
TERMINOLOGIA EM SÉRIE: *HOUSE M.D.*

TERMINOLOGY IN SERIES: HOUSE M.D.

Mila Bang*

Guilherme Fromm**

Resumo: o objetivo deste trabalho é apresentar uma pesquisa sobre as características terminológicas bilíngues de uma série televisiva denominada *House M.D.* Começamos apresentado o projeto geral a qual esta pesquisa está relacionada e explicamos o papel das séries médicas na televisão norte-americana. Na sequência, relacionamos os pontos teóricos da pesquisa, que são terminologia em séries de TV (FROMM, 2011) e a Teoria Comunicativa da Terminologia (de CABRÉ). A metodologia usada para a compilação do *corpus* é a da Linguística de *Corpus* e o programa usado para análise lexical é o Wordsmith Tools (SCOTT, 2008). O banco de dados de gerenciamento terminológico usado foi o VoTec (FROMM, 2007) e alguns dos resultados são apresentados neste artigo, enquanto outros estão disponíveis online.

Palavras-chave: Linguística de *Corpus*; Terminologia Bilingue; VoTec; Série de TV; House M.D.

Abstract: The aim of this article is to present a research about the bilingual terminological features of a TV series called *House M.D.* We start it presenting the general project the research belongs to and continue explaining the role of medical series in the American TV. In the sequence, we present the theoretical characteristics of the research, which are terminology in TV series (FROMM, 2011) and the Communicative Theory of Terminology (from CABRÉ). The methodology used to compile the corpus is from Corpus Linguistics and the program used in the lexical analysis is Wordsmith Tools (SCOTT, 2008). The terminological data bank manager program used was VoTec (FROMM, 2007) and some of the results are presented in this text, while others can be searched online.

Keywords: Corpus Linguistics; Bilingual Terminology; VoTec; TV Series; House M.D.

Introdução

Este artigo é resultado do desenvolvimento de um projeto de pesquisa cuja proposta de estudo se vale da Linguística de *Corpus* para analisar as legendas, tanto em inglês quanto em português, do seriado House M.D., a fim de criar um vocabulário técnico online (disponível em www.ic.voteconline.com.br) de alguns termos médicos

* Graduada em Letras pelo ILEEL/UFU; professora da Cultura Inglesa. E-mail: bangmila@gmail.com

** Professor doutor adjunto do ILEEL/UFU. E-mail: guifromm@ileel.ufu.br

utilizados no seriado. Este estudo faz parte de um projeto maior, intitulado Terminologia em Ficção, que pretende elaborar trabalhos semelhantes em séries de TV que tenham um caráter terminológico.

O objetivo inicial dessa pesquisa era tratar de questões de tradução, comparando as legendas em inglês com as legendas alternativas¹ em português para verificar as discrepâncias nas traduções. Porém, decidimos mudar esse objetivo, tendo em vista a grande quantidade de vocabulário técnico usado na série em questão, e montar um mini vocabulário de termos médicos na plataforma VoTec². Como este trabalho não pretende ser exaustivo, não foram inclusos nessa plataforma todos os candidatos a termos técnicos, apenas cinco – obedecendo à proposta do projeto maior. Acreditamos, também, que esse número seja suficiente para mostrar ao leitor como uma pesquisa terminológica com seriados de televisão pode ser desenvolvida.

1. House M.D. e os seriados médicos

Atualmente, os seriados americanos são um grande sucesso pelo mundo todo e, cada vez mais, fazem parte da diversão cotidiana dos telespectadores. A partir desse fato, surgiu o interesse de aliar, neste trabalho, a metodologia de pesquisa da Linguística de *Corpus* (LC) e a Teoria Comunicativa da Terminologia de Cabré na pesquisa terminológica de uma série em particular.

As legendas de House M.D. foram escolhidas como objeto de estudo não só pela fama que o seriado conquistou em escala mundial, mas também, e principalmente, por utilizar frequentemente a terminologia da área da medicina. Apesar de se tratar de ficção, o seriado lança mão de vários termos específicos médicos para alcançar certa verossimilhança e convencer o espectador.

Em 2004, o seriado em questão foi exibido pela primeira vez, dividido em vinte e dois episódios, incluindo o episódio piloto. Até o seu episódio final (totalizando oito temporadas), exibido em 2012, foi assistido por telespectadores de diferentes partes do mundo, consolidando, assim, o sucesso da série. As séries médicas tornaram-se muito populares talvez por retratarem uma realidade que não é comum aos indivíduos. Estes vão ao hospital quando estão enfermos, mas, geralmente, não sabem como é o

¹ Feitas por tradutores amadores.

² Vocabulário Técnico Online (FROMM, 2007).

funcionamento de um hospital, como é a rotina e o cotidiano das pessoas que trabalham para salvar vidas. Esse pode ser um dos motivos pelo qual o universo da medicina é tão retratado, causando certo fascínio nos telespectadores.

Desde a década de 60, já existiam outros seriados médicos de grande sucesso, tais como *Ben Casey* e *Dr Kildare*. Além desses, nas últimas décadas foram produzidas também outras séries com a mesma temática, como *ER – Plantão Médico*, *Grey's Anatomy*, *Scrubs* e *Nip/Tuck*. Apesar de todas mostrarem o cotidiano de um hospital ou de uma clínica, cada um desses seriados possui um tom diferente, seja cômico ou dramático.

O seriado House M.D., ou simplesmente Dr. House, é uma série estadunidense cujo protagonista, Gregory House, é interpretado pelo ator inglês Hugh Laurie. A série foi criada por David Shore e tem algumas características que foram inspiradas no personagem Sherlock Holmes, de Sir Arthur Conan Doyle, pois o autor da série considera-se um grande fã dessa figura da literatura. Algumas dessas características incluem desde elementos básicos, como o número do apartamento de ambos, 221B, até elementos fundamentais para a série, como o método pelo qual House consegue resolver casos médicos.

Dr. House foi distribuída em 66 países diferentes e considerado o programa mais seguido no mundo no ano de 2008. A série foi nomeada a vinte e três *Emmy Awards* e premiada quatro vezes, incluindo o prêmio de criador e produtor executivo para David Shore. Além disso, recebeu, também, o prêmio *Humanitas* pelo episódio “*Three stories*”, além do Globo de Ouro para a categoria de melhor desempenho em um seriado de televisão para Hugh Laurie, dentre outros.

