construção do conhecimento, por meio de artigos científicos; explicitou o comportamento dos participantes envolvidos na situação, por meio de manuais técnicos; apontou a forma de transmissão de conhecimento de especialistas para leigos que deveria ser realizada por intermédio de cartilhas ou *folders* de divulgação. Contudo, as análises dos dados extraídos dos *corpora* mostrou que há semelhanças importantes entre o manual técnico e a cartilha e *folder* de divulgação. Essas semelhanças culminaram na criação do *template* (modelo), com base na descrição do perfil metafuncional dos textos de popularização da ciência.

Foi possível, portanto, apresentar os pressupostos teóricos básicos da Linguística Sistêmico-Funcional e aplicá-los aos Estudos da Tradução baseados em *corpora*, considerando-se a investigação linguística em duas vertentes teóricas complementares. Uma é baseada em *corpus*, uma vez que, neste estudo, utilizaram-se conceitos e categorias já existentes para a escolha das variáveis de estudo e para chegar a dimensões da linguagem verbal e a tipos de textos através dos dados empíricos disponíveis. A outra é conduzida pelo *corpus*, dado que, para a análise do perfil metafuncional dos tipos de textos, foi necessário construir um modelo de sistematização das escolhas e apresentar padrões e distinções entre os dados constituintes dos *corpora* para o embasamento de formulações de generalizações linguísticas.

A Seção 4.2 ilustra a revisão desses achados à luz da literatura.

## **4.2 Revisão dos Achados à Luz da Literatura**

Após serem retomados os capítulos anteriores, a próxima seção apresenta a discussão dos resultados divididos em quatro blocos, a saber: (i) interface entre as diferentes áreas do conhecimento; (ii) análise de *corpora* comparáveis; (iii) caracterização dos textos sob a perspectiva da Linguística Sistêmico-Funcional; e (iv) descrição do perfil metafuncional no ambiente multilíngue.

### **4.2.1 Interface entre as diferentes áreas do conhecimento**

Esta tese foi realizada com base na interface entre os Estudos da Tradução baseados em *corpora* (MCENERY; XIAO, 2007; GRANGER, 2003), com o aporte da Linguística Sistêmico-Funcional (HALLIDAY; MATTHIESSEN, 2004; EGGINS, 2004; MATTHIESSEN; TERUYA; WU, 2008; FIGUEREDO, 2011), além de contar com o apoio de outras áreas do conhecimento, como a Linguística Computacional e a Estatística.

Por meio da Linguística de Corpus, campo de investigação empírica que estuda a língua em uso, foi possível, por meio da compilação e exploração dos *corpora* (BERBER SARDINHA, 2004), adotar uma abordagem metodológica que garantisse a exploração da linguagem verbal a partir de exemplos reais. Em outras palavras, foi possível realizar um estudo empírico dos *corpora* comparáveis (KENNING, 2010) desta pesquisa. A Linguística Computacional foi essencial para auxiliar no processo de extração e manipulação dos dados dos *corpora* (MCENERY; XIAO, 2007; GRANGER, 2003), por meio de métodos empíricos e técnicas estatísticas, conforme proposta de Biber (1988, 1995; BIBER; CONRAD; REPPEN, 1998; BIBER *et al.*, 1999), retomada por Neumann (2008) e adotada nesta pesquisa. Seguiu-se a vertente baseada em *corpus*, que utiliza conceitos e categorias já existentes para a escolha das variáveis de estudo e para chegar a dimensões da linguagem verbal e a tipos de textos por meio dos dados disponíveis.

Os dados extraídos dos *corpora* por meio do R e anotados pelo Treetagger receberam tratamento estatístico. Foram aplicados três testes, quais sejam:

1. Teste qui-quadrado ( $\chi^2$ ), para testar a significância dos dados obtidos em relação a cada um dos tipos de textos em português e em inglês e verificar se houve associação significativa dos nódulos (itens lexicais e itens gramaticais) analisados com a língua e os tipos de textos;
2. Teste exato de Fisher, também para testar a significância dos dados, em complemento ao teste  $\chi^2$ ; e
3. Teste Z de comparação de proporção (porcentagem), para identificar para quais casos a maior ocorrência de uma classe de palavras está significativamente associada com a língua em um dado tipo de texto.

Os resultados obtidos com a aplicação desses testes possibilitaram identificar quais itens lexicais e itens gramaticais – usados como nódulos – e classes de palavras que se diferenciam significativamente e podem ser mais bem investigados para a caracterização do registro ao qual estão vinculados.

Tanto a Linguística de Corpus quanto a Linguística Sistêmico-Funcional são abordagens que se complementam, enfatizam a língua em uso e apresentam características probabilísticas. Portanto, o diálogo entre a LC e LSF pode ser considerado proveniente da proposta de a análise dos textos originais e traduzidos ser realizada por meio de questões linguísticas, incluindo-se o contexto de produção do significado no processo de análise tanto de textos originais quanto de suas traduções (HALLIDAY; McINTOSH; STREVENS, 1964; CATFORD, 1965; MATTHIESSEN, 2001; PAGANO; VASCONCELLOS, 2005; MUNDAY, 2001). Portanto, esta pesquisa se pautou no estudo de Neumann (2008) no que toca à complementaridade entre: (i) a Linguística de Corpus, empregada para a descrição linguística por meio de métodos empíricos e técnicas estatísticas, e (ii) a Linguística Sistêmico-Funcional, por meio de análises das variações do registro. Além disso, contou com a contribuição que a LC oferece aos Estudos da Tradução por meio de suas ferramentas metodológicas, com o aporte teórico-metodológico da teoria sistêmico-funcional.

#### **4.2.2 Análise de corpora comparáveis**

Os *corpora* compilados para a realização desta pesquisa seguiram categorização dos *corpora* adotada por McEnery e Xiao (2007) e a tipologia proposta por Granger (2003): comparável em duas línguas (inglês e português). A partir desses *corpora* foi possível investigar padrões

de uso linguístico recorrentes em textos na área da Saúde (mais especificamente, da medicina – subárea triagem neonatal para anemia falciforme) em três tipos de texto com vistas à caracterização desses tipos de textos sob a perspectiva dos parâmetros das variáveis do contexto de situação (*i.e.*, campo, sintonia e modo). A identificação de padrões de uso linguístico a partir de um *corpus* comparável permitiu caracterizar os textos do tipo EXPLORAR e HABILITAR em inglês e em português, seguindo-se a proposta da tipologia de textos baseadas nos parâmetros do contexto (MATTHIESSEN; TERUYA; WU, 2008). Contudo, observou-se que o manual técnico e a cartilha e *folder* de divulgação guardam semelhanças entre si em português, o que dificultou a distinção e posterior classificação desses dois tipos de texto no que toca aos dados gerais dos *corpora*. Esse fato apontou para a necessidade de uma análise mais detalhada, com o aporte da Linguística Sistêmico-Funcional, que possibilitasse uma distinção mais clara entre o manual técnico (interação especialista – técnico da área da saúde) e a cartilha e *folder* de divulgação (interação especialista – público leigo) e contribuísse para a categorização dos artigos acadêmicos, manual técnico e cartilha e *folder* de divulgação.

A análise da variação lexical e da densidade lexical, baseadas respectivamente nas relações *type/token* e substantivo/verbo, mostrou uma diferença clara entre os textos na língua inglesa: há mais variação lexical nos artigos acadêmicos e manuais técnicos que nas cartilhas e *folders* de divulgação e há mais densidade lexical nos artigos acadêmicos, seguido dos manuais técnicos e das cartilhas. Em português, contudo, a análise revela uma aproximação entre os o manual técnico e a cartilha e *folder* de divulgação, uma vez que ambas as relações ora analisadas estão bastante próximas. Essas relações podem ser observadas a partir de várias partes dos dois textos. A título de ilustração, os exemplos (1) e (2) a seguir retirados dos excertos do manual técnico e da cartilha e *folder* de divulgação, respectivamente, comprovam a aproximação entre esses dois tipos de texto:

- (1) A infecção deve sempre ser investigada e acompanhada com muito zelo pela equipe de saúde, pois ela é responsável pela primeira causa de mortalidade nesta doença.
- (2) Os episódios de febre devem, portanto, ser encarados como situações de risco, nas quais os procedimentos diagnósticos devem ser aprofundados e a terapia deve ser imediata.

Essas semelhanças apontaram para a criação do *template* (modelo), com base na descrição do perfil metafuncional dos textos de popularização da ciência (BRAGA, 2013), entendendo-se que a cartilha para a anemia falciforme é um tipo textual novo no contexto de cultura e que

guarda características análogas dos textos pertencentes aos processos EXPLICAR e HABILITAR. Esse *template* foi desenvolvido no intuito de propor uma solução para a necessidade de se entender o posicionamento do autor em relação ao leitor para a construção de significados quando da elaboração de uma cartilha.

As diferenças entre os *corpora* em relação à frequência das palavras tanto em uma análise entre tipos de textos como em uma análise entre línguas remeteu à forma como o domínio experiencial foi estabelecido pelo autor do texto. Esses indícios ensejam o estudo das variações dos registros utilizados para construir significados sobre a anemia falciforme. Conforme aponta Neumann (2008), a maior ou menor variação lexical podem estar atreladas à natureza do tipo de registro. No caso das cartilhas e *folders* de divulgação em inglês, por exemplo, é possível que seja utilizada uma linguagem mais padronizada a fim de facilitar a leitura dos textos e evitar mal-entendidos em relação às informações que estão sendo divulgadas sobre a anemia falciforme – mal-entendidos esses que podem emergir em virtude da distância social entre o autor e o leitor de textos desse tipo de texto. Já para as cartilhas e *folders* em português, não se observa a utilização de uma linguagem mais direcionada ao público leigo, pois há características bem próximas às encontradas no manual técnico. Ao analisar as nuvens de palavras, por exemplo, observa-se a ocorrência dos verbos “devem” e “podem”, que ocorrem com frequência tanto nos manuais técnicos quanto nas cartilhas, como foi ilustrado também nos exemplos (1) e (2) acima.

Além disso, conforme apontado por Halliday e Matthiessen (2004), a forma condensada como a informação nos textos científicos é construída, devido à qualidade sinóptica dos tipos de textos, remete a uma maior densidade lexical, como foi o caso especificamente no artigo acadêmico em português e em inglês neste estudo. Parece haver uma compactação da informação dos textos escritos de modo que os pares ou pessoas envolvidas no âmbito acadêmico sejam capazes de entender mais facilmente a linguagem expressa por meio da condensação de significados. Esse resultado pode ser interpretado com base em Halliday (1994, p. 57), que explica que o aumento da densidade lexical se dá não devido à elevada ocorrência de itens lexicais, mas sim pela diminuição da frequência de ocorrência dos verbos. Nesse processo, os verbos se nominalizam, o que faz com que o número de orações diminua por complexo oracional.

Vislumbrando-se a perspectiva de Neumann (2008) sobre a variabilidade de registro, em particular, na utilização da Linguística de Corpus, para a descrição linguística por meio de métodos empíricos e técnicas estatísticas, observou-se nos *corpora* a frequência de ocorrência das diferentes classes de palavras, com a finalidade de identificar padrões mais recorrentes em cada um dos tipos de textos. Para tanto foi realizado o etiquetamento dos itens lexicais e itens gramaticais, com o aporte do programa Treetagger. Esse etiquetamento possibilitou mapear a forma como os itens lexicais e itens gramaticais de cada um dos tipos de textos desta pesquisa foram distribuídos de acordo com as classes de palavras, tornando possível observar a relação significativa entre a frequência de ocorrência das classes de palavras e a língua. Foi esse etiquetamento inclusive que permitiu comparar a relação substantivo/verbo, mencionada anteriormente, e inferir que há mais nominalizações no artigo acadêmico em função de essa relação ser maior. Cabe aqui observar que substantivos, por si sós, não são nominalizações, mas uma alta ocorrência dessa classe em relação ao número de verbos aponta indícios de existência de nominalizações.

Após analisadas as frequências de ocorrências de cada uma das classes de palavras, foram selecionados os nódulos “doença falciforme”, “anemia falciforme”, “*sickle cell disease*” e “*sickle cell anemia/anaemia*”, selecionadas por serem consideradas palavras-chave da área da Saúde em português e em inglês. Em seguida, disponibilizaram-se os dados com a análise dos nódulos com as palavras “dor”, “*pain*”, “paciente”, “*patient*”, “*you*”, “você”, selecionados devido à frequência de ocorrência nos textos. Análise apontou, dentre outros achados, que não há uma relação próxima entre o leitor e autor nas cartilhas e *folders* de divulgação em português, haja vista que não foram encontradas ocorrências de “você”, sendo que em inglês há ocorrências de “*you*”, correspondente formal de “você”. Esse foi também mais um fator apontando para a necessidade de um *template* para esse texto em português.

Considerando-se a interface entre os Estudos da Tradução e a Linguística Comparada nas últimas décadas, conforme aponta Granger (2003), observou-se nesse estudo a importância do *corpus* eletrônico. A autora explica que a utilização de *corpus* aproxima os pesquisadores da Linguística de Corpus e dos Estudos da Tradução ao realizarem pesquisas com os mesmos tipos de dados, utilizando os mesmos programas concordanceadores, contudo com objetivos diferentes. Nota-se, sobremaneira, que a pesquisa baseada em *corpus* tem limitações para os Estudos da Tradução e faz-necessária a manipulação (automática, semiautomática ou manual) dos *corpora* dependendo do objetivo específico de cada pesquisa. No caso desta pesquisa, foi

necessário o aporte da Linguística Sistêmico-Funcional (EGGINS, 2004) uma vez que foram analisados mais detidamente os itens lexicais (substantivo, adjetivo, advérbio e verbo) e os itens gramaticais (conjunção e pronomes) por mostrarem indícios que possam ser utilizados para a categorização dos registros.

Além disso, com o objetivo de consubstanciar as teorias e os dados extraídos dos *corpora* em bases quantitativas/estatísticas (BAKER, 1995) foram realizados alguns testes estatísticos. O teste do  $\chi^2$  foi realizado para investigar se a frequência total dos itens lexicais (substantivos, os adjetivos, os advérbios e os verbos) e dos itens gramaticais (pronomes e conjunções) observada em cada tipo de texto e por língua se desviou ou não da frequência esperada em relação ao número total de *tokens*. Já para se identificar qual classe de palavra está associada com qual língua (em qual das duas línguas a classe de palavra ocorre em maior porcentagem) utilizou-se o teste Z de comparação de proporção (porcentagem), considerando-se um nível de significância de 0,05. No caso dos nódulos “anemia falciforme” e “doença falciforme” analisados em português, o p-valor foi menor que 0,01, rejeitando a hipótese nula. Concluiu-se que, para os tipos de texto em português, houve associação estatisticamente significativa dos nódulos “doença falciforme” e “anemia falciforme” com o tipo de texto. Já para os nódulos “*sickle cell disease*” e “*sickle cell anemia/anaemia*”, essa associação não foi significativa. Tal análise pode apontar para a utilização indiscriminada desses nódulos referentes à construção do significado sobre a doença, conforme afirmam Serjeant, B. E Serjeant, G. (1991, p. 31). Para os demais nódulos (“dor”, “*pain*”, “paciente”, “*patient*”, “*you*”, “você”), o teste  $\chi^2$  sugeriu uma associação significativa apenas dos nódulos “dor” e “*pain*” com o tipo de texto.

