

Identificação de unidades fraseológicas no vocabulário de *Star Trek*: abordagens *corpus-driven* e *corpus-based*

Identification of phraseological units in the vocabulary of *Star Trek*: *corpus-driven* and *corpus-based* approaches

Lucas Maciel Peixoto*

RESUMO: Este artigo analisou unidades fraseológicas a partir do *corpus* formado pelas legendas de todos os filmes e episódios de *Star Trek*, com o objetivo de demonstrar a existência de vocabulário específico em obras de ficção, e possíveis influências desse vocabulário na língua corrente. As bases teóricas foram a Teoria Comunicativa da Terminologia e também a Etnoterminologia, além de estudos sobre a Fraseologia. A Linguística de *Corpus* foi usada como base metodológica, por meio de uma abordagem *corpus-driven* (direcionada por *corpus*) e outra *corpus-based* (baseada em *corpus*). Os dados examinados provieram de um *corpus* de estudo compilado a partir das legendas em inglês dos seriados e filmes de *Star Trek*, e analisado via *WordSmith Tools*, e também de dois *corpora on-line* de língua inglesa: o *Corpus of Contemporary American English* e o *Google Books*. Os resultados do estudo mostraram exemplos de unidades fraseológicas que são usadas dentro do universo da série, inclusive termos específicos. Além disso, também foram encontradas frases criadas pela série e ressignificadas pelos fãs, tornando-se unidades fraseológicas na língua corrente.

PALAVRAS-CHAVE: Fraseologia. Terminologia. Linguística de *Corpus*. Ficção Científica. *Star Trek*.

ABSTRACT: This article analyzed phraseological units from a corpus consisting of the subtitles of all *Star Trek* films and episodes, aiming to demonstrate the existence of specific vocabulary in works of fiction, and possible influences of this vocabulary in everyday language. The theoretical basis was the Communicative Theory of Terminology and also the Ethnoterminology, as well as studies on Phraseology. Corpus Linguistics was used as a methodological basis, by means of a *corpus-driven* approach and a *corpus-based* one. The data examined came from a corpus of English language subtitles from the *Star Trek* series and movies, which was analyzed with *WordSmith Tools*, and also from two English language online corpora: the *Corpus of Contemporary American English* and *Google Books*. The results of the study showed examples of phraseological units which are used within the universe of the series, including specific terms. Additionally, the study also found phrases created by the series and to which new meanings were given by the fans, becoming phraseological units in everyday language.

KEYWORDS: Phraseology. Terminology. Corpus Linguistics. Science Fiction. *Star Trek*.

* Mestrando do Programa de Pós-Graduação em Estudos Linguísticos da Universidade Federal de Uberlândia. Possui Graduação em Letras – Inglês pela mesma universidade.

“This thing you call language, though. Most remarkable. You depend on it for so very much. But is any one of you really its master?”¹

1. Apresentação

Star Trek é uma franquia de ficção científica criada pelo produtor americano Gene Roddenberry na década de 60. A história fictícia tem início no século XXII, quando a humanidade alcança a capacidade de viagem espacial e inicia explorações interestelares, encontrando novos mundos, seres e fenômenos espaciais. Hoje, *Star Trek* assumiu grande proporção como fenômeno cultural, com uma extensa filmografia composta por seis seriados televisivos², que totalizam 716 episódios, e doze filmes. O conteúdo explorado durante as narrativas vai desde batalhas espaciais até a abordagem de conflitos morais e questões filosóficas, dentro de um universo povoado por inúmeras culturas extraterrestres, em um contexto de intenso avanço tecnológico, tudo isso acompanhado por um vocabulário próprio.

É justamente a existência desse vocabulário próprio que motiva a realização de estudos em Lexicologia e Terminologia sobre esse tipo de obra fictícia. Nesse contexto, este trabalho buscou identificar exemplos de unidades fraseológicas (em língua inglesa) no vasto inventário lexical do universo de discurso de *Star Trek*, e encontrar exemplos de expressões que tiveram sua origem no vocabulário da franquia, e que foram, posteriormente, apropriados e ressignificados na língua corrente, passando a ser usadas e entendidas até mesmo por falantes da língua que não são fãs da série. Pretende-se demonstrar que as bases teóricas da Lexicologia e da Terminologia podem ser aplicadas na análise e descrição do vocabulário de obras de ficção científica, e que as abordagens *corpus-driven* e *corpus-based* da Linguística de *Corpus* podem ser usadas de forma complementar para identificar e analisar fraseologismos sob diferentes olhares metodológicos.