Gregory House é um infectologista e nefrologista brilhante, que consegue elaborar excelentes diagnósticos médicos. Para tal, ele se utiliza de métodos não muito convencionais, o que faz com que haja um constante conflito entre ele e seus colegas de trabalho, bem como entre ele e seus pacientes. Frequentemente, o personagem de Hugh Laurie é caracterizado como antissocial, cético, misantropo, mal-humorado, cínico e mesquinho. Apesar de tais características, House é um médico muito respeitado, já que seus instintos e sua habilidade de observar e deduzir o levam a diagnosticar doenças em casos difíceis e já desacreditados, que nenhum outro médico conseguiu resolver.

Esse personagem destaca-se não só pela sua inteligência impressionante, mas também pela sua estranha conduta como profissional. É um médico que evita ao máximo se aproximar de seus pacientes e que não confia nas pessoas, pois de acordo com o seu discurso, “todo mundo mente” (tradução livre de “*Everybody lies*”). House M.D. foi, por muitos anos, um dos seriados de maior audiência nos Estados Unidos; essa popularidade se deve a vários fatores, mas principalmente ao fato de que o rol de defeitos do protagonista é, provavelmente, o que o torna tão humano e genial, constituindo, assim, um dos personagens mais interessantes da televisão.

A maioria dos capítulos inicia-se com uma cena fora do hospital fictício no qual a série se passa, o Hospital Escola Princeton Plainsboro. Essa é a técnica *cold open*, ou mais popularmente conhecida como *teaser*, que costuma ser utilizada para envolver o telespectador com a trama de forma com que faça que ele não perca o interesse pelo programa. Nessas cenas, são mostradas pessoas anônimas que terão sintomas de doenças que serão tratadas posteriormente pela equipe do médico Gregory House. O seriado é apenas ficção, ou seja, não mostra fatos ou casos reais, mas em House, M.D. são retratadas doenças que realmente existem.

2. Revisão da Literatura

A terminologia, cada vez mais, permeia o dia a dia de todos. Todas as áreas de conhecimento e todos os fazeres possuem palavras que são usadas no intuito de facilitar a comunicação daqueles que trabalham dentro dessas áreas. Do astrônomo ao zoólogo, do catador de lixo ao jardineiro, todos usam termos específicos dentro de suas profissões. Esta tendência acaba por migrar para as séries de televisão, que retratam os termos para fornecer uma maior verossimilhança aos seus personagens ficticiais. De um modo geral, o uso da terminologia se destaca em três tipos de séries de televisão (FROMM, 2011):

a. séries com terminologia totalmente ficcional (podem ser procedimentais³ ou serializadas⁴): figuram neste item os programas sobre o fantástico, com personagens e características que existem em mundos criados a partir da imaginação. Exemplos clássicos são as séries sobre monstros, fantasmas, espíritos, mitologias, etc. Aqui, mais do que uma terminologia propriamente dita, temos uma formatação terminológica: os termos não são reais, mas o modo de construir suas definições o são;

b. séries que misturam ficção e ciência (tendem a ser serializadas): campo propício para a ficção científica, nesses programas encontramos terminologia própria de áreas reais como Física, Química, Astronáutica, Astronomia, Engenharias, Biologia, etc. No entanto, como a ficção também se faz presente, termos são criados, dentro dessas áreas, para explicar fenômenos que ainda não podem ser explicados ou, ainda, fenômenos inventados. Essas séries, assim como as do item c, contam com consultores (MARTHE, 2011), advindos dessas respectivas áreas técnicas, que fornecem termos reais (ou, pelo menos, com “aura” de reais) para melhor situarem os enredos;

c. séries que retratam cotidiano profissionais de médicos, investigadores, cientistas forenses, etc. (tendem a ser procedimentais): nesse caso, as terminologias usadas tentam retratar, de forma fiel, aquelas usadas pelos profissionais dessas áreas nos seus afazeres cotidianos.

House M.D., portanto, se encaixaria na opção que retrata o cotidiano de profissionais. Apesar de ser uma trama com um enredo que também envolve relações interpessoais e outros assuntos que não estão relacionados à medicina, esse seriado foi escolhido por ter um roteiro rico em termos médicos, já que é mais especializado (em termos quantitativos) que os outros seriados com essa mesma temática. Como a maior parte do enredo é baseada em investigações médicas, muitos termos específicos da área da medicina, como nomes de remédios, exames, equipamentos, são usados frequentemente. Além disso, médicos e enfermeiras conferem o que foi escrito pelos

³ Séries procedimentais são aquelas que apresentam uma independência entre os episódios; se o telespectador perder algum, não há prejuízo para o entendimento daquele episódio em específico que está assistindo (<http://www.vayatele.com/diccionario/diccionario-telefilo-procedimental>).

⁴ Séries serializadas são aquelas que demandam uma maior atenção por parte do telespectador. A trama se desenrola por toda uma temporada da série, ou seja, perder um episódio significa perder parte da compreensão geral (<http://locoxelcine.com/series-de-tv/terminologia-de-las-series/>).

roteiristas, o que faz com que “as terminologias usadas retratem, de forma fiel, àquelas usadas pelos profissionais dessas áreas nos seus afazeres cotidianos” (FROMM, 2011).

Como uma das características dos seriados de cotidianos de profissionais é a explicitação dos termos da área em destaque, já que a verossimilhança é necessária para a construção de um bom enredo, é importante definir o que consideramos um termo. Seguimos, aqui, as ideias propostas por Maria Tereza Cabré na sua Teoria Comunicativa de Terminologia (TCT). Segundo Krieger e Finatto (2004, p. 35), comentando Cabré,

De acordo com o princípio comunicativo, uma unidade lexical pode assumir o caráter de termo em função de seu uso em um contexto e situação determinados. Consequentemente, o conteúdo de um termo não é fixo, mas relativo, variando conforme o cenário comunicativo em que se inscreve. Tais proposições levam a TCT a postular que a priori não há termos, nem palavras, mas somente unidades lexicais, tendo em vista que estas adquirem estatuto terminológico no âmbito das comunicações especializadas.