#### **4.2.3 Caracterização dos textos sob a perspectiva da Linguística Sistêmico-Funcional**

As contribuições da sistêmica para esta pesquisa se concentram, aqui, em mostrar organização dos textos dos *corpora* com base na tipologia de textos (MATTHIESSEN; TERUYA; WU, 2008) e amparada nos estudos de texto, por meio das variáveis do contexto de situação e seus respectivos parâmetros (EGGINS, 2004). Partiu-se do pressuposto de que há diferença, em função do registro, entre os padrões de uso linguístico, mapeados de forma semiautomática de acordo com a metodologia de estudos de *corpora*. Esses padrões foram identificados sob a perspectiva dos parâmetros das variáveis do contexto – campo, sintonia e modo – por meio de significados vinculados às metafunções ideacional, interpessoal e textual (EGGINS, 2004).

Para verificar como se deu o processo de construção do significado dos três tipos de textos que compuseram os *corpora* desta pesquisa, considerou-se que as especificidades de cada registro são passíveis de análises semiautomáticas, por meio da extração de itens lexicais e itens gramaticais mais frequentes, selecionados como nódulos. A análise desses nódulos permitiu identificar indícios para caracterização dos textos com base nas variáveis do contexto de situação (*i.e.*, campo, sintonia e modo) e seus respectivos parâmetros dos processos sociossemióticos EXPLORAR e HABILITAR (MATTHIESSEN; TERUYA; WU, 2008). A análise mais detida de alguns desses nódulos mostrou ser possível identificar algumas características inerentes aos processos sociossemióticos nos quais os tipos de textos estão instanciados. A ocorrência de pronomes como “*you*”, por exemplo, é indicativa de trocas de mercadorias (bens e serviços e informações) que devem ser seguidas à risca no caso do processo HABILITAR. Já a ocorrência de palavras como “falciforme”, “anemia” e “doença” explica, informa e transmite o conhecimento, no caso dos três tipos de texto, ou seja, o autor, na condição de detentor do conhecimento, elabora um artigo acadêmico, ou um manual técnico ou uma cartilha e *folder* de divulgação que transmitam a informação sobre a doença de acordo com o tipo de texto, EXPLORAR e HABILITAR. Já palavras como o substantivo “dor” e “paciente” são indicativos de orientação para o leitor que não detém o conhecimento das informações, porém está interessado em realizar procedimentos de saúde, de acordo com as prescrições do autor, detentor e divulgador do conhecimento, como no caso do processo sociossemiótico HABILITAR. As escolhas ideacionais e interpessoais analisadas nesta pesquisa apontam a forma de construção do significado nos diferentes tipos de texto, em consonância com os resultados da pesquisa de Eggins (2004).

Em suma, a partir da análise dos padrões de uso linguístico, foi possível apontar as diferenças e semelhanças entre os diferentes tipos de textos, considerando-se a investigação das variáveis campo (relacionada ao tipo de ação social em uma dada situação), sintonia (relacionada ao tipo de relação entre os participantes na situação) e modo (relacionada à forma como a mensagem é realizada). Os resultados obtidos sugerem que, apesar de pertencerem ao mesmo domínio experiencial e guardarem características em comum, os tipos de textos estudados nesta pesquisa apresentam características prototípicas passíveis de serem caracterizadas inter-registro e interlândia tanto por meio de dados empíricos quanto por meio dos parâmetros das variáveis campo, sintonia e modo, de acordo com a tipologia dos processos sociossemióticos EXPLORAR e HABILITAR. Esses resultados apontam a Linguística de Corpus propicia a caracterização dos textos, pois a análise dos dados empíricos dos *corpora* apontam diferenças,

em função do registro (Neumann, 2008), em relação à variação lexical, densidade lexical, frequência de ocorrência de palavras e mapeamento da forma como as palavras foram distribuídas em classes de palavras.

#### ***4.2.4 Descrição do perfil metafuncional no ambiente multilíngue***

Os achados desta pesquisa, no que toca à diferenciação dos textos à luz da Linguística Sistêmico-Funcional, por meio da descrição do perfil metafuncional dos textos, são resultantes de uma aplicação prática das pesquisas de Halliday e Matthiessen (1999) e de Figueredo (2011). Esta pesquisa pode ser considerada como uma das primeiras mais sistemática do ponto de vista de uma aplicação prática da gramática sistêmico-funcional para o português brasileiro (FIGUEREDO, 2011). Ao mesmo tempo, os dados desta pesquisa podem ser adicionados a futuras sistematizações dessa gramática, o que pode permitir compreender melhor a produção linguística no ambiente multilíngue, além de propor modelos para descrição de perfis metafuncionais.

A compilação dos dados a partir dos sistemas de TRANSITIVIDADE, MODO, e TEMA e MENSAGEM possibilitou a organização de tabelas (Anexo 2) cujos dados foram utilizados para a criação de gráficos e dendrogramas. A partir desses gráficos e dendrogramas foi possível ilustrar graficamente o perfil metafuncional dos três diferentes tipos de textos em português e em inglês considerando-se as três metafunções (*i.e.*, ideacional, interpessoal e textual). A análise de cada um dos gráficos permitiu observar a organização interna dos textos de forma simbólica, apontando a diferença no perfil metafuncional de cada um deles. Já a análise dos dendrogramas permitiu identificar aglomerações dos dados obtidos por meio da análise das funções dos sistemas de MODO, TRANSITIVIDADE e TEMA. Tanto os gráficos quanto os dendrogramas ilustram um modelo de como a língua funciona no ambiente multilíngue.

Em suma, a descrição do perfil metafuncional de cada um dos tipos de textos possibilitou:

1. A descrição da forma como os textos realizam significados por meio das variáveis no contexto de cultura (MARTIN, 1992);
2. A construção de uma proposta de modelo de sistematização das escolhas metafuncionais dos textos no ambiente multilíngue (FIGUEREDO, 2011);

3. A caracterização dos registros por meio das escolhas metafuncionais de textos do tipo EXPLORAR e HABILITAR (MATTHIESSEN; TERUYA; WU, 2008; FIGUEREDO, 2011) em português e inglês; e
4. A sugestão de possíveis modificações em textos do tipo HABILITAR com base em pesquisas anteriores, (COULTHARD, 1992). Elaborou-se, portanto, um *template*, mostrando passo a passo a criação de um texto desde a primeira escolha de significado, considerando-se a especialização e variedade que a Linguística Sistêmico-Funcional proporciona para realizar cada micro contexto de situação específico por meio das variáveis do contexto de situação: campo, sintonia e modo. Apesar de serem necessárias intervenções de especialistas da área no texto criado, conclui-se que uma cartilha possa ser elaborada a partir de um *template* (modelo) com base no perfil metafuncional de um texto de popularização da ciência. Justifica-se a escolha por um texto de popularização da ciência considerando-se a cartilha e *folder* de divulgação específica para a anemia falciforme ser um novo registro no contexto de cultura, não havendo, portanto, um *corpus* paralelo do qual possam se extrair dados passíveis de serem comparados. Além disso, observou-se que textos da popularização da ciência são associados aos processos sociosemióticos EXPLICAR e HABILITAR. Considerou-se, por conseguinte nesta pesquisa, que tais processos seriam uma base para a elaboração da cartilha, que apesar de estar mais claramente associada ao processo HABILITAR, também guarda analogia com textos do processo sociosemiótico EXPLICAR.

Esta tese apresentou resultados dos dados provenientes dos *corpora* com base em duas abordagens metodológicas complementares, ou seja, a Linguística de Corpus e a Linguística Sistêmico-Funcional. Assim, foi possível contribuir na relação entre os Estudos da Tradução e os estudos linguísticos com base em *corpora*, com o aporte da Linguística Sistêmico-Funcional.

Terminadas as discussões sob a luz da teoria, o Capítulo 5 traz a conclusão acerca desta pesquisa.

# CONCLUSÃO

## 5 CONCLUSÃO

**E**sta tese foi desenvolvida com o intuito de promover a relação entre Linguística de Corpus e os Estudos da Tradução com subsídios da Linguística Sistêmico-Funcional.

Deu consecução aos seus objetivos gerais e específicos, nomeadamente:

1. Contribuiu para estreitar a relação entre os Estudos da Tradução, a Linguística de Corpus e os estudos linguísticos sistêmico-funcionais, possibilitando tanto a descrição quanto a explicação da língua em uso nos contextos sociais, além de sua proposta de incorporação da dimensão semântica, mais especificamente no ambiente multilíngue, para estudar como textos que instanciam os tipos de textos – artigo acadêmico, manual técnico e cartilha e *folders* de divulgação – são produzidos em português e em inglês;
2. Elaborou modelo para a tradução e produção textual no ambiente multilíngue no âmbito de Projeto de Cooperação entre o LETRA/FALE/UFMG e o NUPAD/FM/UFMG ao investigar padrões de uso linguístico (SINCLAIR, 1991; BERBER SARDINHA, 2004) em três tipos de textos com vistas à caracterização desses registros;
3. Classificou os tipos de textos como pertencentes aos processos sociossemióticos EXPLORAR e HABILITAR (MATTHIESSEN; TERUYA; WU, 2008), com base nas variáveis do contexto de situação (*i.e.*, campo, sintonia e modo) e seus respectivos parâmetros com base na tipologia desses processos sociossemióticos, enfocando a linguagem dentro do contexto de cultura (EGGINS, 2004)
4. Analisou a distribuição metafuncional (escolhas ideacionais, interpessoais e textuais) em cada um dos tipos de textos, propondo a descrição sistêmico-funcional (FIGUEREDO, 2011) dos sistemas de TRANSITIVIDADE (relacionado à variável campo, metafunção ideacional), MODO (relacionado à variável sintonia, metafunção interpessoal) e TEMA e MENSAGEM (relacionados à variável modo, metafunção textual). Essa descrição metafuncional facilitou o entendimento do processo de construção do conhecimento, por meio de artigos científicos; explicitou o comportamento dos participantes envolvidos na situação, por meio de manuais técnicos;

apontou a forma de transmissão de conhecimento de especialistas para leigos por intermédio de cartilhas ou *folders* de divulgação; e

5. Propôs a criação de um *template* (modelo de produção textual) para a elaboração de cartilha, a partir nos resultados da análise da cartilha e *folder* de divulgação, com vistas a sua futura utilização como um primeiro passo para a elaboração de um pipeline (percurso ou linha de produção textual), com base no perfil metafuncional do texto de popularização da ciência no âmbito de Projeto de Cooperação entre o LETRA/FALE/UFMG e o NUPAD/FM/UFMG. Vale lembrar que embora haja textos aos quais se atribui o rótulo de cartilha no contexto de cultura brasileiro, por serem de outros domínios experienciais que o de anemia falciforme não poderiam subsidiar a tradução de cartilhas originalmente escritas em inglês. Logo, a escolha por um texto de popularização da ciência se deu pelo fato de a cartilha específica para a anemia falciforme ser um novo registro no contexto de cultura, não havendo, portanto, um *corpus* paralelo do qual pudessem ser extraídos dados. Espera-se que esse *template* possa ser utilizado como o primeiro passo para a organização de um *pipeline* (*i.e.*, linha de produção textual), além de auxiliar tradutores e pesquisadores na produção de cartilhas e *folders* de divulgação na área da Saúde (mais especificamente, a medicina, subárea triagem neonatal para anemia falciforme) ;

Esta tese apresentou, portanto, resultados de um estudo de padrões de uso linguístico (SINCLAIR, 1991; BERBER SARDINHA, 2004) em textos da área da Saúde (mais especificamente da Medicina – subárea triagem neonatal para anemia falciforme), em *corpora* comparáveis compilados com base nas variáveis contextuais descritas pela Linguística Sistemico-Funcional, em três tipos de texto com distintas configurações explícitas de sintonia: (i) artigos acadêmicos (interação especialista – especialista) associados ao processo sociosemiótico EXPLORAR; (ii) manuais técnicos (interação especialista – técnico da área da saúde) e (iii) cartilhas e *folders* de divulgação (interação especialista – leigo), ambos associados ao processo sociosemiótico HABILITAR. Com isso foi possível descrever e contrastar padrões linguísticos nos três tipos de texto estudados com vistas à caracterização de registros orientada para a tradução e produção textual no ambiente multilíngue. O estudo validou a LSF como uma teoria abrangente, que permite (i) as análises e descrições textuais necessárias para comparar diferentes textos em diferentes línguas e (ii) o uso de recursos computacionais que tornam possível a análise de *corpora* em termos probabilísticos.

Observou-se que a replicação de metodologias com enfoque em *corpora* (MCENERY; XIAO, 2007; GRANGER, 2003), além de gerar resultados válidos e confiáveis linguística e estatisticamente, contribui para corroborar metodologias emergentes (como a análise de *corpora* comparáveis por meio do ambiente de programação R, com poder estatístico, do Treetagger, etiquetador morfossintático, e do UAM CorpusTool, para análise textual por meio de esquemas) e também ampliar o interesse da academia pela produção de ferramentas que tenham aplicação prática para profissionais de outras áreas. Portanto, como sugestão para pesquisas futuras, evidencia-se a possibilidade de caracterizar os textos também a partir das generalizações dos tipos de texto em contexto de cultura, por meio da busca e identificação de padrões de uso linguístico em *corpora* comparáveis que direcionem a tradução e elaboração de textos em inglês e em português.

No caso dos resultados apresentados na presente tese, a extração de co-ocorrência entre os itens lexicais e os itens gramaticais extraídos dos *corpora* comparáveis através do R apontaram padrões de uso linguístico indicativos de diferenças características dos tipos de textos analisados. No que toca à Linguística Sistêmico-Funcional, as diferenças e semelhanças entre os tipos de texto estudados responderam ao impacto das variáveis do contexto (*i.e.*, campo, sintonia e modo) que se realizam no nível da léxico-gramática (EGGINS, 2004). Dessa forma, foi possível localizar os tipos de texto na tipologia dos textos com base nos parâmetros do contexto (MATTHIESSEN; TERUYA; WU, 2008), aqui classificados enquanto pertencentes aos processos sociossemióticos EXPLORAR (artigo acadêmico), HABILITAR (manual técnico e cartilhas e *folders* de divulgação).