2. Considerações teóricas

Unidades fraseológicas (UFs) são o objeto de estudo da Fraseologia, que, por sua vez, é considerada uma subdisciplina da Lexicologia, a área da Linguística responsável pelo estudo

¹ Kollos, embaixador alienígena, em *Star Trek: The Original Series*, 3ª temporada, 7º episódio.

² *The Original Series* (1966 a 1969), *The Animated Series* (1973 a 1974), *The Next Generation* (1987 a 1994), *Deep Space Nine* (1993 a 1999), *Voyager* (1995 a 2001) e *Enterprise* (2001 a 2005).

do léxico. Uma das atribuições da Lexicologia é a descrição e análise das lexias, que foram assim definidas por Silva (2006):

Lexias são elementos lexicais ou lexemas — unidades funcionais significativas de comportamento lingüístico [*sic passim*] que se opõem ao morfema e à palavra e que assumem o papel central na distinção das partes do discurso. Além disso, as lexias são formas e estruturas lingüísticas de natureza diferente. Suas características comuns consistem em que elas estão acumuladas no léxico, na parte da consciência lingüística que abrange as unidades denominativas, e em que elas exercem uma função denominativa para fenômenos da realidade. (SILVA, 2006, p. 11)

As lexias podem ser classificadas em simples, compostas e complexas. Resumidamente, lexias simples são as palavras simples, ou seja, palavras com um único radical, acrescidas ou não de afixos. Lexias compostas são agrupamentos de lexias simples por aglutinação ou justaposição, algumas vezes com o uso de hífen, gerando integrações semânticas. Lexias complexas, por sua vez, são sequências lexemáticas de alta produtividade e estrutura relativamente fixa (SILVA, 2006).

Lexias complexas podem ocorrer na língua sob a forma de UFs. Por sua complexidade, é difícil obter uma definição formal de UF, e vários autores apresentam ideias divergentes sobre o assunto. Exemplos de fenômenos lingüísticos que costumam ser considerados UFs são: expressões idiomáticas, provérbios, locuções, estruturas características de determinados contextos de comunicação, como abertura e fechamento de textos jurídicos e cartas, entre outros (WELKER, 2004).

Independentemente do que se considera ou não uma UF, é de comum acordo que são construções lexicais fixas que ocorrem com frequência na língua. Um falante nativo do português, por exemplo, provavelmente já ouviu expressões como *estar nas nuvens*, que significa estar feliz ou satisfeito, ou *beco sem saída*, que representa uma situação problemática para a qual não há solução aparente. Frequentemente (mas não necessariamente), esse tipo de expressão possui versões em outras línguas, como *to be on cloud nine* e *dead end*, idiomatismos em inglês que correspondem, respectivamente, aos mencionados acima, em português, com significados semelhantes. Nota-se que são expressões fixas, ou seja, sempre aparecem com as mesmas palavras ou derivações, e possuem um significado único, isto é, a ideia de estar feliz e satisfeito, nesse caso, só é obtida com a frase completa *estar nas nuvens*. Observadas isoladamente, as palavras contidas em *estar / nas / nuvens* não levam a essa interpretação.

Portanto, uma característica importante de uma UF é o fato de que “o significado do todo é diferente da soma dos significados das partes” (WELKER, 2004, p. 165).

Segundo Biderman (2005), expressões idiomáticas como os exemplos citados no parágrafo anterior são o caso mais extremo de UF, mas não são o único exemplo. Também podem ser consideradas UFs: construções substantivo + adjetivo, como *caixa preta* e *mesa redonda*, nas quais o adjetivo é obrigatório e fixo e constitui uma unidade com o substantivo, de forma que *caixa vermelha* ou *mesa quadrada*, por exemplo, são sintagmas livres que não possuem o mesmo valor semântico das UFs; construções com verbos, como *levar calote* ou *ter compromisso*; construções como os *phrasal verbs* da língua inglesa, como *fall for (somebody)* ou *look forward to (something)*, que possuem significado fixo diferente da soma dos significados individuais dos verbos e preposições que os compõem; e vários outros exemplos.

Além de integrar o léxico da língua corrente e possuir significado amplamente conhecido pelos falantes, as UFs também podem ocorrer em contextos específicos de comunicação, como no vocabulário técnico-científico comum em áreas especializadas do conhecimento. Por exemplo: a página da *internet* do Supremo Tribunal Federal conta com um glossário jurídico que contém termos como *transitar em julgado* e *efeito vinculante*, expressões próprias da área que fazem sentido para os profissionais que nela atuam; na economia, a expressão do latim *coeteris paribus (tudo o mais é constante)* é usada em análises de mercado nas quais se considera uma variável de estudo sem que as demais variáveis sofram alterações; no ambiente corporativo, *pensar fora da caixa* significa ter ideias não convencionais ou propor soluções inovadoras.