Acreditamos que House M.D., mais do que qualquer outra série já exibida na televisão que retrate o cotidiano de profissionais diversos, foi um destaque na exposição terminológica. Os telespectadores aprenderam, através dos dramas mostrados em cada episódio, nomes de doenças, processos de diagnóstico, como funcionam equipamentos hospitalares, a farmacologia disponível para o tratamento de enfermidades, etc. De um modo geral, esse seriado seria um protótipo (talvez, compartilhando o posto com *Jornada nas Estrelas*) de como empregar terminologia real em um produto ficcional.

3. Metodologia de pesquisa

A metodologia da Linguística de *Corpus* nos possibilitou fazer a análise da série através de suas legendas em inglês e em português. De acordo com Berber Sardinha, a Linguística de *Corpus*

ocupa-se da coleta e da exploração de *corpora*, ou conjuntos de dados linguísticos textuais coletados criteriosamente com o propósito de servirem para a pesquisa de uma língua ou variedade linguística. Como tal, dedica-se à

exploração da linguagem por meio de evidências empíricas, extraídas por computador. (2004, p. 3)

Como a Linguística de *Corpus* trabalha com uma abordagem empirista, é muito útil na análise de sistemas linguísticos, levando em consideração os contextos nos quais os falantes os utilizam. Portanto, percebe-se que é extremamente importante formar um *corpus* com amostras de linguagem natural usada em situações reais e autênticas.

Ainda de acordo com Sardinha (2004), há quatro pré-requisitos para a formação de um *corpus*: i) o *corpus* deve ser resultado da compilação de textos autênticos, em linguagem natural, não podendo ter sido produzidos para um objetivo específico de pesquisa ou em linguagem artificial; ii) os textos devem ter sido escritos por falantes nativos de certa língua; iii) os textos componentes do *corpus* têm que ser escolhidos de forma criteriosa, levando em consideração as condições de naturalidade e autenticidade, dentre outros; iv) os textos devem ser representativos de uma variedade linguística ou de um idioma.

Além desses pré-requisitos, um *corpus* deve ser o maior possível, a variedade deve ser limitada ao máximo e o conteúdo deve ser restrito a apenas um autor para que o *corpus* possua uma função representativa. Por isso, neste trabalho, compilamos as legendas de sete das oito temporadas da série House M.D., para compor um *corpus* representativo desse tipo de vocabulário específico.

3.1. Compilação de corpora

Para realizar este trabalho, foi necessário compilar um *corpus* que nos serviu como objeto de estudo. O *corpus*, tanto em inglês quanto em português, é composto por legendas do seriado House, M.D. que foram baixadas de um sítio eletrônico⁵ que disponibiliza legendas para vários seriados. É necessário ressaltar que as legendas traduzidas para o português utilizadas nesta pesquisa foram feitas por tradutores amadores, ou seja, não utilizamos as legendas oficiais do seriado. Decidimos trabalhar com essas legendas alternativas por questões práticas: além de ser possível fazer o *download* desse material em minutos, outra vantagem, “em termos de trabalho com

⁵Legendas TV: www.legendas.tv

programas de Análise Lexical, é que essas legendas já vêm no formato .txt (ou .srt, facilmente convertidas em .txt), melhor lido por esses programas” (FROMM, 2011).

Após a compilação, as legendas foram salvas em pastas diferentes, sendo uma para cada temporada, como podemos ver na figura abaixo.

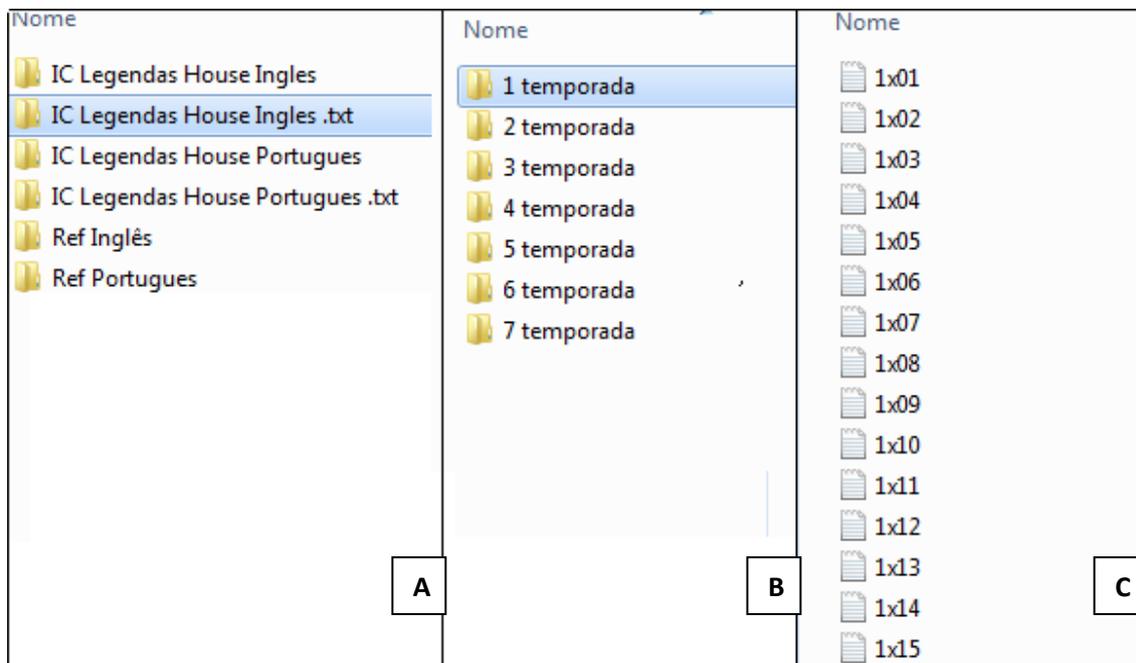


Figura 1. Diretórios com as legendas de House.

A figura 1A representa o diretório geral, no qual foram salvas todas as legendas, em inglês e em português, bem como o *corpus* de referência (usado para o levantamento das palavras-chave) também nas duas línguas. A figura 1B representa as subpastas nas quais foram salvas as legendas dos episódios de cada temporada separadamente. A figura 1C representa o diretório com todas as legendas da primeira temporada.