A descrição do perfil metafuncional dos textos revelou que os processos mais significativos para a construção do mundo das experiências em todos os tipos de textos em português e em inglês foram os processos relacionais e materiais. O Modo que se destacou para todos os tipos de textos em português e em inglês foi o Indicativo Declarativo. Todos os tipos de texto em português e em inglês estão organizados, em função da semântica, a partir de mensagens iniciais, mensagens de continuidade e descontinuidade (mudança e desvio) e, em função da gramática, em temas que se diferem para os textos em português e em inglês. No que toca aos tipos de temas, os textuais e o *default* foram os mais frequentes em todos os tipos de texto em português; o tema ângulo-fonte foi encontrado apenas no artigo acadêmico em português; os temas simples e múltiplos (textual e tópico) foram os mais frequentes em todos os tipos de

texto em inglês; o tema múltiplo (interpessoal e tópico) ocorreu apenas na cartilha e *folder* de divulgação em inglês.

Em suma, a identificação de padrões a partir de um *corpus* comparável e a descrição dos perfis metafuncionais dos tipos de texto permitem caracterizar os três tipos de texto estudados nesta pesquisa e são o ponto de partida para se fomentar a construção de um *pipeline*. Pondera-se, neste estudo, as especificidades de cada registro que são passíveis de análises semiautomáticas e podem subsidiar a elaboração desse *pipeline* com base na interface entre os Estudos da Tradução e a Linguística de Corpus, com o aporte da Linguística Sistêmico-Funcional. Para a construção do *pipeline* são necessárias três etapas. A primeira etapa consiste na análise gramatical que leva à descrição do perfil metafuncional de cada tipo de texto analisado, podendo gerar *templates* que sirvam como modelos para a elaboração de diferentes tipos de textos na área da Saúde, por exemplo. Esses *templates* devem estar inseridos no *pipeline* para a consecução das etapas subsequentes. A segunda etapa refere-se a um *pipeline* semântico (significado da língua de partida na língua de chegada, por exemplo) e a última etapa consiste na seleção de um dicionário de terminologia. As duas últimas etapas são consideradas complementares à primeira.

Como sugestões de pesquisas futuras, os resultados obtidos neste estudo poderão ser examinados com vistas à:

1. Descrição metafuncional de novos tipos de textos, com vistas à produção textual e tradução no ambiente multilíngue, para verificar se corrobora os achados desta pesquisa;
2. Automatização dos processos de análise dos sistemas de TRANSITIVIDADE, MODO, e TEMA e MENSAGEM, no intuito de se criar uma distribuição de dados numéricos contínuos (e não apenas categóricos), considerando-se o nível de delicadeza de cada um desses sistemas da léxico-gramática. Dessa forma será possível propor uma representação geométrica que possibilitará a categorização de diversos tipos de textos;
3. Extração de termos técnicos e fraseologias especializadas (BEVILACQUA, 2004) KRIEGER; BEVILACQUA, 2005) em cada um dos tipos de texto com o aporte da Linguística Computacional para fins de mineração de dados dos *corpora*. Esses termos técnicos e fraseologias poderão ser utilizados para compor o banco de dados

que subsidiária o desenvolvimento do *pipeline* previsto como desdobramento desta pesquisa. Esse *pipeline* poderá servir para direcionar a tradução e elaboração de documentos em diferentes tipos e textos, no sentido de automatizar alguns processos, além de auxiliar na uniformização dos padrões de uso linguístico (BERBER SARDINHA, 2004) e acelerar a produção multilíngue de textos;

4. Criação de novos *templates* para consolidar a aplicação do perfil metafuncional em diferentes tipos de textos; e
5. Replicação da metodologia por meio do ambiente de programação R, do etiquetador morfossintático Treetagger e do programa UAM CorpusTools.

Como limitações deste estudo apontam-se:

1. A intervenção muito ativa do pesquisador para a realização das análises dos sistemas de TRANSITIVIDADE, MODO, e TEMA e MENSAGEM;
2. A elaboração de um único um *template* que permitiu a elaboração de um texto do tipo HABILITAR, com base no perfil metafuncional prototípico identificado;
3. A utilização de gráficos e dendrogramas, enquanto proposta de contribuição metodológica a partir de dados categóricos, pode ser entendida mais como uma análise exploratória e não conceitual, principalmente por ser uma técnica pouco explorada nos estudos linguísticos e por se tratar de dados categóricos.

Finalmente, vale lembrar que esta pesquisa é parte de um projeto conjunto desenvolvido entre o Laboratório Experimental de Tradução da Faculdade de Letras da UFMG (LETRA) e o Núcleo de Ações e Pesquisa em Apoio Diagnóstico da Faculdade de Medicina da UFMG (NUPAD/FM/UFMG)

# **REFERÊNCIAS BIBLIOGRÁFICAS**

## REFERÊNCIAS

- BAKER, M. Corpus linguistics and translation studies: implications and applications. In: BAKER, M.; FRANCIS, G; TOGNINI-BONELLI, E. (Ed.). *Text and technology: in honour of John Sinclair*. Amsterdã: John Benjamins Publishing Company, 1993. p. 233-250.
- BAKER, M. Corpora in translation studies: an overview and some suggestions for future research. *Target*, Amsterdã, v. 7, n. 2, p. 223-243, 1995.
- BAKER, M. Corpus-based translation studies: the challenges that lie ahead. In: SOMERS, H. (Ed.). *Terminology, LSP and translation: studies in language engineering in honour of Juan C. Sager*. Amsterdã: John Benjamins Publishing Company, 1996. p. 177-186.
- BERBER SARDINHA, T. *Linguística de corpus*. Barueri: Manole, 2004.
- BEVILACQUA, C. R. Fraseologia: perspectiva da língua comum e da língua especializada. *Revista Língua & Literatura*, São Paulo, v. 6/7, n. 10/11, p. 73-86, 2004/2005.
- BIBER, D. *Variation across speech and writing*. Cambridge: Cambridge University Press, 1988.
- BIBER, D. M. Methodological issues regarding corpus-based analyses of linguistic variation. *Literary and Linguistic Computing*, Oxford, v. 5, n. 4, p. 257-269, 1990.
- BIBER, D. Representativeness in corpus design. *Literary and Linguistic Computing*, Oxford, v. 8, n. 4, p. 243-257, 1993.
- BIBER, D. *Dimensions of register variation: a cross-linguistic comparison*. Cambridge: Cambridge University Press, 1995.
- BIBER, D.; CONRAD, S.; REPPEN, R. *Corpus linguistics: investigating language structure and use*. Cambridge: Cambridge University Press, 1998.
- BIBER, D.; JOHANSSON, S.; LEECH, G.; CONRAD, S.; FINEGAN, E. *Longman grammar of spoken and written English*. Essex: Pearson Education Limited, 1999.
- BRAGA, A. B. C. *A popularização da popularização da ciência: gramaticalização da interação, avaliação, representação da experiência, construção e organização discursiva*. 2013. 101 f. Monografia (Graduação em Letras) – Faculdade de Letras, Universidade Federal de Ouro Preto, Mariana, 2013.
- CAFFAREL, A.; MARTIN, J. R.; MATTHIESSEN, C. M. I. M; Introduction: Systemic functional typology. In: CAFFAREL, A.; MARTIN, J. R.; MATTHIESSEN, C. M. I. M. (Ed.) *Language typology: a functional perspective*. Amsterdã: John Benjamins, 2004. p. 797-825.
- CATFORD, J. *A linguistic theory of translation: an essay in applied linguistics*. Londres: Oxford University Press, 1965.
- COULTHARD, M. On analyzing and evaluating written Text. In: COULTHARD, M. (Ed.). *Advances in written text analysis*. Nova York: Routledge, 1992. p. 1-11.
- DA SILVA, I. A. L. *Conhecimento experto em tradução: aferição da durabilidade de tarefas tradutórias realizadas por sujeitos não tradutores em condições empírico-experimentais*. 2007. 273 f.

Dissertação (Mestrado em Linguística Aplicada) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2007.

EGGINS, S. Genre: context of culture in text. In: EGGINS, S. *An introduction to systemic functional linguistics*. 2. ed. Londres: Continuum, 2004. p. 54-84.

FIGUEREDO, G. P. Uma descrição sistêmico-funcional da estrutura do grupo nominal em português orientada para os estudos linguísticos da tradução. 2007. 106 f. Dissertação (Mestrado em Linguística Aplicada) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2007.

FIGUEREDO, G. P.; PAGANO, A. S. Uma abordagem linguística do texto literário traduzido: aspectos da construção do romance *They shoot horses, don't they?* e sua tradução para o português, sob a perspectiva da organização temática. *Graphos*, João Pessoa, v. 11, p. 85-103, 2009.

FIGUEREDO, G. Introdução ao perfil metafuncional do português brasileiro: contribuições para os estudos multilíngues. 2011. 383 f. Tese (Doutorado em Linguística Aplicada) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2011.

FIRTH, J. *Papers in Linguistics – 1934-1951*. Oxford: Oxford University Press, 1957.

FRANKENBERG-GARCIA, A. *Corpora e tradução*. Material didático utilizado na primeira Escola de verão da Linguateca, Módulo 3, 2006. Disponível em: <<http://www.linguateca.pt/escolaverao2006/>>. Acesso em: 18 jul. 2012.

GRANGER, S. The corpus approach: a common way forward for Contrastive Linguistics and Translation Studies? In: GRANGER, G.; LEROT, J.; PETCH-TYSON, S. (Ed.). *Corpus-based approaches to Contrastive Linguistics and Translation Studies*. Nova York: Rodopi, 2008. p. 17-29.

GRIES, S. T. Testing independent relationships. In: CHAPELLE, C. A. (Ed.). *The encyclopedia of applied linguistics*. Oxford: Wiley-Blackwell, no prelo.

GRIES, S. T. *Statistics for linguistics with R: a practical introduction*. 2. ed. Berlim: Walter de Gruyter GmbH, 2013.

HALLIDAY, M. A. K. *Language as social semiotic: the social interpretation of language and meaning*. Londres: Edward Arnold, 1978.

HALLIDAY, M. A. K. Towards a theory of good translation. In: STEINER, E.; YALLOP, C. (Ed.). *Exploring translation and multilingual text production: beyond content*. Berlim: Mouton de Gruyter, 2001. p. 13-18.

HALLIDAY, M. A. K. *Linguistic studies of text and discourse*. Londres: Continuum, 2002. (The collected works of M. A. K. Halliday, v. 2).

HALLIDAY, M. A. K.; HASAN, R. *Cohesion in English*. Essex: Longman, 1976.

HALLIDAY, M. A. K. *An introduction to functional grammar*. 1. ed. Londres: Arnold, 1985.

HALLIDAY, M. A. K. The construction of knowledge and value in the grammar of scientific discourse, with reference to Charles Darwin's the origin of species. In: COULTHARD, M. (Ed.). *Advances in written text analysis*. Nova York: Routledge, 1994. p. 136-156.

HALLIDAY, M. A. K.; MATTHIESSEN, C. M. I. M. *Construing experience through meaning: a language-based approach to cognition*. Londres: Continuum, 1999.

HALLIDAY, M. A. K.; MATTHIESSEN, C. M. I. M. *An introduction to functional grammar*. 3. ed. Londres: Arnold, 2004.

HALLIDAY, M. A. K.; McINTOSH, A.; STREVEN, P. *The linguist sciences and language teaching*. Londres: Longmans, 1964.

HANSEN, S. The nature of translated text: an interdisciplinary methodology for the investigation of the specific properties of translations. 2002. 241 f. Tese (Doutorado) – Universidade do Sarre, Saarbrücken, 2002.

HASAN, R. *Linguistics, language and verbal art*. Geelong: Deakin University Press, 1985.

HASAN, R. Text and context. In: CLORAN, C.; BUTT, D.; WILLIAMS, G. (Org.). *Ways of saying: ways of meaning*. Londres: Cassel, 1996. p. 152-183.

HOLMES, J. S. The name and nature of Translation Studies. In: Holmes, J. S. *Translated Papers on Literary Translation and Translation Studies*. Amsterdã: Rodopi, 1972. p. 67-80.

JOHANSSON, S. On the role of corpora in cross-linguistic research. In: JOHANSSON, S.; OKSEFJELL, S. (Ed.). *Corpora and cross-linguistic research: theory, method, and case studies*. Amsterdã: Rodopi, 1998. p. 3-24.

KENNING, M. M. What are parallel and comparable corpora and how can we use them? In: KEENE, A.; MCCARTHY, M. (Ed.) *The Routledge handbook of corpus linguistics*. Madison: Routledge, 2010. p. 487-500.

KRIEGER, M. G. Do reconhecimento de terminologias: entre o linguístico e o textual. In: ISQUERDO, A. N.; KRIEGER, M. G. (Org.). *As ciências do léxico*. Campo Grande: Editora UFMS, 2003. v. 2, p. 15-27.

KRIEGER, M. G.; BEVILACQUA, C. R. A pesquisa terminológica no Brasil: uma contribuição para a consolidação da área. *Debate Terminológico*, 2005. v. 1, p. 1-11.

LACERDA, P. G.; PAGANO, A. S. Aspectos da tradução de textos de divulgação da ciência: um estudo de duas retextualizações para o português brasileiro e europeu. Belo Horizonte: *Revista de Estudos da Linguagem*, v. 16, p. 145-164, 2008.

LIMA, K. C. S. Conhecimento experto em tradução: orientação e revisão em tarefas tradutórias executadas por pesquisadores expertos não tradutores. 2008. 187 f. Dissertação (Mestrado em Linguística Aplicada) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2008.

MALINOWSKI, B. *The dynamics of culture change: an inquiry into race relations in Africa*. New Haven: Yale University Press, 1945.

MCENERY, T. XIAO, R. Parallel and comparable corpora: What are they up to? In: JAMES, G. (Ed.). *Corpus Linguistics and Translation Studies*. Clevedon: Multilingual Matters, 2007. Cap. 10, p. 1-13.

MARTIN, J. R. *English text: system and structure*. Amsterdã: John Benjamins Publishing Company, 1992.

MARTIN, J. R. Tenderness: Realisation and Instantiation in a Botswanan Town. NØRGAARD, N. (Ed.). *Systemic Functional Linguistics in Use*. Odense Working Papers in Language and Communication v. 29, 2008.

MARTIN, J. R.; ROSE, D. *Working with discourse: meaning beyond the clause*. 2. ed. Londres: Continuum, 2007.

MATTHIESSEN, C. M. I. M. The environments of translation. In: STEINER, E.; YALLOP, C. (Ed.). *Exploring translation and multilingual text production: beyond content*. Berlim: Mouton de Gruyter, 2001. p. 11-26.