O significado dos últimos exemplos citados não é amplamente conhecido pelos falantes da língua, uma vez que essas expressões não estão presentes na língua corrente, mas, sim, no vocabulário específico de determinadas especialidades, de conhecimento restrito aos especialistas. Portanto, essas UFs podem ser analisadas como termos, que são o objeto de estudo da Terminologia.

Em linhas gerais, a Terminologia lida com a existência de vocabulário específico nas diversas áreas do conhecimento científico. A teoria da Terminologia que fundamenta este trabalho é a Teoria Comunicativa da Terminologia – TCT (CABRÉ, 1993). De acordo com Sales (2007), a TCT surge como um contraponto à Teoria Geral da Terminologia – TGT, a teoria clássica da Terminologia proposta por Eugen Wüster, que se pautava sobre uma perspectiva prescritiva e normativa dos termos, o que oferecia pouca flexibilidade quanto ao seu contexto

de uso. A TCT, em contrapartida, busca abordar o termo de forma descritiva, levando em consideração sua dimensão textual e discursiva. Isso traz a relativização do que pode ser considerado um termo, de acordo com o contexto em que aparece.

Portanto, a ideia de termo trazida pela TCT estabelece novas diretrizes para o trabalho terminológico, sobretudo a identificação e descrição de termos a partir do seu contexto de uso. Levando em consideração essas novas diretrizes, Tagnin e Bevilacqua (2013) afirmam que, a partir do final da década de 1980, iniciou-se uma interação produtiva entre a Linguística de *Corpus* (LC) e a Terminologia, uma vez que, para se encontrar os contextos de uso de um termo, observam-se suas ocorrências em textos especializados. Dentro dessa perspectiva, isso é feito por meio do levantamento de *corpora* especializados e a posterior identificação, análise e descrição de termos, usando a LC como base metodológica e considerando as situações comunicativas em que se encontram (BEVILACQUA, 2013).

2.1. A Etnoterminologia e o vocabulário de *Star Trek* como texto especializado

Um dos objetivos principais deste estudo é demonstrar que o discurso especializado não está presente apenas nas ciências ou nas atividades profissionais técnicas, mas, também, em obras de ficção. Para tanto, é válido considerar a definição de texto especializado trazida por Oliveira e Müller (2013). Segundo os autores, o texto especializado é um instrumento de comunicação usado por especialistas em determinada área do conhecimento humano, em um contexto caracterizado pela presença de termos específicos. Os autores definem as linguagens especializadas como “um subconjunto léxico que em determinadas situações comunicacionais adquire traços peculiares de significado” (OLIVEIRA, MÜLLER, 2013, p. 52).

O saber especializado e os traços peculiares de significado aos quais se referem os autores, quando inseridos em um universo de discurso etno-literário, são o foco da Etnoterminologia, uma subdisciplina da Terminologia segundo a qual “as unidades lexicais do universo de discurso etno-literário têm um estatuto próprio e exclusivo” (BARBOSA, 2006, p. 48). Barbosa explica que as unidades lexicais próprias do contexto etno-literário possuem valor semântico social e carregam, também, um valor cultural particular a determinado universo de discurso. Como exemplo, a autora cita a história do *Bumba-meu-boi*, na qual a unidade lexical *boi* possui um significado particular. Não se trata do animal encontrado na natureza, mas, sim, de um ser fantástico que integra o mito. Segundo Barbosa, essa particularidade do texto etno-literário justifica a formalização da Etnoterminologia como disciplina científica.

Este trabalho propõe que essa base teórica pode ser aplicada à análise e descrição do vocabulário encontrado na ficção científica. Segundo Bang e Fromm (2013), unidades lexicais terminológicas aparecem em obras ficcionais porque há, nessas obras, uma preocupação com a verossimilhança da narrativa, dos personagens, e do contexto em geral. No modelo proposto pelos autores, *Star Trek* se enquadra no tipo de obra que mistura ficção e ciência na criação do seu vocabulário:

[Nessas obras,] 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. (BANG, FROMM, 2013, p. 5)

Assim, pode-se dizer que o vocabulário de *Star Trek* tem características similares ao vocabulário presente nos textos especializados e etno-literários. Há um vocabulário criado pelos produtores da série e absorvido pelos fãs, que originou formas de expressão tanto dentro quanto fora do contexto da narrativa fictícia. Analogamente ao que ocorre nos universos etno-literários e áreas do conhecimento técnico-científico, há em *Star Trek* um universo de discurso e cultura particulares, acompanhado por um vocabulário específico. Os fãs da série dominam esse vocabulário e fazem uso dele para se expressar e se comunicar uns com os outros, dentro de determinada situação comunicacional. Esse vocabulário será analisado nas seções seguintes.