Depois de organizar as pastas, pensamos ser necessário editar as legendas, já que todas têm marcação de tempo antes das falas, como pode ser visto na figura 2.

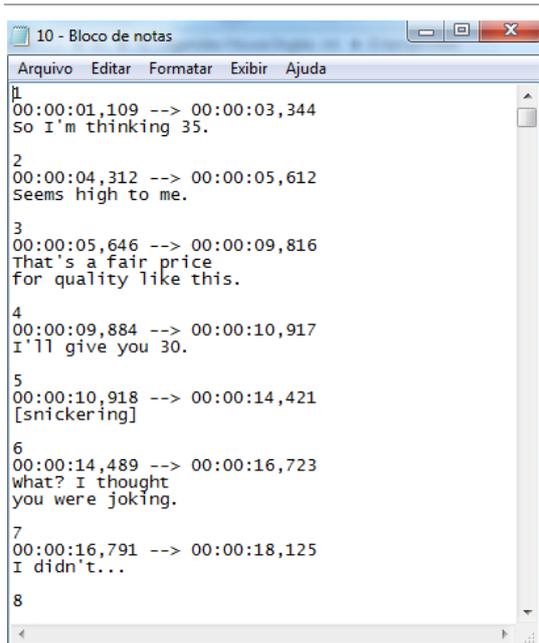


Figura 2. Marcação de tempo nas legendas.

Porém, um teste foi feito para verificar se essas marcações realmente influenciavam ou não no resultado das listas de palavras. Com isso, foi possível identificar que os números da marcação aparecem nas listas de palavras como erro (o programa de análise lexical, por padrão, não inclui números em suas listagens) e, conseqüentemente, não alteram o resultado das listas de palavras. Assim, não foi necessário continuar com a limpeza das legendas.

Em seguida, começamos a trabalhar com o programa computacional *Wordsmith Tools* (SCOTT, 2008). Esse programa é fácil de usar, prático e oferece recursos que podem facilitar o trabalho dos estudiosos da linguagem. Uma vez compilado o *corpus*, é possível obter todas as informações necessárias para a análise do mesmo.

Seguindo a proposta de Teixeira (2008), podemos classificar o nosso *corpus* da seguinte maneira:

Tabela 1. *Corpus* House M.D.

Modo	Escrito (legendas disponíveis na Internet)
Tempo	Sincrônico (legendas de 2004 - 2011)
Seleção	Amostragem (linguagem de textos da área de ficção) Estático (seleção não renovável)
Conteúdo	Especializado (legendas de uma série de cotidiano profissional)
Autoria	Língua nativa (inglês americano) e tradução em português
Disposição Interna	Paralelo (original e traduções)
Finalidade	Estudo (análise terminológica)

3.2. Análise do corpora e delimitação dos candidatos a termos

O programa que utilizamos para analisar o *corpus* deste trabalho foi o *Wordsmith Tools*. Esse programa de análise lexical possui três ferramentas: 1) *Wordlist*: elabora uma listagem de palavras; 2) *Keywords*: através da análise contrastiva do *corpus* de estudo com um *corpus* de referência geral, obtêm-se a lista de palavras-chave; 3) *Concord*: exibe as linhas de concordância (os contextos nos quais os candidatos a termos estão inseridos) numa tela *KWIK* (palavras chave em contexto).

Ao selecionar as legendas de House na ferramenta *Wordlist* (a elaboração da listagem de palavras é procedimento básico para a aplicação das outras duas ferramentas), optamos pelo uso de uma *stop list*⁶. Este procedimento já eliminou, nas listagens em inglês e português, todas as palavras gramaticais⁷ das duas línguas.

Depois de criar as listas de palavras, utilizamos a ferramenta *Keywords* para verificar quais são as palavras-chave do *corpus* de estudo – as possíveis candidatas a se tornarem termos a serem analisados. Para isso, comparamos o nosso *corpus* com um *corpus* de referência⁸, que é sempre maior. Isso porque, de acordo com Berber Sardinha (1999),

O tamanho do *corpus* de referência influencia a quantidade de palavras chave obtidas. Os tamanhos críticos de corpora de referência são 2, 3 e 5 vezes o tamanho do *corpus* de estudo. Corpora de referência com estas dimensões retornam significativamente mais palavras chave do que corpora de tamanhos menores. (p. 10, 11)

Feita essa comparação, obtivemos os candidatos a termos do seriado House, como podemos ver (parcialmente, já que esta lista tem quinhentas linhas) na figura abaixo.

⁶ Relação de palavras que não serão analisadas, pelo programa de análise lexical, quando da elaboração de suas listagens.

⁷ Pronomes, conjunções, numerais, interjeições, artigos, preposições e grande parte dos advérbios.

⁸ Neste trabalho, utilizamos o ANC (*American National Corpus*) e o BP (Banco do Português).

N	Key word	Freq	%	RC. Freq	RC. %	Keyness	Lemmas	Set
1	#	1,047,086	67.49	0		6,174,346.00	0.0000000000	
2	TO BE	20,239	1.11	0		105,551.34	0.0000000000	
3	GONNA	2,616	0.14	3,813	0.02	5,517.51	0.0000000000	
4	SHE'S	1,893	0.10	5,156	0.02	2,445.47	0.0000000000	
5	PATIENT	1,062	0.06	2,239		1,725.62	0.0000000000	
6	DR	883	0.05	1,734		1,517.08	0.0000000000	
7	PAIN	778	0.04	1,549		1,322.56	0.0000000000	
8	WRONG	1,176	0.06	3,934	0.02	1,213.94	0.0000000000	
9	BRAIN	902	0.05	2,402	0.01	1,191.14	0.0000000000	
10	BLOOD	1,086	0.06	3,706	0.02	1,094.36	0.0000000000	
11	BIOPSY	332	0.02	228		1,006.29	0.0000000000	
12	HEART	973	0.05	3,287	0.01	992.56	0.0000000000	
13	À	492		0		947.34	0.0000000000	
14	LUNGS	306	0.02	207		932.67	0.0000000000	
15	SICK	528	0.03	1,032		910.35	0.0000000000	
16	GOTTA	377	0.02	441		901.23	0.0000000000	
17	DOCTOR	556	0.03	1,220		874.20	0.0000000000	
18	LIVER	475	0.03	859		867.47	0.0000000000	
19	SYMPTOMS	450	0.02	773		852.19	0.0000000000	
20	FEVER	286	0.02	266		765.08	0.0000000000	
21	MRI	212	0.01	90		749.38	0.0000000000	
22	SYMPTOM	232	0.01	159		703.71	0.0000000000	

Figura 3. Palavras-chave do seriado *House M.D.*, na língua inglesa.