MATTHIESSEN, C.; TERUYA, K.; WU, C. Multilingual studies as a multi-dimensional space of interconnected language studies. In: WEBSTER, J. (Ed.). *Meaning in context: implementing intelligent applications of language studies*. Londres: Continuum, 2008. p. 147-189.

MATTHIESSEN, C.; TERUYA, K.; LAM, M. *Key terms in Systemic-Functional Linguistics*. Londres: Continuum, 2010.

MUNDAY, J. *Introducing translation studies*. Londres: Routledge, 2001.

NEUMANN, S. Corpus design. In: NEUMANN, S. *Linguistic properties of translations: a corpus-based investigation for the language pair English-German*. Deliverable n. 1. DFG project STE 840/5-1, 2005.

NEUMANN, S. *Contrastive register variation: a quantitative approach to the comparison of English and German*. Habilitationsschrift. Saarbrücken: Philosophische Fakultät II, Universität des Saarlandes, 2008.

NUNES, L. P. *As conjunções but e mas em textos ficcionais originais e traduzidos: uma abordagem tridimensional com base na Linguística Sistêmico-Funcional*. 2010. 187 f. Dissertação (Mestrado em Estudos Linguísticos) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2010.

NUNES, L. P. *Relações coesivas e estruturais em cópulas combinado: uma análise de conjunções em textos originais e traduzidos*. 94 f. Texto de Qualificação (Doutorado em Linguística Aplicada) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2013.

O'DONNELL, M. The UAM CorpusTool: software for corpus annotation and exploration. In: BRETONES CALLEJAS, C. M. (Ed.). *Applied Linguistics Now: understanding language and mind*. Alméria: Universidad de Améria, 2008. p. 1433-1447.

PAGANO, A. S. América Latina, tradução e pós-colonialismo. *Alfa*, São Paulo, v. 44, p. 157-167, 2000.

PAGANO, A. S. Organização temática e tradução. In: ALVES, F.; MAGALHÃES, C.; PAGANO, A. (Org.). *Competência em tradução: cognição e discurso*. Belo Horizonte: Editora da UFMG, 2005. p. 247-299.

PAGANO, A. S. Abordagens sistêmicas da tradução. In: CALDAS-COULTHARD, C.; SCLIAR-CABRAL, L. (Org.). *Desvendando discursos: conceitos básicos*. Florianópolis: Editora da UFSC, 2007. p. 255-288.

PAGANO, A. S. *Modelagem sistêmico-funcional da tradução e da produção textual multilíngue*. 2012. 130 f. Tese (Concurso Público para Professor Titular) – Faculdade de Letras, Universidade Federal de Minas Gerais, Belo Horizonte, 2012.

PAGANO, A. S.; FIGUEREDO, G. P. Gramaticalização da dor em português e espanhol. In: VIANA, V.; TAGNIN, S. (Org.). *Corpora no ensino de línguas estrangeiras*. São Paulo: Hub Editorial, 2011. p. 269-300.

PAGANO, A. S.; FIGUEREDO, G. P.; BRAGA, C. N. O. Estórias/Stories: aspectos temáticos da Tradução de Guimarães Rosa para o inglês. In: DUARTE, L. P. (Org.). *Veredas de Rosa III*. Belo Horizonte: CESPUC, 2007. p. 162-172.

PAGANO, A. S.; VASCONCELLOS, M. L. Estudos da Tradução no Brasil: reflexões sobre teses e dissertações elaboradas por pesquisadores brasileiros nas décadas de 1980 e 1990. *DELTA*, São Paulo, v. 19, n. especial, 2003. Disponível em: <[http://www.scielo.br/scielo.php?italic=sci\\_arttext&pid=S0102-44502003000300003&lng=en&nrm=iso](http://www.scielo.br/scielo.php?italic=sci_arttext&pid=S0102-44502003000300003&lng=en&nrm=iso)>. Acesso em: 23 maio 2013.

PAGANO, A. S.; VASCONCELLOS, M. L. Explorando interfaces: Estudos da Tradução, Linguística Sistêmico-Funcional e linguística de corpus. In: ALVES, F.; MAGALHÃES, C.; PAGANO, A. S. (Ed.). *Competência em tradução: cognição e discurso*. Belo Horizonte: Editora da UFMG, 2005. p. 177-207.

SCHMID, H. Probabilistic Part-of-speech tagging using decision trees. In: INTERNATIONAL CONFERENCE ON NEW METHODS IN LANGUAGE PROCESSING, Manchester, UK, 1994. *Proceedings...* Atenas: Tehnografia Digital Press, 2009. p. 50-58.

SERJEANT, G; SERJEANT, B. *Sickle cell disease*. Nova York: Oxford University Press, 1991.

SINCLAIR, J. M. *Corpus, concordance and collocation*. Oxford: Oxford University Press, 1991.

SINCLAIR, J. M. Preface. In: GHADDESSY, M.; ROBERT, A.; ROSEBERRY, L. (Ed.). *Small corpus studies and ELT: theory and practice*. Amsterdã: John Benjamins, 2001. p. vii-xv.

SIQUEIRA, A. L; TIBURCIO, J. D. *Estatística na área da Saúde: conceitos, metodologia, aplicações e prática computacional*. Belo Horizonte: Coopmed, 2011.

STEINER, E. *Explicitation, its lexicogrammatical realization, and its determining (independent) variables – towards an empirical and corpus-based methodology*. Saarbrücken: Institut für Angewandte Sprachwissenschaft sowie Übersetzen und Dolmetschen – Universität des Saarlandes, 2005a. 43p. Relatório.

TEICH, E. Contrastive linguistics and translation studies revisited. In: GIL, A. (Org.) *Modelle der Translation: Grundlagen für Methodik, Bewertung, Computermodellierung*. Berlin: Lang, 1999. p. 507-521.

TOGNINI-BONELLI, E. *Corpus linguistics at work*. Amsterdã: John Benjamins, 2001.

URE, J. *Text types classified by situational factors*. Atlanta: MS, 1989.

VELA, M.; HANSEN-SCHIRRA, S. The use of multi-level annotation and alignment for the translator. In: ASLIB TRANSLATING AND THE COMPUTER CONFERENCE. Londres. *Proceedings der ASLIB Translating and the computer 28 conference*. 28 nov. 2006. Disponível em: <[http://fr46.uni-saarland.de/croco/publication\\_en.html](http://fr46.uni-saarland.de/croco/publication_en.html)>. Acesso em: 14 jun. 2012.

# **ANEXOS**

## Anexo 1. Relação dos *Scripts* Utilizados na Pesquisa

```

#####
##### Limpar texto e Ilha de Palavras: para uma palavra #####
#####

x=scan(file=file.choose(),what="char",sep="\n",skip=0,quiet=F) # importa o texto para o R
## algumas limpezas necessárias
x=gsub("[.,:()]", "", x, ignore.case=T, perl=T)
x=gsub("\'", "", x, ignore.case=T, perl=T) ## elimina todas as aspas dos textos
x=gsub("[%]", "por_cento", x, ignore.case=T, perl=T)
x=gsub("-", "_", x, ignore.case=T) # replaces all the hyphens for underscore
x=gsub(" ", "", x, perl=T) # this gets rid of EXTRA spaces
x=gsub("\\W$", "", x, perl=T) ## this gets rid of empty spaces at the end of statements
x=gsub("-", "_", x, ignore.case=T) # replaces all the hyphens for underscore
x=tolower(x) ## tudo minúsculo.
x=gsub("sickle cell disease", "sickle_cell_disease", x, perl=T) # anemia falciforme as one word

##Frequência das palavras:
y=strsplit(x, " ")
y_data=data.frame(table(unlist(y)))
y_data=y_data[order(y_data$Freq,decreasing=T),]
y_data
####

## Análise das ilhas -- você sóprecisa decidir quantas palavras e os termos a serem analisados
## Análise com UMA palavra ANTES e UMA depois de Anemia Falciforme (concordance_1)
list.frames=c()
for(i in 1:length(y)){# go through the list
  current.line=y[[i]] # access each line
  for(j in 1:(length(current.line)-1)){ #go up to the antepenultimate words, thus -1

    list.frames=append(list.frames,paste(current.line[j],current.line[(j+1)],current.line[(j+2)],current.line[(j+3)],current.line[(j+4)],
current.line[(j+5)], current.line[(j+6)],current.line[(j+7)],current.line[(j+8)],current.line[(j+9)],current.line[(j+10)],collapse=" ")) # print into
a file

    }# end of loop go through the line
  }#end of loop go through the list

match=c()
list.frames.split=strsplit(list.frames, " ")
for(k in 1:length(list.frames.split)){
  match=append(match,ifelse(list.frames.split[[k]][6]=="sickle_cell_disease",list.frames.split[k], "DELETE"))
}
match=match[which(match!="DELETE",)]

## Concordance
target=c()
A1=c()
A2=c()
A3=c()
A4=c()
A5=c()
D1=c()
D2=c()
D3=c()
D4=c()
D5=c()
for(o in 1:length(match)){# go through the match list
  A1=append(A1,match[[o]][1]) # print into a file
  A2=append(A2,match[[o]][2])
  A3=append(A3,match[[o]][3])
  A4=append(A4,match[[o]][4])
  A5=append(A5,match[[o]][5])
  target=append(target,match[[o]][6])
  D1=append(D1,match[[o]][7])
  D2=append(D2,match[[o]][8])
  D3=append(D3,match[[o]][9])
  D4=append(D4,match[[o]][10])
  D5=append(D5,match[[o]][11])
}# end of loop go through the line

concordance_anemia=data.frame(A1,A2,A3,A4,A5,target,D1,D2,D3,D4,D5)
concordance_anemia 7

```

```

#####
##### Ilha de Palavras: para duas palavras #####
#####

SEM_UNDERSCORE="predomina a"
COM_UNDERSCORE="doença_falciforme"

# importa o texto para o R:
x=scan(file=file.choose(),what="char",sep="\n",skip=0,quiet=F)

## algumas limpezas necessárias:
x=gsub("[,;:()]", "", x, ignore.case=T, perl=T)
x=gsub("\", "", x, ignore.case=T, perl=T) ## elimina todas as aspas dos textos
x=gsub("%]", "por_cento", x, ignore.case=T, perl=T)
x=gsub("-", "_", x, ignore.case=T) x=gsub(" ", "", x, perl=T) # this gets rid of EXTRA spaces
x=gsub("\\W$", "", x, perl=T) ## this gets rid of empty spaces at the end of statements
x=gsub("-", "_", x, ignore.case=T) # replaces all the hyphens for underscore
x=tolower(x) ## tudo minúsculo.
x=gsub(SEM_UNDERSCORE, COM_UNDERSCORE, x, perl=T)

##Frequência das palavras
y=strsplit(x, " ")
y_data=data.frame(table(unlist(y)))
y_data=y_data[order(y_data$Freq,decreasing=T),]
y_data
####

## Análise das ilhas -- precisa decidir quantas palavras e os termos a serem analisados
## Análise com UMA palavra ANTES e UMA depois de Anemia Falciforme (concordance_1)
list.frames=c()
for(i in 1:length(y)){# go through the list
  current.line=y[[i]] # access each line
  for(j in 1:(length(current.line)-1)){ #go up to the antepenultimate words, thus -1

    list.frames=append(list.frames,paste(current.line[j],current.line[(j+1)],current.line[(j+2)],current.line[(j+3)],current.line[(j+4)],
current.line[(j+5)], current.line[(j+6)],current.line[(j+7)],current.line[(j+8)],current.line[(j+9)],current.line[(j+10)],collapse=" ")) # print into
a file

    }# end of loop go through the line
  }#end of loop go through the list

match=c()
list.frames.split=strsplit(list.frames, " ")
for(k in 1:length(list.frames.split)){
  match=append(match,ifelse(list.frames.split[[k]][6]==COM_UNDERSCORE,list.frames.split[k], "DELETE"))
}
match=match[which(match!="DELETE"),]

## Concordance
target=c()
A1=c()
A2=c()
A3=c()
A4=c()
A5=c()
D1=c()
D2=c()
D3=c()
D4=c()
D5=c()
for(o in 1:length(match)){# go through the match list
  A1=append(A1,match[[o]][1]) # print into a file
  A2=append(A2,match[[o]][2])
  A3=append(A3,match[[o]][3])
  A4=append(A4,match[[o]][4])
  A5=append(A5,match[[o]][5])
  target=append(target,match[[o]][6])
  D1=append(D1,match[[o]][7])
  D2=append(D2,match[[o]][8])
  D3=append(D3,match[[o]][9])
  D4=append(D4,match[[o]][10])
  D5=append(D5,match[[o]][11])
}# end of loop go through the line

concordance_anemia=data.frame(A1,A2,A3,A4,A5,target,D1,D2,D3,D4,D5)
concordance_anemia

write.table(concordance_anemia, "C:/Users/Kelen/Documents/popular.txt", row.names=F, sep="\t", quote=F) ## abre no Excel, mas tem
sempre que mudar o nome do arquivo, senão salva com o mesmo nome