Assim que fizemos as listas de palavras-chave, selecionamos alguns termos específicos para serem inseridos no banco de dados posteriormente. Dois critérios foram adotados para fazer essa escolha: 1) ser um termo específico da área médica; 2) existir tanto nas legendas em inglês quanto nas legendas em português. Como o vocabulário online é bilíngue, o segundo critério foi excludente, pois se um termo não existe nas duas línguas, não é possível obter contextos⁹ para criar as definições em português e em inglês.

Com os termos escolhidos, partimos para a utilização da ferramenta *Concord*. Essa ferramenta permite que tenhamos acesso à listagem das ocorrências de um termo e ao texto que está ao seu redor. Assim, conseguimos material linguístico o suficiente para elaborar uma definição e preencher a microestrutura do verbete quando da sua inserção no VoTec. Na figura 4, temos o exemplo do termo *clot* (coágulo, em português):

⁹ A proposta do VoTec é a criação da microestrutura dos termos única e exclusivamente a partir das informações advindas dos exemplos retirados do *corpus* (FROMM, 2007).

The screenshot shows the Concord software interface with a concordance table for the term 'clot'. The table has columns for N, Concordance, Set, Tag, Word #, Sen, Para, Para, lead, lead, Sec, Sec, File, and %. The concordance lines are numbered 1 through 22, each showing a snippet of text from a source file and the percentage of occurrences. The text snippets include phrases like 'The clot must have broken up', 'Maybe the wound threw a clot', 'Not a clot. Maybe a lipoma?', 'Angio showed a clot', 'There's the clot', 'and pull the clot out', 'patient's blood won't clot', 'And whose blood won't clot', 'To break down, clot off', 'Hanna, you've got a clot in your lung', 'I had a blood clot, and the muscle was dying', 'get any oxygen unless we can bust the clot', 'It's not a clot', 'It's not busting the clot', 'confirmed the spot on Julia's lung was a clot', 'I highly doubt it's more relevant than the clot', 'What about a blood clot?', 'The rough flow causes a clot, explains the heart and head', 'He had a g.I. Infarction caused by a clot', 'The clot is a very bad sign', and 'Clot in his brain could kill one of them'.

N	Concordance	Set	Tag	Word #	Sen	Sen	Para	Para	lead	lead	Sec	Sec	File	%
1	284 00:10:51,741 -> 00:10:52,708 The clot must have broken up 285 00:10:52,			2,663	22537%		026%				026%		6.txt	52%
2	00:09:21,458 Maybe the wound threw a clot. 153 00:09:21,525 -> 00:09:22,659			1,365	13800%		021%				021%		22.txt	21%
3	00:09:31,168 -> 00:09:32,635 Not a clot. Maybe a lipoma? 157 00:09:32,			1,392	14200%		021%				021%		22.txt	22%
4	:58,461 or go into his head and get the clot out. 513 00:25:58,546 -> 00:26:00,			4,845	46796%		053%				053%		15.txt	64%
5	,282 -> 00:22:04,565 Angio showed a clot 444 00:22:04,616 -> 00:22:06,567			4,207	40443%		055%				055%		15.txt	55%
6	:28:32,566 -> 00:28:33,766 There's the clot. 577 00:28:38,939 -> 00:28:41,274			5,439	52600%		071%				071%		15.txt	71%
7	:27:02,009 -> 00:27:05,478 and pull the clot out. 538 00:27:05,529 -> 00:27:07,			5,073	49235%		056%				056%		15.txt	66%
8	-> 00:10:56,145 patient's blood won't clot. 250 00:10:56,212 -> 00:10:57,312			2,494	17600%		027%				027%		7.txt	27%
9	-> 00:14:45,007 And whose blood won't clot. 352 00:14:45,074 -> 00:14:46,675			3,534	24000%		039%				039%		7.txt	39%
10	54,105 -> 00:28:56,440 To break down, clot off. 672 00:28:56,507 -> 00:28:59,			6,417	55596%		073%				073%		6x04.txt	73%
11	-> 00:35:47,108 Hanna, you've got a clot in your lung. 624 00:35:47,176 ->			5,803	64477%		039%				039%		21.txt	89%
12	:01,095 -> 00:29:05,999 I had a blood clot, and the muscle was dying. 541 00:			4,994	56157%		077%				077%		21.txt	77%
13	get any oxygen unless we can bust the clot. 628 00:35:55,584 -> 00:35:57,885			5,845	64900%		030%				030%		21.txt	90%
14	00:36:32,417 -> 00:36:35,285 It's not a clot. 635 00:36:35,353 -> 00:36:37,287			5,909	65700%		031%				031%		21.txt	91%
15	166 -> 00:36:13,534 It's not busting the clot. 632 00:36:13,601 -> 00:36:15,701			5,888	65400%		031%				031%		21.txt	91%
16	confirmed the spot on Julia's lung was a clot. 535 00:23:11,398 -> 00:23:13,733			5,163	45300%		056%				056%		18.txt	56%
17	I highly doubt it's more relevant than the clot. 538 00:23:17,738 -> 00:23:19,839			5,197	45500%		057%				057%		18.txt	56%
18	-> 00:14:24,491 What about a blood clot? 332 00:14:24,558 -> 00:14:26,292			3,121	28400%		041%				041%		15.txt	41%
19	00:11:40,931 The rough flow causes a clot, explains the heart and head. 291			2,714	23847%		029%				029%		12.txt	28%
20	421 He had a g.I. Infarction caused by a clot 681 00:25:50,488 -> 00:25:52,723			6,201	56540%		058%				058%		10.txt	68%
21	718 00:27:07,265 -> 00:27:08,932 The clot is a very bad sign. 719 00:27:09,			6,545	59558%		071%				071%		10.txt	72%
22	him. 541 00:24:45,194 -> 00:24:47,729 Clot in his brain could kill one of them			5,486	47627%		059%				059%		5x24.txt	69%

Figura 4. Linhas de concordância para o termo *clot*.