```

```
#####
Comandos para criação do gráfico "Nuvem de Palavras" com palavras em português
#####
```

```
### Após a instalação do ambiente de programação R é preciso carregar os seguintes pacotes que são necessários para construção do gráfico:
```

```
require(tm)
require(wordcloud)
require(RColorBrewer)
```

```
### Para carregar um arquivo:
```

```
x=scan(file=file.choose(),what="char",sep="\n",skip=0,quiet=F)
```

```
### Preparação e limpeza dos corpora:
```

```
y=strsplit(x," ")
a=data.frame(table(unlist(x)))
a[,1]
ap.corpus<- Corpus(DataframeSource(data.frame(as.character(a[,1]))))
ap.corpus<- tm_map(ap.corpus, removePunctuation) # Remoção de pontuação
ap.corpus<- tm_map(ap.corpus, tolower) # Coloca todas as palavras em letra minúscula
```

```
ap.corpus<- tm_map(ap.corpus, function(x) removeWords(x,c("nosso","duas","devese","dessa","esses","trinta","atal","tais","com",
"para","suas","nove","1998","elas","este","estas","estes","três","eou","além","e","ou","sua","pela","sem",
"seus","isso","pela","for","após","ser","anos","uma","mais","entre","aos","mas","essa","pois","ou","se",
"são","seis","até","dois","sete","como","seu","ela","para","dez","seis","cinco","com","que","por","de",
"dos","das","não","das","dos","para","nos","nas"))) # Exclusão de palavras que não são de interesse na análise.
```

```
ap.tdm <- TermDocumentMatrix(ap.corpus)
ap.m <- as.matrix(ap.tdm)
ap.v <- sort(rowSums(ap.m),decreasing=TRUE)
ap.d <- data.frame(word = names(ap.v),freq=ap.v) # Verifica quais palavras apareceram com maior frequência
table(ap.d$freq)
pal2 <- brewer.pal(8,"Dark2")
```

```
### Se no gráfico estiver aparecendo poucas palavras ou se as dimensões ficarem ruins, altere o primeiro termo da "scale"!!!
```

```
wordcloud(ap.d$word,ap.d$freq, scale=c(5,.2),min.freq=3,
max.words=Inf, random.order=FALSE, rot.per=.15, colors=pal2) ### Se no gráfico estiver aparecendo poucas palavras, alterar o primeiro termo da "scale"!!!
title("Acréscimo_um_nome_desejado_ao_gráfico") ### Adicionar um título ao gráfico
windows() ### Abre uma nova janela gráfica
```

```
#####
Comandos para criação do gráfico "Nuvem de Palavras" com palavras em inglês
#####

### É necessário carregar os pacotes abaixo para a construção do gráfico:
require(tm)
require(wordcloud)
require(RColorBrewer)

### Escolha o corpora em inglês:
x=scan(file=file.choose(),what="char",sep="\n",skip=0,quiet=F)

### Preparação dos corpora:
y=strsplit(x," ")
a=data.frame(table(unlist(x)))
a[,1]
ap.corpus<- Corpus(DataframeSource(data.frame(as.character(a[,1]))))
ap.corpus<- tm_map(ap.corpus, removePunctuation) ### Remoção de pontuação
ap.corpus<- tm_map(ap.corpus, tolower) ### Coloca todas as palavras em letra minúscula.
ap.corpus<- tm_map(ap.corpus, function(x) removeWords(x, stopwords("english")))
ap.tdm <- TermDocumentMatrix(ap.corpus)
ap.m <- as.matrix(ap.tdm)
ap.v <- sort(rowSums(ap.m),decreasing=TRUE)
ap.d <- data.frame(word = names(ap.v),freq=ap.v) # Verifica quais palavras apareceram com maior frequência.
table(ap.d$freq)
pal2 <- brewer.pal(8,"Dark2")

### Se no gráfico estiver aparecendo poucas palavras ou se as dimensões ficarem ruins, modificar os termos em "scale"!!!
wordcloud(ap.d$word,ap.d$freq, scale=c(6,.4),min.freq=3,
max.words=Inf, random.order=FALSE, rot.per=.15, colors=pal2)
title("Acrescentar_um_nome_desejado_ao_gráfico") ### Adicionar um título ao gráfico
windows() ### Abre uma nova janela gráfica
```

```

#####
#### Contador de Classes de Palavras ####
#####

rm(list=ls(all=T))

## Muda o diretório, especificando o local onde se encontram os corporas
setwd("//sql/Pasta_Base$/fernando/Corpora-Anemia Falciforme/Textos Etiquetados")

## lista todos os arquivos que estão no endereço indicado
filenames=list.files(path="//sql/Pasta_Base$/fernando/Corpora-Anemia Falciforme/Textos Etiquetados")
for (a in filenames){
name=gsub(" ", "_",a) ## substitui espaço vazio por underline
name=gsub(".txt","",name) ## retira o escrito .txt do nome dos bancos...
a=paste("//sql/Pasta_Base$/fernando/Corpora-Anemia Falciforme/Textos Etiquetados/",a,sep="")
assign(name,read.csv2(a, header=FALSE,sep=""))
}

## Até aqui será o mesmo comando para todos os textos:##

tudo=rbind(Artigo_científico_etiquetado_portugues) ## Substituir pelo nome de interesse dos corpora
dados=tudo
dados=na.omit(dados)
colnames(dados)=c("palavras","tags") ## atribui nomes às colunas do banco.

dados=dados[dados$tags!="LS",] # retira os LS ##
dados=dados[dados$tags!="PDT",] # retira os PDT retirar
dados=dados[dados$tags!="POS",] # retira os POS
dados=dados[dados$tags!="PRP+P",] # retira os PRP+P
dados=dados[dados$tags!="PRP+PRP",] # retira os PRP+PRP
dados=dados[dados$tags!="PRP+NOM",] # retira os PRP+NOM
dados=dados[dados$tags!="RP",] # retira os RP
dados=dados[dados$tags!="G",] # retira os G
dados=dados[dados$tags!="IL",] # retira os IL
dados=dados[dados$tags!="X",] # retira os X
dados=dados[dados$tags!="EX",] # retira os EX
dados=dados[dados$tags!="P+P",] # retira os P+P
dados=as.data.frame(lapply(dados,function(x)x[,drop=TRUE])) ## delete unused levels
tag=gsub(" CONJSUB", "CONJSUB",dados$tags)
tag=gsub(" IN ", "IN",tag)
tag=gsub("\\bL\\b","ADJ",tag)
dados$tags=tag
#dados[238576,2]="ADJ" ## corrige a única entrada sem tag
dados$tags=as.factor(dados$tags)
dados$tags=gsub("\\+","_",dados$tags)
dados$tags=gsub("\\$","S",dados$tags)
dados$tags=gsub("\\bIN\\b","preposicao",dados$tags)
dados$tags=gsub("\\bCONJSUB\\b","preposicao",dados$tags)
## Classificação
dados$tags=gsub("\\bJJ\\b","adjetivo",dados$tags) ## substitui a classificação especificada pela designação adjetivo, advérbio, substantivo
etc.
dados$tags=gsub("\\bJJR\\b","adjetivo",dados$tags)
dados$tags=gsub("\\bJJS\\b","adjetivo",dados$tags)
dados$tags=gsub("\\bADJ\\b","adjetivo",dados$tags)
dados$tags=gsub("\\bRB\\b","adverbio",dados$tags)
dados$tags=gsub("\\bRBR\\b","adverbio",dados$tags)
dados$tags=gsub("\\bRBS\\b","adverbio",dados$tags)
dados$tags=gsub("\\bWRB\\b","adverbio",dados$tags)
dados$tags=gsub("\\bADV\\b","adverbio",dados$tags)
dados$tags=gsub("\\bFW\\b","substantivo",dados$tags)
dados$tags=gsub("\\bNN\\b","substantivo",dados$tags)
dados$tags=gsub("\\bNNS\\b","substantivo",dados$tags)
dados$tags=gsub("\\bNP\\b","substantivo",dados$tags)
dados$tags=gsub("\\bNPS\\b","substantivo",dados$tags)
dados$tags=gsub("\\bF\\b","substantivo",dados$tags)
dados$tags=gsub("\\bB\\b","substantivo",dados$tags)
dados$tags=gsub("\\bNOM\\b","substantivo",dados$tags)
dados$tags=gsub("\\bS\\b","substantivo",dados$tags)
dados$tags=gsub("\\bIN\\b","preposicao",dados$tags)
dados$tags=gsub("\\bTO\\b","preposicao",dados$tags)
dados$tags=gsub("\\bPRP\\b","preposicao",dados$tags)
dados$tags=gsub("\\bPRP_DET\\b","preposicao",dados$tags)
dados$tags=gsub("\\bPRP_ADV\\b","preposicao",dados$tags)
dados$tags=gsub("\\bCC\\b","conjuncao",dados$tags)
dados$tags=gsub("\\bCONJ\\b","conjuncao",dados$tags)

```

```

dados$tags=gsub("\\bDT\b","determinante",dados$tags)
dados$tags=gsub("\\bDET\b","determinante",dados$tags)
dados$tags=gsub("\\bPP\b","pronome",dados$tags)
dados$tags=gsub("\\bWDT\b","pronome",dados$tags)
dados$tags=gsub("\\bWP\b","pronome",dados$tags)
dados$tags=gsub("\\bWPS\b","pronome",dados$tags)
dados$tags=gsub("\\bPPS\b","pronome",dados$tags)
dados$tags=gsub("\\bP\b","pronome",dados$tags)
dados$tags=gsub("\\bPR\b","pronome",dados$tags)
dados$tags=gsub("\\bCD\b","numero",dados$tags)
dados$tags=gsub("\\bCARD\b","numero",dados$tags)
dados$tags=gsub("\\bUH\b","interjeicao",dados$tags)
dados$tags=gsub("\\bI\b","interjeicao",dados$tags)
dados$tags=gsub("\\bMD\b","verbo",dados$tags)
dados$tags=gsub("\\bVB\b","verbo",dados$tags)
dados$tags=gsub("\\bVBD\b","verbo",dados$tags)
dados$tags=gsub("\\bVBG\b","verbo",dados$tags)
dados$tags=gsub("\\bVBN\b","verbo",dados$tags)
dados$tags=gsub("\\bVBP\b","verbo",dados$tags)
dados$tags=gsub("\\bVBZ\b","verbo",dados$tags)
dados$tags=gsub("\\bVVG\b","verbo",dados$tags)
dados$tags=gsub("\\bVV\b","verbo",dados$tags)
dados$tags=gsub("\\bVHN\b","verbo",dados$tags)
dados$tags=gsub("\\bVHZ\b","verbo",dados$tags)
dados$tags=gsub("\\bVHD\b","verbo",dados$tags)
dados$tags=gsub("\\bVHG\b","verbo",dados$tags)
dados$tags=gsub("\\bVVD\b","verbo",dados$tags)
dados$tags=gsub("\\bVVZ\b","verbo",dados$tags)
dados$tags=gsub("\\bVH\b","verbo",dados$tags)
dados$tags=gsub("\\bVVN\b","verbo",dados$tags)
dados$tags=gsub("\\bVVP\b","verbo",dados$tags)
dados$tags=gsub("\\bVHP\b","verbo",dados$tags)
dados$tags=gsub("\\bV\b","verbo",dados$tags)
dados$tags=gsub("\\bV_P\b","verbo",dados$tags)
dados$tags=gsub("\\bV_P_d\b","verbo",dados$tags)
data.frame(table(dados$tags)) ## tabela os resultados obtidos.

```

```

dim(Artigo_cientifico_etiquetado_ingles) ## para saber quantas palavras há no corpora - sem eliminação de nenhuma palavra.
dim(Artigo_cientifico_etiquetado_portugues)
dim(Tecnico_etiquetado_portugues)
dim(Tecnico_etiquetado_ingles)
dim(Popular_etiquetado_portugues)
dim(Popular_etiquetado_ingles)

```

```
#####  
# Teste qui-quadrado de associação dos textos com as classes de palavras  
#####
```

```
## Artigo acadêmico (ac) ##
```

```
ac=matrix(c(366,566,129,120,689,977,80,77,1361,1389,596,599),6,2,T)  
chisq.test(ac)
```

```
## Cartilha e folder de divulgação (tp) ##
```

```
tp=matrix(c(277,544,245,129,438,674,196,80,921,1132,666,601),6,2,T)  
chisq.test(tp)
```

```
## Manual Técnico (tt) ##
```

```
tt=matrix(c(318,540,134,97,418,679,75,92,1094,1038,519,565),6,2,T)  
chisq.test(tt)
```

```
#####  
#####===== Comandos de Gráfico 3D =====#####  
#####  
require(scatterplot3d) # carrega pacote para realização de gráfico 3D.  
dado=read.table("Tabela Artigo inglês.txt",T) # abre banco para gráfico 3D.  
attach(dado) #Fixa o nome das variáveis.  
  
x=Modo # associa à letra X a variável Modo.  
y=Transitividade # associa à letra Y a variável Transitividade.  
z=Tema # associa à letra Z a variável Tema.  
  
scatterplot3d(x, y, z, type="p",highlight.3d=F, #  
col.axis="blue",col.grid="lightblue", main="", pch=20,  
xlab="Modo", ylab="Transitividade",zlab="Tema")
```

```
#####  
#####==== Comandos de Dendrograma ====#####  
#####  
ArtigoPortuguês=read.table("Tabela Artigo Português.txt",T) # abreo banco  
ArtigoPortuguês=ArtigoPortuguês[,-1]# retira a coluna de numeração das orações  
ArtigoPortuguês=ArtigoPortuguês[,-4]# retira a coluna 'mensagens'  
  
Artigo_Português=dist(ArtigoPortuguês) # calcula distâncias Euclidianas  
  
cluster.s=hclust(Artigo_Português,method="complete")  
plclust(cluster.s,ylab="Distância") # plota o dendrograma
```

## Anexo 2. Tabelas Construídas com Base nos Dados do UAM CorpusTool

### Artigo inglês

Oração	Modo X	Transitividade Y	Tema Z	Mensagem W
1	1	3	1	1
2	1	2	1	2
3	1	1	2	2
4	1	1	1	2
5	1	2	1	2
6	1	2	1	3
7	1	2	3	1
8	1	2	1	2
9	1	2	1	2
10	1	2	1	2
11	1	2	1	2
12	1	2	1	2
13	1	2	1	1
14	1	2	1	2
15	1	2	3	2
16	1	2	1	2
17	1	2	1	1
18	1	1	1	2
19	1	1	1	2
20	1	2	1	2
21	1	2	1	2
22	1	2	1	3
23	1	2	4	1
24	1	2	1	3
25	1	2	1	1
26	1	3	1	2
27	1	2	1	2
28	1	2	1	2
29	1	2	1	2
30	1	2	1	2
31	1	2	3	2
32	1	2	3	2
33	1	1	1	1
34	1	2	1	2
35	1	2	1	2
36	1	2	3	2
37	1	2	1	2
38	1	2	3	2
39	1	2	1	1
40	1	2	1	2
41	1	2	3	2
42	1	2	1	2
43	1	2	1	2
44	1	2	1	2
45	1	2	1	2
46	1	5	1	3
47	1	2	1	2
48	1	3	1	2
49	1	1	1	2
50	1	2	1	3
51	1	2	1	2
52	1	2	1	2
53	1	2	1	3
54	1	1	1	2
55	1	1	1	2
56	1	2	1	2
57	1	2	3	2

## Artigo português

Oração	Modo X	Transitividade Y	Tema Z	Mensagem W
1	1	2	1	1
2	1	2	2	2
3	1	2	1	2
4	1	2	3	2
5	1	2	1	2
6	1	2	3	2
7	1	2	1	2
8	1	2	1	3
9	1	4	1	2
10	1	1	3	2
11	1	2	1	2
12	1	2	1	2
13	1	1	3	2
14	1	5	1	2
15	1	2	2	3
16	1	2	1	2
17	1	2	3	2
18	1	1	2	5
19	1	2	2	3
20	1	2	2	4
21	1	2	1	3
22	1	2	1	2
23	1	1	3	2
24	1	1	3	2
25	1	1	2	1
26	1	1	3	2
27	1	4	3	2
28	1	1	3	2
29	1	4	3	2
30	1	4	3	2
31	1	1	3	2
32	1	4	2	2
33	1	1	3	3
34	1	2	1	2
35	1	2	3	2
36	1	2	1	2
37	1	2	1	2
38	1	2	1	2
39	1	1	1	3
40	1	1	1	1
41	1	1	3	2
42	1	1	1	2
43	1	1	1	2
44	1	2	3	2
45	1	2	1	2
46	1	1	3	2
47	1	1	3	2
48	1	1	1	2
49	1	2	1	2
50	1	2	3	4
51	1	2	1	2
52	1	2	1	1
53	1	2	1	2
54	1	2	1	1
55	1	2	1	2
56	1	2	1	2
57	1	2	1	1
58	1	2	1	4
59	1	2	3	3
60	1	2	1	2
61	1	1	3	2
62	1	2	3	2
63	1	2	3	2