A partir dessas linhas de concordância, selecionamos alguns contextos a fim de nos ajudar na criação de definições para os termos. Nas legendas do seriado *House*, não conseguimos identificar nenhum contexto definitório, isto é, nenhum dos contextos dos termos escolhidos era seguido do verbo *ser*, nas legendas em português, e *to be*, nas legendas em inglês¹⁰. Sendo assim, foi necessário trabalhar com os contextos explicativos (AUBERT, 1996). Geralmente, quando o médico Gregory House está trabalhando para diagnosticar um paciente, a sua equipe dá várias sugestões de doenças de acordo com os sintomas sentidos pelo paciente. Assim, é possível criar definições a partir de explicações e não de definições. Na figura 5, temos o termo *seizure*. Para conseguir criar uma definição, tivemos que trabalhar com hipóteses e deduções feitas pelos médicos da série já que, em momento algum, o termo *seizure* aparece definido.

¹⁰ Os contextos definitórios costuma ser seguidos por esses verbos e/ou vírgulas, dois pontos, parênteses. Esses contextos já apresentariam uma definição pronta.

Exemplo	Conceito
“And now you are talking about a seizure disorder. Which is also not a toxin. But which is most likely caused by an old injury to the brain.” / “Seizure right after a fight obviously suggests concussion.”	A reaction that can be caused by an injury to the brain or by concussions.
“But a hemispherectomy would completely stop the right-brain seizure activity and he would no longer need to take his anticonvulsant medication.”	It can be treated by a hemispherectomy.
“I drink any more, Im gonna explode. / You are not going to explode. You are just gonna have a seizure, which will prove that your kidneys are not working...”	It is a reaction caused when your kidneys are not working properly.
“Serious bacterial food poisoning explains seizure and vomiting.”	It is a reaction which can be caused by bacterial food poisoning.
“The brain is too fried during a seizure to form memories.”	It is a state in which the brain is incapable of holding memories.

Figura 5. Exemplo de conceitos primários criados a partir de contextos explicativos.

4. Criação de um vocabulário técnico online: a plataforma VoTec

Depois da coleta dos contextos, partimos para a criação do vocabulário e a inserção dos termos na plataforma VoTec¹¹. Na figura 6, temos uma lista de termos ainda não aprovados pelo administrador do banco de dados.

The screenshot shows the VoTec web application interface. At the top, it says 'Vocabulário Técnico Online' and 'Tela Cheia | English'. Below that, it says 'Bem Vindo, Mila Bang'. There is a message: 'Seu último acesso foi em: 21/11/2011 19:05'. Below this are four buttons: 'Trocar Senha', 'Novo Termo', 'Cadastro de Fontes', and 'Sair'. The main section is titled 'Fichas ainda não aprovadas' and contains a table with the following data:

Termo	Língua	Coletado em	Cadastrada por	Ações
Coágulo		22/10/2011	Mila Bang	editar
Clot		22/10/2011	Mila Bang	editar
Convulsão		02/10/2011	Mila Bang	editar
Seizure		02/10/2011	Mila Bang	editar
Antibiótico		25/09/2011	Mila Bang	editar
Antibiotic		25/09/2011	Mila Bang	editar

Figura 6. Página inicial do VoTec (acesso restrito).

¹¹ Essa plataforma foi desenvolvida por Fromm (2007) para ser utilizada em sua tese de Doutorado. Porém, como o projeto desenvolvido é de Iniciação Científica, um novo site foi criado especialmente para os orientandos que trabalham com Terminologia em Ficção.

Para cadastrar um termo, basta clicar em “Novo Termo”. Acessando um termo, temos uma tela na qual cadastramos os contextos encontrados nas legendas para esse termo em análise. A partir desses contextos, são criados conceitos primários. Na figura 7, podemos ver os contextos retirados das linhas de concordância e os conceitos elaborados para o termo *seizure*.

Vocabulário Técnico Online Tela Cheia | English

Cadastro de Contextos para Seizure

[Voltar ao Painel](#) [Próximo Passo](#)

Novo Contexto

Passo 2

Dados do Contexto

Exemplo*:

Conceito*:

Fonte*: [Cadastrar Nova](#)

Fonte

Data de Coleta*: (dia/mês/ano ex.: 18/03/2007)

Contextos Cadastrados

Exemplo	Conceito	Fonte	Ações
"And now you are talking about a seizure disorder. Which is also not a toxin. But which is most likely caused by an old injury to the brain." / "Seizure right after a fight obviously suggests concussion."	A reaction that can be caused by an injury to the brain or by concussions.	LTV 02/10/2011	editar - excluir
"But a hemispherectomy would completely stop the right-brain seizure activity and he would no longer need to take his anticonvulsant medication."	It can be treated by a hemispherectomy.	LTV 02/10/2011	editar - excluir
"I drink any more, Im gonna explode. / You are not going to explode. You are just gonna have a seizure, which will prove that your kidneys are not working..."	It is a reaction caused when your kidneys are not working properly.	LTV 02/10/2011	editar - excluir
"Serious bacterial food poisoning explains seizure and vomiting."	It is a reaction which can be caused by bacterial food poisoning.	LTV 02/10/2011	editar - excluir
"The brain is too fried during a seizure to form memories."	It is a state in which the brain is incapable of holding memories.	LTV 02/10/2011	editar - excluir
"It is Whipples Disease. Explains the seizure and the stomach pain."	It is a symptom caused by the Whipples disease.	LTV 02/10/2011	editar - excluir
"I need some help in here! Your patient is seizing. / Why would he have a seizure? / Because he has cancer!"	It is one of the symptoms of patients who have cancer.	LTV 02/10/2011	editar - excluir

Contextos Cadastrados: 7

03/12/2011 19:52 © 2007 FFLCH - ICMC Jr.

Figura 7. Contextos e conceitos do termo *seizure*.

Na tela seguinte, começa o trabalho de preenchimento das abas destacadas em vermelho na figura 8. Os campos preenchidos constituirão a microestrutura do verbete.

Vocabulário Técnico Online Tela Cheia | English

Termo: Amyloidosis

Passo Anterior Salvar Sair sem salvar

Contextos **Conceito Final / Definição**

Contextos

	Exemplo	Conceito	Fonte
1	“Could be amyloidosis. Amyloid buildup hits the organs, tissue... her muscles would not have atrophied.”	It is a disease which does not cause the muscles to atrophy.	LTV 22/10/2011
2	“The problem was not the surgery. Abnormally deposited proteins. It is amyloidosis. That would explain the blinking and the low heart rate.”	It is a disease that drops the heart rate and may cause constant blinking.	LTV 22/10/2011
3	“Yes. Can I go test that theory now? / Waste of time. He needs a bone marrow transplant for the amyloidosis.”	It can be treated with a bone marrow transplant.	LTV 22/10/2011

Dados **Traços Distintivos** Semântica Termo Equivalente Termos Remissivos Informações Enciclopédicas

Traços Distintivos

Nova Coluna – Não use aspas duplas

	A	B	C	D	E	F	G	H	I	J	K
1	disease	do not	cause	muscles	atrophy						
2	disease					drops	heart rate	constant b...			
3	disease								treated	transplant	bone ma
4	disease		cause				falling he...				
5	disease		cause				kidney fai...				
6	disease		cause				seizures				

Figura 8. Preenchimento das abas.

Nem todos os campos de todas as abas serão preenchidos, pois os contextos nem sempre fornecem todas as informações. Neste trabalho, só não preenchemos a aba *Termos Remissivos*¹². Na primeira aba, *Dados*, informamos as características morfossintáticas do termo e também a posição que ocupa no *corpus*. Na aba *Traços Distintivos* (em destaque na figura 8), inserimos os semas retirados dos conceitos previamente elaborados (na horizontal – cada número de linha representa um exemplo cadastrado no banco de dados), procurando relacioná-los (na vertical – semas idênticos ou sinônimos). As informações disponíveis nessa aba é que nos ajudam a criar o conceito/definição final do termo.

¹² Por se tratar da elaboração de apenas cinco termos, não há condições de montar remissivas dentro do trabalho.

Na aba *Semântica*, é possível informar se o termo é dicionarizado ou não (porém, essa informação não é disponibilizada na página de acesso livre da plataforma por questão de direitos autorais); além disso, é possível, também, estabelecer relações semânticas presentes nos exemplos (hiponímia, hiperonímia, co-hiponímia, antonímia e sinonímia). A aba seguinte, *Termo Equivalente*, nos permite ligar o termo em inglês ao termo equivalente em português. Na aba *Informações Enciclopédicas*, inserimos o conceito do termo analisado encontrado em enciclopédias online que não exigem direitos autorais, como a *Wikipedia*.

Após inserirmos todas essas informações, criamos um conceito final (figura 9), baseado na Análise Componential (ILARI,2003) da aba *Traços Distintivos*. A diferença entre o Conceito Final (elaboração primária a partir dos semas) e a Definição (resultado final, disponibilizado na página de consulta) é o grau de elaboração que o pesquisador pretende dar ao produto. No presente caso, optamos por não alterar as informações dentro dos campos.

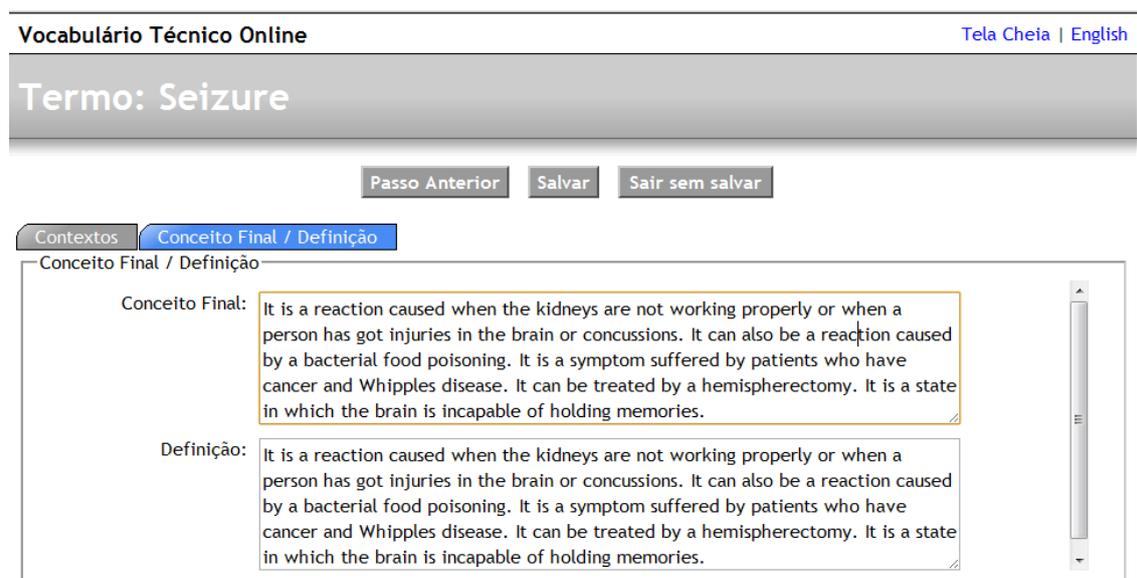


Figura 9. Conceito final do termo *seizure*.

5. Resultado

Ao terminar de executar todos os procedimentos descritos acima e depois de receber a aprovação do administrador, o termo ficará disponível no VoTec. Para que o consulente possa procurar um termo, é necessário selecionar a área desejada

(Terminologia em Ficção), e depois escolher o seriado (House), como apresentado na figura 10.