## Manual inglês

Oração	Modo X	Transitividade Y	Tema Z	Mensagem W
1	1	2	1	1
2	1	1	3	2
3	1	4	1	2
4	1	2	1	2
5	1	2	1	1
6	1	2	3	2
7	1	4	1	3
8	1	2	3	2
9	1	2	3	2
10	1	2	1	1
11	1	2	1	2
12	1	2	1	2
13	1	2	1	2
14	1	2	1	4
15	1	2	1	2
16	1	1	1	2
17	1	2	1	1
18	1	2	3	2
19	1	2	1	2
20	1	2	1	4
21	1	2	1	2
22	1	2	1	2
23	1	2	1	2
24	1	3	1	1
25	1	2	1	2
26	1	2	1	2
27	1	2	1	2
28	1	2	1	2
29	1	2	1	4
30	1	2	1	2
31	1	2	1	1
32	1	2	3	2
33	1	2	1	2
34	1	3	1	1
35	1	2	1	2
36	1	2	3	2
37	1	1	1	4
38	1	2	1	2
39	1	2	1	2
40	1	2	1	2
41	1	2	1	2
42	1	4	1	1
43	1	4	3	2
44	1	4	1	2
45	1	2	3	2
46	1	2	1	4
47	1	2	1	2
48	1	2	3	2
49	1	1	1	2
50	1	1	3	2
51	1	2	1	2
52	1	2	1	2

## Manual Português

Oração	Modo X	Transitividade Y	Tema Z	Mensagem W
1	1	2	1	1
2	1	2	1	2
3	1	2	3	2
4	1	2	1	3
5	1	3	1	4
6	1	2	1	2
7	1	2	3	2
8	1	2	1	1
9	1	2	1	2
10	1	3	1	2
11	1	2	1	2
12	1	1	1	3
13	1	2	3	2
14	1	1	3	2
15	1	2	1	2
16	1	1	1	2
17	1	1	1	2
18	1	2	1	1
19	1	2	1	2
20	1	2	1	2
21	1	2	1	2
22	1	2	1	2
23	1	2	1	2
24	1	2	1	2
25	1	2	3	2
26	1	1	1	4
27	1	2	1	2
28	1	1	1	2
29	1	2	1	2
30	1	2	1	2
31	1	3	1	1
32	1	2	3	2
33	1	2	1	4
34	1	2	3	2
35	1	2	1	2
36	1	2	1	2
37	1	2	1	4
38	1	2	1	5
39	1	1	1	2
40	1	2	1	2
41	1	1	1	2
42	1	1	3	2
43	1	2	1	2
44	1	1	3	2
45	1	2	1	2
46	1	2	1	4
47	1	1	1	2
48	1	1	1	3
49	1	1	1	2
50	1	3	1	1
51	1	2	3	2
52	1	2	1	2
53	1	2	1	2
54	1	2	1	2
55	1	2	1	4

Oração	Modo X	Transitividade Y	Tema Z	Mensagem W
56	1	3	1	1
57	1	2	1	2
58	1	1	1	2
59	1	1	1	2
60	1	1	3	2
61	1	2	1	4
62	1	2	1	2
63	1	1	3	2
64	1	2	1	4
65	1	2	1	2
66	1	2	3	2
67	1	2	1	2
68	1	2	1	2
69	1	2	1	2
70	1	2	3	2
71	1	2	1	2
72	1	6	1	4
73	1	2	1	2
74	1	2	3	2
75	1	2	1	2
76	1	2	1	2
77	1	2	1	2
78	1	2	1	2

*Cartilha inglês*

Oração	Modo X	Transitividade Y	Tema Z	Mensagem W
1	1	2	3	1
2	1	2	1	2
3	2	1	2	4
4	1	2	1	2
5	1	2	1	2
6	1	2	1	1
7	1	2	1	2
8	1	2	1	4
9	1	2	1	2
10	1	2	2	2
11	1	2	1	1
12	1	2	1	2
13	1	2	3	2
14	1	2	1	3
15	1	2	1	2
16	1	2	1	3
17	1	1	1	2
18	1	2	1	1
19	1	2	1	2
20	1	2	1	2
21	1	2	1	2
22	1	2	1	2
23	1	2	1	4
24	1	2	1	2
25	1	2	5	2
26	1	2	1	2
27	1	2	1	1
28	1	1	1	2
29	1	1	1	2
30	1	1	1	2
31	1	2	1	2
32	1	2	3	3
33	1	1	1	2
34	1	1	3	2
35	1	2	1	2
36	1	2	1	4
37	1	1	1	2
38	1	2	1	2
39	1	2	3	2
40	1	1	1	3
41	1	2	3	2
42	1	1	1	1
43	1	3	1	2
44	1	1	1	4
45	1	2	3	2
46	1	2	1	2
47	1	2	1	2
48	1	3	1	1
49	2	2	1	2
50	1	2	1	2

Cartilha português

Oração	Modo X	Transitividade Y	Tema Z	Mensagem W
1	1	3	1	1
2	1	2	1	2
3	1	1	3	2
4	1	1	1	2
5	1	2	1	2
6	1	2	1	3
7	1	2	3	1
8	1	2	1	2
9	1	2	1	2
10	1	2	1	2
11	1	2	1	2
12	1	2	1	2
13	1	2	1	1
14	1	2	1	2
15	1	2	3	2
16	1	2	1	2
17	1	2	1	1
18	1	1	1	2
19	1	1	1	2
20	1	2	1	2
21	1	2	1	2
22	1	2	1	3
23	1	2	4	1
24	1	2	1	3
25	1	2	1	1
26	1	3	1	2
27	1	2	1	2
28	1	2	1	2
29	1	2	1	2
30	1	2	1	2
31	1	2	3	2
32	1	2	3	2
33	1	1	1	1
34	1	2	1	2
35	1	2	1	2
36	1	2	3	2
37	1	2	1	2
38	1	2	3	2
39	1	2	1	1
40	1	2	1	2
41	1	2	3	2
42	1	2	1	2
43	1	2	1	2
44	1	2	1	2
45	1	2	1	2
46	1	5	1	3
47	1	2	1	2
48	1	3	1	2
49	1	1	1	2
50	1	2	1	3
51	1	2	1	2
52	1	2	1	2
53	1	2	1	3
54	1	1	1	2
55	1	1	1	2
56	1	2	1	2
57	1	2	3	2

## Legenda (textos em português)

**Modo**

- 1 – Modo Indicativo Declarativo - Oferta de Informação
- 2 – Imperativo - Demanda de Bens e Serviços

**Transitividade**

- 1 – Processo Material
- 2 – Processo Relacional Atributivo
- 3 – Processo Relacional Identificativo
- 4 – Processo Mental
- 5 – Processo Verbal
- 6 – Processo Existencial

**Tema**

- 1 – Tema *Default - Default*
- 2 - Ângulo: Fonte - *Default*
- 3 – Textual - *Default*
- 4 – Interpessoal - *Default*

**Mensagem**

- 1 – Inicial Arranjo Explícito Não Focalizado
- 2 – Fase Continuidade Arranjo-*Default* Não Focalizado
- 3 – Fase Descontinuidade Mudança Arranjo *Default* Não Focalizado
- 4 – Fase Descontinuidade Desvio Arranjo *Default* Não Focalizado
- 5 – Inicial Arranjo Implícito Não Focalizado

## Legenda (textos em inglês)

**Modo**

- 1 – Modo Indicativo Declarativo - Oferta de Informação
- 2 – Imperativo - Demanda de Bens e Serviços

**Transitividade**

- 1 – Processo Material
- 2 – Processo Relacional Atributivo
- 3 – Processo Relacional Identificativo
- 4 – Processo Existencial

**Tema**

- 1 – Simple – Não Marcado
- 2 – Simple – Marcado
- 3 – Múltiplo – Textual e Tópico – Não Marcado
- 4 – Múltiplo – Textual e Tópico – Marcado
- 5 – Múltiplo – Interpessoal e Tópico – Não Marcado

**Mensagem**

- 1 – Inicial Arranjo Explícito Não Focalizado
- 2 – Fase Continuidade Arranjo-*Default* Não Focalizado
- 3 – Fase Descontinuidade Mudança Arranjo *Default* Não Focalizado
- 4 – Fase Descontinuidade Desvio Arranjo *Default* Não Focalizado

### Anexo 3. Textos que Compuseram os *Corpora* da Descrição do Perfil Metafuncional

#### Artigo Acadêmico

Avaliação e medição criteriosas da dor falciforme – e de qualquer dor - são passos fundamentais para a programação das ações necessárias para seu alívio ou abolição, e para julgamento da efetividade das mesmas. Para os métodos usados para a avaliação devem ser os mesmos utilizados para qualquer síndrome dolorosa. A autora e outros estudiosos consideraram adequado o uso de autorrelatos durante as crises e nos seus intervalos, e enfatizaram a importância dos diários domésticos que, segundo ela, superam as distorções próprias dos relatos retrospectivos, são úteis na coleta de dados para intervenções clínicas, e favorecem estratégias de autocontrole e reconhecimento de possíveis fatores desencadeantes, propiciando sentimentos de controle do paciente sobre a doença. A forma da dor ser tratada tem implicações futuras na maneira como os pacientes conviverão com ela e com a vida. Há necessidade de se acreditar nos pacientes quando dizem estar sentindo dor, e de se instalar imediatamente analgesia adequada. As palavras de são simples e significativas: “no tratamento da crise dolorosa falciforme as queixas dos pacientes precisam ser levadas a sério e a terapia analgésica iniciada prontamente, com intensidade proporcional à gravidade da queixa dolorosa”. Como esclarecimento a receios quanto ao uso de analgesia potente nos episódios graves, de pesquisadores do declararam administrar narcóticos parenterais nos episódios graves de dor. Segundo como quase todos os episódios são manejados em casa, a avaliação da dor e das medidas de convivência com ela precisa focalizar seu manejo satisfatório em ambiente domiciliar. Essa avaliação criteriosa deve ser parte essencial de serviços de cuidados comunitários centrados na dor, que efetivamente garantem a recuperação e melhoram a qualidade de vida. Para desde a mais tenra idade a criança e seus familiares devem aprender a lidar com a dor e tratá-la, com o intuito de se alcançar menor interferência nas atividades diárias. Para os pesquisadores da Fundação Hemominas, a educação da família e do paciente em relação ao manuseio da dor é essencial. Para que se conheça e reduza o impacto real da dor sobre a função psicológica e a qualidade de vida, os episódios tratados parcial ou totalmente nos domicílios – assim como as estratégias adotadas para controle – devem ser considerados e incluídos nos estudos, cujos resultados serão inadequados se não se levar em conta essa realidade. No processo educativo é preciso conferir poder ao paciente e aos familiares, no sentido de deterem o conhecimento sobre a doença. O aconselhamento educacional e genético tem efeito positivo no prognóstico, já que os pacientes saberão identificar e evitar fatores desencadeantes, e procurar atendimento médico quando necessário. Segundo a criança com dor falciforme deve ser estimulada e treinada a identificar e desenvolver suas habilidades e acreditar em suas potencialidades, sem deixar de reconhecer suas limitações e dominar as estratégias para lidar com elas. Nas palavras dos autores, “a criança deve ser incentivada a acreditar que há mais nela do que sua doença falciforme”. Apesar de variada gravidade clínica, um contingente significativo de pacientes sofre as consequências de uma doença crônica e grave, exacerbada por crises que lhes acrescentam sofrimento, dor intensa e hospitalizações, agravando suas dificuldades escolares e/ou no trabalho e trazendo-lhes marcantes limitações no lazer e no prazer. Sabem-se possíveis portadores de disfunções orgânicas e candidatos à morte prematura por complicações, previsíveis ou não, como septicemias e tromboembolismos. É frequente a insatisfação e a insegurança com o próprio corpo, com frequente deficiência no crescimento e atraso no desenvolvimento sexual. A possibilidade de procriação pode lhes trazer mais receios do que alegria, já que faz perpetuar o “traço” e o estigma da doença. No estado de Minas Gerais, o Programa Estadual de Triagem Neonatal de Minas Gerais PETN-MG detectou a proporção de uma criança com anemia falciforme a cada nascidos vivos. Em determinadas situações, como hipóxia, desidratação e acidose, as moléculas de Hb S formam polímeros alongados que modificam o citoesqueleto das hemácias, dando origem à forma característica de foice (falcização). A polimerização da Hb S causa diversas modificações físico-químicas nas hemácias e é o evento primário indispensável para a patogênese molecular da anemia falciforme. Esse processo leva à hemólise crônica das hemácias, causando um quadro de anemia persistente. Além disso, as hemácias falcizadas se tornam mais aderentes e ligam-se umas às outras: a hemácias não falcizadas, a células nucleadas e ao endotélio vascular, causando vaso-oclusão. Os fenômenos vaso-oclusivos e a hemólise acentuada são os principais determinantes das manifestações clínicas da anemia falciforme. Apesar da alteração principal estar restrita às hemácias, trata-se de doença sistêmica cujos efeitos podem incidir sobre qualquer órgão. As manifestações clínicas variam desde morte precoce na infância a ausência de sintomas com expectativa de vida quase normal. Por ser uma anomalia da globina beta, os sintomas comumente só surgem após o sexto mês de vida, quando a produção da cadeia já é claramente predominante em relação à cadeia gama da HbF. Alguns exemplos de manifestações clínicas da anemia falciforme são: litíase biliar, crise aplástica, crises de dor, síndrome torácica aguda, acidente vascular cerebral, priapismo, sequestro esplênico, úlceras nas pernas e necrose óssea avascular. O diagnóstico laboratorial da anemia falciforme é baseado na mobilidade eletroforética mais lenta da Hb S em relação à Hb A normal. As principais técnicas utilizadas são a eletroforese de hemoglobina, focalização isoeletrica ou cromatografia líquida de alta performance (HPLC). Técnicas de diagnóstico molecular para detecção da substituição do nucleotídeo que dá origem à HbS também estão disponíveis. Fatores ambientais, como o local onde vive o paciente, prevalência de doenças infectocontagiosas, condições socioeconômicas e acesso e qualidade da assistência médica, podem influenciar no fenótipo da doença. No Brasil, a maioria dos indivíduos acometidos possui baixa renda e necessita dos cuidados médicos fornecidos pelo Sistema Único de Saúde. Apesar de ser causada pela substituição de um único nucleotídeo, o fenótipo da anemia falciforme é o produto de vários genes. A redução do nível da HbS e/ou o aumento do nível da HbF podem influenciar no processo de falcização e modificar o fenótipo da doença. Além disso, vários processos na vasculatura e órgãos influenciam no fenômeno de vaso-oclusão e também levam à diversidade de gravidade da doença

Fonte: MARTINS, P. R. J.; MORAES-SOUZA, H.; SILVEIRA, T. B. Morbimortalidade em doença falciforme. *Revista Brasileira de Hematologia e Hemoterapia.*, São Paulo, v. 32, n. 5, p. 378-383, 2010. (excerto)

### Artigo Acadêmico

**WHAT'S KNOWN ON THIS SUBJECT:** Sickle cell disease (SCD) is a multisystem disease with hemolytic and infectious complications that contribute to high morbidity and mortality rates in children. However, a comprehensive review of the literature for the prevention and management of pediatric SCD-related complications is lacking. **WHAT THIS STUDY ADDS:** This article presents a comprehensive review of the literature for the care provided to children with SCD. The results will enable practitioners and researchers to identify areas with the strongest evidence for clinical practice and those that need additional research. Sickle cell disease (SCD) is one of the most common genetic disorders in the United States and affects ~70 000 to 100 000 children and adults, predominantly of African and Hispanic descent. SCD is a multisystem disease characterized by vascular injury caused by vaso-occlusion and hemolytic and infectious complications that contribute to high morbidity and mortality rates in this population. Many severe complications of SCD occur in childhood, including sepsis, stroke, acute chest syndrome, and painful episodes that require hospitalization. Universal newborn screening and advances in routine care, including penicillin prophylaxis and immunization against serious bacterial infections, have significantly decreased mortality in young children with SCD. In addition, transcranial Doppler screening has been shown to identify children at higher risk of first-time stroke that can be treated prophylactically with chronic transfusion therapy. However, research studies are lacking in key areas for children with SCD, including studies of disease-modifying treatments. For example, the Multicenter Study of Hydroxyurea was published in 1995 and a significant decrease was shown in painful episodes and acute chest syndrome in adults, and a single-center study revealed decreased mortality with long-term use. However, only 1 small randomized controlled trial (RCT) has been published on the use of hydroxyurea in children to date. The purpose of this study was to review the quality of the literature for preventive interventions and treatment of complications for children with SCD to facilitate the use of evidence-based medicine in clinical practice and identify areas in need of additional research. In this study we have focused on level I pediatric SCD. **METHODS:** Evidence Acquisition: We identified topics important to the care of children with SCD by using guidelines, review articles and clinical expertise. In addition, we specifically searched for screening, diagnostic, and treatment interventions used in routine SCD care to ensure their inclusion in this study. The 28 topics identified are listed in Table 1. Individual topics were organized into 3 categories: (1) acute events; (2) chronic conditions; and (3) routine health care maintenance and disease-modifying treatments for SCD. Acute SCD-related events were defined as those that require urgent or emergent care. Chronic conditions included (1) SCD-specific complications that typically do not require immediate attention and persist for = 3 months (eg, avascular necrosis) and (2) chronic illnesses prevalent in the pediatric SCD population (eg, asthma). Finally, routine health care maintenance and disease-modifying treatments were defined as the comprehensive care needed for children with SCD. We searched for these topics in Ovid Medline for Evidence Level of Articles Selected for Full Review. For each topic, we used several key words to identify relevant articles. We also added the term "sickle cell" to ensure that all of the articles captured for a topic were related to SCD and its specific genotypes (eg, sickle cell anemia, hemoglobin SC disease). We limited each search to "children 0–18 years" and "human" (Appendix 1). In addition to abstracts identified electronically, we performed an ad hoc search of bibliographies of articles selected for this review. **Study Selection:** Each abstract was independently reviewed by 2 authors (Drs Kavanagh, Wang, and Sprinz). Abstracts were selected for full-article review if they pertained to the key topics identified, used experimental (eg, RCT) or quasi-experimental (eg, cohort studies with comparison groups) research methods, and enrolled children in the study. If the research methods were unclear in the abstract, the article was reviewed in detail. Studies were excluded if they enrolled adults only, were in vitro or animal studies, gene or genome-wide association studies, non-English-language abstracts or articles, commentaries or editorials, or case reports. We also excluded longitudinal cohort studies that presented outcomes aggregated over 1 or more decades without description of the care processes. Although these studies could provide important information on the historical natural history of SCD, they might not reflect outcomes that could be achieved under current standards of care. **Data Extraction:** Articles selected from the review of abstracts were retrieved for full article evaluation. Data were abstracted by each reviewer using a common form. The abstraction form included fields for the number of participants, population of interest, intervention(s) studied, and study design. Differences in coding were reconciled by the pair of reviewers (Drs Kavanagh, Wang, and Sprinz) in consultation with the third reviewer as needed. All references and abstracted data were maintained in Microsoft Excel 2003 (Redmond, WA). **Quality of the Literature:** The hierarchy of research design was assigned following the US Preventive Services Task Force ratings. RCTs and meta-analyses were categorized as level I evidence. Level II evidence included well designed controlled trials without randomization, cohort and case-control studies, multiple time series analyses, and results from uncontrolled experiments. Level III evidence was defined as studies reflecting the opinions of respected authorities on the basis of clinical experience, descriptive studies or case reports, and reports of expert committees. **RESULTS:** We identified 3188 abstracts for screening. Of these, 2867 abstracts were eliminated because of duplicate references, topic, study design, population, non-English-language abstract or article, outdated reviews or unable to be retrieved. There were 321 (10% of abstracts screened) articles selected for full review (Appendix 2), of which 26 were rated as level I evidence. We identified 25 RCTs and 1 meta-analysis from which we extracted summary data. Eighteen of the 28 topics selected for this literature review did not have level I evidence published during this 15-year period. Of the 321 articles selected for full review, 111 and 184 articles were categorized as level II and level III evidence, respectively.

Fonte: KAVANAGH, P. L.; SPRINZ, P. G.; VINCI S. R.; BAUCHNER, H.; WANG, C, J. Management of children with sickle cell disease: a comprehensive review of the literature. *Pediatrics*, Boston, v. 128, n. 6, p. 1552-1574, 2011. (excerto)

## Manual Técnico

Qualquer tipo de pneumonia exige hospitalização para antibioticoterapia venosa, suportes clínico e ventilatório. Os demais pacientes (não graves) devem ser abordados com a mesma propedêutica, porém o tratamento pode ser ambulatorial com reavaliação em 48 horas. Caso esteja indicada a antibioticoterapia empírica (febre sem foco), pode-se utilizar ampicilina, sendo crianças adulto via oral ou amoxicilina criança via oral e adulto via oral. É a segunda causa mais comum de morte em crianças menores de anos de idade. A etiologia é desconhecida; contudo, infecções virais parecem preceder à maioria dos episódios. Na anemia falciforme, o SEA pode ocorrer nos primeiros meses de idade, sendo menos frequente após os anos. Pode acontecer, acima dessa faixa etária, em pacientes nos quais a esplenomegalia é persistente (Hb SC e S-talassemias). Os sinais clínicos principais são: o aumento súbito do baço e a redução intensa da Hb, podendo evoluir para choque hipovolêmico. A hospitalização deve ser imediata. A correção da hipovolemia com soluções cristaloides pode ser feita, enquanto se aguarda a transfusão de hemácias. Normalmente, basta uma transfusão, pois a maior parte do sangue sequestrado retorna para a circulação à medida que a esplenomegalia regride. A esplenectomia deverá ser programada após duas crises de sequestro esplênico ou após um primeiro episódio grave. Em caso de queda do hematócrito (Htc), este deve ser corrigido para os níveis basais do paciente, com transfusão de concentrado de hemácias (CHM), até o limite de por transfusão. Em casos graves, realiza-se a exsanguineotransfusão. Pacientes com dor torácica, tosse, dispneia, hipoxemia, prostração e alteração radiológica recente do tórax provavelmente têm síndrome torácica aguda. Devem receber tratamento hospitalar por se tratar de uma complicação multifatorial, com alto potencial de gravidade. A constitui a principal causa de morte, em qualquer faixa etária, principalmente em adolescentes e adultos. Sua fisiopatologia compreende pneumonia, infartos, atelectasias e falcização intrapulmonar. A propedêutica infecciosa está indicada, com realização de hemograma, dosagem de proteína C reativa, hemocultura e punção pleural se houver derrame. O tratamento imediato objetiva a correção da hipoxemia, a elevação dos níveis da hemoglobina e a redução da Hb S. A saturimetria, com oximetria de pulso, deve ser instituída. A oxigenioterapia está indicada para manter a saturação de 8 MmHg. A gasometria arterial deve ser realizada para confirmação da hipoxemia. Inicia-se a antibioticoterapia venosa em doses habituais devido à alta probabilidade de infecção bacteriana: ampicilina: criança: em doses; adulto: em doses. Analgésicos opioides devem ser usados com cautela (por causa do risco de depressão respiratória). O acidente vascular cerebral (AVC) é uma das mais graves complicações da doença falciforme. Quando isquêmico, resulta de infarto em áreas irrigadas pelas artérias cerebrais do polígono de Willis. O AVC hemorrágico, mais comum em adultos, responde por dos casos e apresenta mais morbimortalidade. Ocorre em consequência de ruptura de pequenos vasos, a partir de neoformações vasculares ou de aneurismas. O AVC isquêmico ocorre, principalmente, em pacientes com Hb SS, sendo raro naqueles com Hb SC e S-talassemias. As crianças a partir dos anos de idade são mais afetadas, com incidência de até os anos. Alterações de consciência, déficits neurológicos focais, convulsões, paresias, afasia, confusão mental e cefaleia de grande intensidade ou duração são sinais e sintomas sugestivos de acidente vascular cerebral. O tratamento do episódio agudo deve ser realizado em regime hospitalar. Os pacientes com sintomas sugestivos precisam ser submetidos à avaliação neurológica imediata. Deverá ser feita uma tomografia computadorizada do crânio, sem contraste. Se a mesma não apresentar alterações, será repetida no período de dois a quatro dias, com contraste, após preparo hematológico. A ressonância magnética está indicada, se disponível. Se a segunda tomografia também não apresentar sinais de AVC, está indicada arteriografia, a ser realizada com os prévios cuidados necessários aos procedimentos com contrastes, nesses pacientes. O rastreamento para doença cérebro-vascular pelo doppler transcraniano deve ser iniciado aos anos de idade e repetido anualmente. Os casos com risco médio e alto devem receber acompanhamento e tratamento adequados. A exsanguineotransfusão parcial deve ser iniciada imediatamente após estabilização do paciente, para redução do nível de Hb S para menos de. Os pacientes devem ser monitorizados, devido à gravidade do quadro. A principal causa de aplasia medular eritrocítica na doença falciforme é a infecção pelo parvovírus, que acomete principalmente crianças na faixa etária de a anos, devido ao seu tropismo pelas células precursoras eritropoiéticas. O quadro inclui febre variável, palidez e fraqueza, podendo evoluir para falência cardíaca em consequência da acentuação da anemia. Observa-se redução acentuada dos níveis de hemoglobina e da contagem de reticulócitos. Trata-se de um evento transitório. O tratamento consiste na estabilização hemodinâmica pela transfusão de hemácias com monitorização até a elevação dos reticulócitos. O priapismo é a ereção persistente e dolorosa do pênis. Pode ocorrer na doença falciforme em todas as faixas etárias. O tratamento objetiva reverter as ereções indesejáveis, aliviar a dor e preservar a função sexual. Priapismo repetitivo: o tratamento é, frequentemente, domiciliar. São recomendados banhos mornos, hidratação oral, esvaziamento da bexiga, analgésicos e exercícios físicos. Caso não haja melhora em poucas horas, o paciente deverá ser hospitalizado, para hidratação e Priapismo agudo prolongado: as possíveis etiologias são infecções particularmente da próstata, traumas, medicações com efeitos no sistema nervoso central, abuso de álcool, uso de maconha ou após atividade sexual. O quadro exige imediata intervenção urológica: aspiração cavernosa de sangue com injeção de fenilefrina e, em última instância, o tratamento cirúrgico visando à drenagem sanguínea por *shunts* ou por operações abertas. Frequentemente é necessária sondagem vesical. A analgesia potente com opioide está indicada. Concentrados de hemácias estão indicados se não houver melhora nas primeiras horas, podendo ser necessária a analgesia venosa, sendo mais frequente após os anos de idade. Existem três formas clínicas: Ereção dolorosa prolongada, que não detumescer por mais de algumas horas. Tipicamente, pode permanecer por dias ou semanas. É seguida por impotência parcial ou completa. Persistente, frequentemente sem dor, com aumento do pênis ou endurecimento que persiste por semanas a anos. Esse padrão normalmente se desenvolve após ataque prolongado e está frequentemente associado à impotência parcial ou completa. Repetitivo: ereção dolorosa reversível, com detumescência ocorrendo em poucas horas.