The screenshot shows the 'Vocabulário Técnico Online' interface. At the top, it says 'Terminologia em Ficção' and 'House' are selected in dropdown menus. Below the search bar, there are options for 'Tipos de Exibição' (Normal, Descritiva) and 'Tipos de Consulta' (Total, Tradutor, Modular). On the left, there are 'Consultas Externas' like Corpus NILC, Google, Answers.com, Wikipedia, and CORTEC. The main content area shows search results for an empty search term. It is divided into two sections: 'Português' and 'English'. Each section has a heading 'Resultados da busca por: ""' and a list of terms with their respective areas and links to the series 'House'. The Portuguese results include Amiloidose, Antibiótico, Convulsão, Coágulo, and Tomografia. The English results include Amyloidosis, Antibiotic, Clot, Scan, and Seizure. At the bottom right, there is a timestamp: '16/04/2013 07:46 © 2007 Guilherme Fromm - ICMC Jr.'

Figura 10. Termos disponíveis para o seriado House.

É importante frisar que

os termos não necessariamente são uma tradução (embora o *corpus*, nesse caso, seja paralelo). O site nos permite visualizar o termo em contraste nas duas línguas, mas a construção dos mesmos em cada língua é independente, dependendo dos exemplos coletados. (FROMM, 2011)

Ao selecionar o termo, será apresentada a definição final, bem como um exemplo dentro de um contexto, a posição na ordem de frequência e o número de ocorrências, as informações semânticas e as informações enciclopédicas.

The screenshot shows the 'Vocabulário Técnico Online' interface. At the top, there are navigation links for 'Tela Cheia', 'English', and 'Ajuda'. Below this, there are dropdown menus for 'Terminologia em Ficção' and 'House', and a search box with a 'Buscar' button. On the left side, there are three main sections: 'Tipos de Exibição' (Normal, Descritiva), 'Tipos de Consulta' (Total, Tradutor, Modular), and 'Consultas Externas' (Corpus NILC, Google, Answers.com, Wikipedia, CORTEC). The main content area displays two language options: Portuguese and English. Under Portuguese, there is a link 'Voltar ao resultado da busca' and a detailed definition of 'Amiloidose' from House's s.f.s. dictionary, including corpus frequency and encyclopedic information. Under English, there is a link 'Go back to search results' and a detailed definition of 'Amyloidosis' from House's n.m/f.s. dictionary, also including corpus frequency and encyclopedic information.

Figura 11. Tela final do termo *amiloidose/amyloidosis* disponível para consulta.

A página de consulta oferece dois modos de exibição (normal e descritiva) e três modos de consulta da microestrutura: total, tradutor e modular. Conforme o consulente seleciona as opções na barra lateral, a visualização da forma e do conteúdo da microestrutura dos verbetes é alterada.

Considerações finais

O objetivo inicial deste estudo, que era verificar as traduções dos termos técnicos em inglês para o português, foi mudado ao longo da pesquisa. Isso porque o material coletado nos levou a considerar que seria mais válido utilizar o nosso *corpus* para desenvolver um trabalho de cunho terminológico, resultando em um mini vocabulário online. Embora não tenha sido possível elaborar um vocabulário extenso com vários termos¹³, por questão de tempo, esta pesquisa serviu como ponto de partida para a criação de um projeto voltado prioritariamente para os alunos de Iniciação Científica: Terminologia em Ficção. Ao acessar a página de consulta, o leitor poderá verificar os trabalhos já concluídos (*Farscape*, *CSI*, *Sobrenatural* e *Lei e Ordem*) e um ainda em fase de pesquisa (*Jornada nas Estrelas*).

¹³ O dicionário criado tem apenas dez termos, sendo cinco em português e cinco em inglês.

Um fato relevante desta pesquisa, e do projeto como um todo, foi aliar o interesse que o aluno tem por uma série televisiva de sua predileção e a vontade de pesquisar. Os passos da pesquisa são facilmente replicáveis (existem vários programas de análise lexical, como o WordSmith Tools¹⁴ e vários ambientes de gestão terminológica, como o VoTec¹⁵) e podem despertar o interesse na pesquisa terminológica/terminográfica em outras áreas do conhecimento.

Referências

AUBERT, F. H. *Introdução à metodologia da pesquisa terminológica bilíngüe*. São Paulo: Humanitas, 1996.

BERBER SARDINHA, T. *Usando WordSmith Tools na investigação da linguagem*. Programa de Estudos Pós-Graduados em Linguística Aplicada e Estudos da Linguagem. LAEL, PUCSP. 1999.

_____. *Linguística de Corpus*. Barueri: Manole, 2004.

FROMM, G. *VoTec: a construção de vocabulários eletrônicos para aprendizes de tradução*. 2007. 210 f. Tese (Doutorado) – Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo, São Paulo, 2007.

_____. *Ficção, Tradução, Terminografia e Linguística de Corpus: confluências*. In: XIII Simpósio Nacional de Letras e Linguística e III Simpósio Internacional de Letras e Linguística, 2011, Uberlândia. Anais do SILEL. Volume 2, Número 2. Uberlândia: EDUFU, 2011.

ILARI, R. *Introdução ao estudo do léxico – brincando com as palavras*. 2ª ed. São Paulo: Contexto, 2003.

KRIEGER, M. das G.; FINATTO, M. J. B. *Introdução à terminologia: teoria e prática*. São Paulo: Contexto, 2004.

MARTHE, M. A cura pela razão. In: *Revista Veja*, n. 2208, p. 130-131, 16/03/2011.

SCOTT, M. *WordSmith Tools version 5*. Liverpool: Lexical Analysis Software, 2008

TEIXEIRA, E. D. *A Linguística de Corpus a serviço do tradutor: proposta de um dicionário de culinária voltado para a produção textual*. São Paulo, 2008. Tese (Doutorado – Programa de Pós-Graduação em Estudos Linguísticos e Literários em Inglês – Departamento de Letras Modernas). Faculdade de Filosofia, Letras e Ciências Humanas, Universidade de São Paulo.

¹⁴ Como o AntConc, disponibilizado gratuitamente em: <http://www.antlab.sci.waseda.ac.jp/software.html>.

¹⁵ Veja a análise de Wilkens *et alii* disponível em: <http://seer.ufg.br/riterm/article/view/29877>.

Wikipedia, Enciclopédia Livre. Wikimedia Foundation. *Gregory House*. Disponível em: http://en.wikipedia.org/wiki/Gregory_House. Acesso em: 22/09/10.

Artigo recebido em abril de 2013.

Aceito em julho de 2013.