Fonte: BRASIL. Ministério da Saúde. Secretaria de Atenção à Saúde. Departamento de Atenção Especializada. *Manual de eventos agudos em doença falciforme*. Brasília : Editora do Ministério da Saúde, 2009. (excerto)

### Manual Técnico

The majority of mild to moderate pain episodes abate spontaneously with bed rest, intake of extra fluids and analgesia and patients manage these at home using pain medications. In some cases there is increasing and excruciating pain which require medical intervention and hospitalization. Untreated complications can cause long term handicap and can prove fatal. A number of factors may precipitate a sickle cell crisis, these include: Hypoxia. Acidosis. Dehydration. Infection. Extreme Fatigue. Trauma. Temperature Changes (sudden). Stress / Anxiety. Increase physical / physiological demand (Pregnancy, physical exercise). Although these may precipitate a sickling crisis many occur with no known specific cause or identifiable precipitating factor. There are a number of short and long term complications associated with sickle cell disease and although many of these can be minimised with good self and medical management and activities to prevent and treat complications however long term and chronic damage to tissues and body organs is often inevitable and these shorten the life span of those with sickle cell disease. It is recommended that children with sickle cell disease have the full course of childhood immunisations; twice daily low dose penicillin is also recommended, folic acid may or may not be recommended, other specialist medication and immunisations are also advised. As part of routine care it is also recommended that children have Transcranial Doppler (TCD) scanning of the brain from the age of three years and every year thereafter; this is in order to identify narrowing of blood vessels and areas of small cerebral infarctions which are signs of Trans Ischaemic Attacks [TIA]) that may be missed because they have not resulted in an obviously debilitating stroke but in time is likely to cause a stroke. TCDs are done routinely in order to identify children who will benefit from having long term blood transfusion in an attempt to prevent the occurrence of a stroke and a consequent serious handicap. In adults the commonest cause of hospital admission is acute vaso-occlusive sickle cell crisis and the most common cause of death is sickle lung often precipitated by an acute chest infection. Chest infections may not be apparent radiologically until the condition worsens. Laboured or rapid respiration may be the only indication that is observed by a vigilant family member or nurse. This observation may be the only warning sign, which determines whether the patient receives the urgent medical attention required to prevent a fatality. It is always important for family members and carers to observe for mood changes, unresponsiveness or behaviour that appears unusual and take immediate action as this may indicate poor oxygenation of the brain increasing hypoxia or a stroke. For example, a patient who is normally compliant and co-operative may become unresponsive because of lack of oxygen to the brain. Leg Ulcers: Leg ulceration is one of the less well understood complications of SCD. It affects a proportion of older children and adults with sickle cell disease especially in the Caribbean and in tropical areas of the world. Leg ulcers are much less common in the UK and the USA but it does occur, especially in patients with less severe forms of sickle cell disease such as Hb. It accounts for a great deal of pain, absence from school and work and reduced physical activity. Leg ulcers are probably caused when minor cuts and abrasions fail to heal properly because of poor circulation to the lower limbs. Treatment consists of careful cleaning and dressing over a period of weeks or even months, but there remains a high risk of recurrence. Skin grafting may be required at a later stage. Ulcers are one of the few visible and unsightly complications of sickle cell disease, increasing the potential for stigmatization in those affected and leading to embarrassment and inhibition about dress and social activities especially because it is seen most commonly among teenagers and young adults; this being a vulnerable developmental phase of life. Priapism: Priapism is a persistent and abnormal erection of the penis accompanied by pain, tenderness and swelling. It may or may not be associated with sexual arousal, but unlike a normal sexual erection priapism is not relieved by sexual activity. Sickling causes priapism by occluding the circulatory spaces and small blood vessels of the penis and obstructing the drainage of blood from the corpus cavernosa. Episodes can last for hours, days or even weeks and can be among the most frightening and disturbing complications of sickle cell disease. Prolonged episodes can lead to partial, or complete, impotence. In extreme causes, exchange blood transfusion and even surgery may be needed to drain blood from the penis to relieve engorgement and pain. Early presentation and therapeutic intervention yield a successful outcome. Management of People with Sickle Cell Disease: There is no specific treatment for individuals with sickle cell disease; however, there are prophylactic measures that help to limit the factors that may trigger sickling crisis episodes and complications. There are national clinical guidelines for acute and community management of children and adults with sickle cell disease and these set out national standards that are measurable and will go some way to improving care. The day- care unit provide opportunities for patients to have rapid assessment and access to specialist care and is a useful mode of entry for those experiencing an uncomplicated non-emergency acute painful crisis. The units are often managed by highly advanced and skilled nurses, and may or may not include a specialist nurse, but in high prevalence areas they are managed by sickle cell / thalassaemia specialist nurses. On arrival in the day-care unit an initial vital signs, general observations and pain assessment is conducted by the nurse and the approved local policy pain medication is administered following which the patient will be assessed by the medical team and specific investigations and treatment is implemented. The patient is often managed quicker when attending the day care unit compared to attending A&E. Aggressive treatment of moderate to severe pain is consistent and the pain managed effectively.

Fonte: ANIONWU, E.; TANGAYI, S.; STREETLY, A. *Caring for people with sickle cell disease and thalassaemia syndromes: a framework for nursing staff*. Londres: Royal College of Nursing, 2011. (excerto)

### Cartilha

A doença falciforme é uma alteração genética, caracterizada por um tipo de hemoglobina mutante designada por hemoglobina S que provoca a distorção dos eritrócitos, fazendo-os tomar a forma de foice” ou meia-lua. O termo doença falciforme define as hemoglobinopatias nas quais pelo menos uma das hemoglobinas mutantes é a HbS. As doenças falciformes mais frequentes são a anemia falciforme, a S talassemia ou microdrepanocitose e as duplas heterozigoses Hb SC e Hb SD. Para o diagnóstico seguro de uma das três situações acima é de fundamental importância conhecer a forma de herança da doença falciforme. Na maioria dos casos, os pais de pacientes com doença falciforme são portadores assintomáticos dessa alteração genética. A situação mais comum se verifica quando dois portadores assintomáticos de falciforme, com patrimônio genético representado pela hemoglobina associada à hemoglobina S (Hb S) e cuja representação universal é Hb AS, se unem, constituindo uma prole. O exemplo a seguir mostra a probabilidade deste casal gerar filhos sem a doença falciforme, gerar portadores assintomáticos e com doença falciforme. A geração de uma pessoa com doença falciforme do tipo S Beta talassemia ou microdrepanocitose ocorre na seguinte situação: Da mesma forma, a geração de uma pessoa com doença falciforme de dupla heterozigose entre Hb S e um outro tipo de hemoglobina mutante, como por exemplo a Hb C, se verifica conforme situação abaixo: O portador assintomático de falciforme, também conhecido por portador do traço de Hb S ou heterozigoto para a Hb S, não é anêmico, não tem anormalidades físicas e tem uma vida normal. Os portadores de doença falciforme, por outro lado, podem apresentar sintomatologia importante e graves complicações. A Hb S tem uma característica química especial que em situações de ausência ou diminuição da tensão de oxigênio provoca a sua polimerização, alterando drasticamente a morfologia do eritrócito que adquire a forma de foice. Estes eritrócitos falcizados dificultam a circulação sanguínea provocando vaso-oclusão e infarto na área afetada. Conseqüentemente, esses problemas resultam em isquemia, dor, necrose e disfunções, bem como danos permanentes aos tecidos e órgãos além da hemólise crônica. Este processo fisiopatológico devido à presença de HbS é observado nas seguintes situações, em ordem decrescente de gravidade: anemia falciforme, HbS Beta talassemia, HbSC e Hb SD. É importante destacar que a frequência de portadores assintomáticos de falciforme no Brasil é de a, conforme a intensidade da população negra inserida na região. Como se sabe, a HbS teve origem no continente africano, e sua introdução no Brasil ocorreu notadamente durante o período da escravidão. Dados da Triagem Nacional sinalizam nascidos por ano com a doença falciforme e com o traço. O diagnóstico laboratorial da doença falciforme é realizado pela detecção da Hb S e da sua associação com outras frações. Assim, a técnica mais eficaz é a eletroforese de hemoglobina em acetato de celulose ou em agarose, em pH alcalino. A eletroforese em pH alcalino permite análises qualitativas e quantitativas das frações, conforme esquema abaixo: Observa-se que a amostra é um padrão normal, a amostra representa o portador assintomático de falciforme, a amostra a anemia falciforme, a amostra caracteriza o aumento de, a amostra exemplifica a dupla heterozigose entre a e a, e a amostra também é o exemplo de dupla heterozigose entre, que apresentam migrações similares em eletroforese alcalina. Justamente por causa deste último caso há necessidade de introduzir uma outra técnica eletroforética, que se realiza em agar ou agarose utilizando o pH ácido. Neste, a se separa da migrando na mesma posição que a, permitindo assim confirmar o diagnóstico de. Cabe ressaltar que é extremamente rara a presença de em nossa população, cuja heterozigose é prevalente numa relação de caso para cada mil pessoas analisadas, e a na ordem de caso para cada milhão. Assim, em geral, a ausência da eletroforese em pH ácido não traz dificuldades no diagnóstico da grande maioria dos casos de doença falciforme. Por outro lado, é necessário a dosagem de Hb Fetal pela desnaturação alcalina e da Hb por eluição especialmente para o caso. Em o Ministro da Saúde lançou o Programa Nacional de Triagem Neonatal por meio da Portaria Ministerial n.º 822. O Programa consiste em três fases: Fase 1 – Fenilcelonúria e Hipertiroidismo Congênito. Fase 2 – Inclui Hemoglobinopatias e Fase3 – Inclui Fibrose Cística. Este exame é realizado na primeira semana de vida da criança, em sangue total colhido do calcanhar (Teste do Pezinho). Finalmente, para um diagnóstico laboratorial completo é importante a realização do hemograma. A seguir, apresentamos um quadro sobre as principais características laboratoriais dos diferentes tipos de doença falciforme. As crises dolorosas são as complicações mais frequentes da doença falciforme e comumente constituem a sua primeira manifestação. Elas são causadas pelo dano tissular isquêmico secundário à obstrução do fluxo sanguíneo pelas hemácias falcizadas. A redução do fluxo sanguíneo causa hipóxia regional e acidose, que podem exacerbar o processo de falcização aumentando o dano isquêmico. Essas crises de dor duram normalmente de quatro a seis dias, podendo, às vezes, persistir por semanas. Hipóxia, infecção, febre, acidose, desidratação e exposição ao frio extremo podem precipitar as crises algicas. Os pacientes mais velhos citam que a depressão e exaustão física podem ser fatores precipitantes das crises. Os pacientes podem apresentar dor severa nas extremidades, abdômen e nas costas. A primeira manifestação de dor na maioria das crianças é a dactilite (ou síndrome mão-pé). Outras manifestações músculo-esqueléticas podem ser simétricas ou não, ou mesmo migratórias com eventual presença de aumento de volume, febre, eritema e calor local tornando, às vezes, difícil o diagnóstico diferencial com osteomielite, artrite séptica, sinovite e febre reumática. A dor abdominal pode simular abdômen agudo cirúrgico ou infeccioso, ou processos ginecológicos. É importante lembrar que, em crianças, as pneumonias, principalmente as de base, podem cursar com dor abdominal. Considerar como fatores de risco: febre maior que, desidratação, palidez, vômitos recorrentes, aumento de volume articular, dor abdominal, sintomas pulmonares agudos, sintomas neurológicos, priapismo, processos algicos que não se resolvem com analgésicos comuns. O tratamento consiste em eliminar os fatores precipitantes, garantir o repouso, assegurar uma boa hidratação (muitas vezes é necessário hidratação parenteral) e analgesia adequada.

Fonte: BRASIL. Ministério da Saúde. Secretaria de Atenção à Saúde. Departamento de Atenção Especializada. *Conduitas básicas na doença falciforme*. Brasília: Ministério da Saúde, 2006. (excerto)

### Cartilha

When no haemoglobin A is produced the condition is known as sickle beta zero thalassaemia, sometimes written as SBETA Thal. This behaves more like sickle cell anaemia. For more information about beta thalassaemia contact the Thalassaemia Society or one of the specialist Centres. SBETA Thal occurs if your child has inherited a haemoglobin gene from one parent and a persisting haemoglobin gene from the other parent. It is sometimes written, HbS/HPFH and very rarely causes any health problems. Sickle haemoglobin D disease: is a more unusual form of sickle cell disease and occurs if your child has inherited haemoglobin\_S from one parent and haemoglobin from the other parent, sometimes written, HbSD. This form can be as serious as sickle cell anaemia but could be milder. In this book we will mainly write about sickle cell disease as if it is one condition although we recognize that the different types are not all the same in the way they affect a person's health. Some are milder than others and children with the same sort of sickle cell disease may have different experiences of sickle cell disease and medical problems. Where there are obvious differences or where there are specific health problems, this will be made clear. Having a trait is commonly known as being a carrier and does not affect the person in any way. It is not a form of sickle cell disease and will never turn into sickle cell disease. If a person is born with a trait they are healthy and will always have a trait. Sickle cell trait means that a person has inherited one normal haemoglobin\_A gene and one sickle haemoglobin gene from their parents, sometimes written as HbAS. Other common haemoglobin traits seen in the include haemoglobin trait, where a person has inherited one normal haemoglobin and one unusual haemoglobin gene from their parents. The importance of knowing if one has sickle cell trait or any other unusual haemoglobin trait is because the unusual haemoglobin gene can be passed on to children through genetic inheritance. How this inheritance works is explained further on page 4. It is thought that the sickle cell haemoglobin gene was a small change (mutation) in the haemoglobin gene many thousands of years ago in countries where malaria was common. This is why we find haemoglobin in people whose ancestors come from Africa, Asia, the Middle and Far East and the Mediterranean. Sickle cell trait is found in approximately: It is also seen very rarely in White English people and other Northern Europeans because of intermarriage and subsequent gene mixing between different ethnic groups. It is important to realize that inheritance of sickle cell or any other unusual haemoglobin is not to do with the colour of a person's skin; it is to do with the genes they have inherited from their parents. In areas of the world where malaria is a danger people who have sickle cell trait were more likely to survive malaria than those who have the usual hemoglobin S. This advantage meant that haemoglobin was more likely to be passed down through the generations by those with sickle cell trait. Unfortunately those who inherited two of the sickle cell genes and have sickle cell anaemia were not protected from malaria, and were also likely to have the medical problems which are described later in this book. All the other unusual haemoglobin genes that have been described here probably also offer some protection against malaria. During the first three to six months of life, your child may not show signs of having sickle cell disease, because at birth there is a high level of baby haemoglobin, sometimes written, and a very low level of sickle the first year of life the haemoglobin level gets less as the child starts making more haemoglobin. Some children continue to make higher levels of haemoglobin even into adulthood. This can be checked by doing a special blood test. The longer your child goes on making haemoglobin the better, because it means he will be making less haemoglobin and is less likely to have many sickling crises. When a child is making a lot more Haemoglobin their red blood cells will not live in the circulation as long as cells that contain the usual Haemoglobin S. The body tries to keep up by making more red blood cells but it usually cannot keep up completely and your child becomes anaemic. Your child may look pale and the palm of the hand and the lips will be paler than your own. This sort of anaemia is known as a haemolytic anaemia and is not the same as the sort of anaemia caused by lack of iron. For this reason, iron tonics or medicines should not be given unless prescribed by your child's doctor. Folic acid tablets or medicine may be prescribed by your doctors because the body uses folic acid when making new red blood cells. In the United Kingdom, most children get enough folic acid from their normal diet and extra folic acid tablets are usually not necessary. When the red blood cells come to the end of their useful life they are broken down in the body. One of the substances that are produced is a yellow pigment called bilirubin. The liver clears the bilirubin from the body, but if there is a lot of bilirubin the liver may not be able to clear it all away and the yellow pigment may appear in the eyes, a condition known as jaundice. Some children may always have slightly yellowish eyes, even when they are well. Others may only become jaundiced when they are unwell, for example with coughs and colds or if they have a painful crisis. The spleen is an organ that lies on the left side of the tummy under the rib cage. One of the first things that your doctor may notice is that your child's spleen is big and can be felt just below the rib cage. The spleen helps to clear infection from the body and also clears up old or damaged blood cells.

Fonte: National Health Service (NHS). Sickle Cell and Thalassaemia Screening Programme. *Information for mums and dads: your baby carries a gene for unusual haemoglobin*. Londres: APoGI, 2010. (excerto)