3. *Corpus* de estudo e metodologia

Este estudo buscou apresentar exemplos de UFs que ilustrem o vocabulário em questão, por meio da análise de um *corpus* de estudo composto pelas legendas de todos os episódios e filmes de *Star Trek* produzidos até o momento da redação deste texto. As legendas foram obtidas por meio de *download* em repositórios de legendas feitas por legendistas não profissionais, na *internet*. Esse tipo de legenda foi escolhido porque pode ser obtido de forma rápida e gratuita, o que é importante devido ao grande número de filmes e episódios que existem na série. O *corpus* de estudo possui 3.070.626 *tokens*, e é composto por porções de fala transcritas em legendas; é sincrônico (limita-se ao período de tempo durante o qual se deu a produção da série *Star Trek*: 1966 a 2013) e em língua inglesa. Os textos foram analisados pelo programa de análise lexical *WordSmith Tools* (SCOTT, 2008), em busca de informações sobre a frequência de ocorrência das UFs de interesse, e também em busca de contextos para

exemplificar o uso pragmático dessas UFs. Também foram analisadas informações advindas de *corpora* disponíveis *on-line*: o *Corpus of Contemporary American English* – COCA³ (DAVIES, 2014) e o *Google Books*⁴ por meio da ferramenta *Google Ngram Viewer*⁵ (MICHAEL, et al, 2010). Ao manusear esses *corpora*, foram usados dois caminhos diferentes: uma abordagem *corpus-based* (baseada em *corpus*) e outra *corpus-driven* (direcionada por *corpus*).

A abordagem *corpus-based* é tradicionalmente usada para “explicar, testar ou exemplificar teorias e descrições que foram formuladas antes de os grandes *corpora* terem se tornado disponíveis para informar o estudo da língua” (TOGNINI-BONELLI, 2001, p. 65⁶). Assim, o ponto inicial de uma análise *corpus-based* é a existência prévia de determinada formulação sobre a língua, que pode ser fruto de uma teoria anterior ou mesmo da reflexão intuitiva e da curiosidade. Partindo-se disso, usa-se a metodologia da LC para contrastar essa formulação prévia com dados empíricos provenientes de um *corpus*. Geralmente, a observação desses dados ratifica parcialmente a hipótese inicial e, ao mesmo tempo, costuma explicitar casos em que ela não se aplica. Cabe ao linguista escolher a melhor maneira de lidar com esses casos excepcionais. Tognini-Bonelli (2001) explica que, para isso, existem três caminhos possíveis, a saber: (i) isolamento, ou seja, manter a hipótese inicial intacta e isolar os casos não explicados, considerando-os exceções; (ii) padronização, ou seja, reformular parcialmente a hipótese inicial, tornando-a mais geral ou ampla para que abarque os novos casos encontrados, enriquecendo-a; e (iii) instanciação, ou seja, “inserir os dados em um sistema de possibilidades abstratas, um conjunto de escolhas paradigmáticas disponível em qualquer ponto, no texto” (TOGNINI-BONELLI, 2001, p. 74⁷). Esses três caminhos não são mutuamente excludentes, podendo ser usados em conjunto de forma complementar. A noção principal é a de que o uso de um *corpus* permite a verificação de evidências explícitas para confirmar, refutar ou aprimorar uma teoria ou formulação linguística ou uma simples indagação sobre determinado fenômeno linguístico.

³ <http://corpus.byu.edu/coca/>

⁴ <http://books.google.com/>

⁵ <https://books.google.com/ngrams>

⁶ Todas as traduções são de minha autoria. No original: the term *corpus-based* is used [...] mainly to expound, test or exemplify theories and descriptions that were formulated before large corpora became available to inform language study.

⁷ No original: [...] building the data into a system of abstract possibilities, a set of paradigmatic choices available at any one point in the text.

A abordagem *corpus-driven*, por sua vez, “busca derivar categorias linguísticas sistematicamente a partir dos padrões recorrentes e das distribuições de frequência que emergem da língua em contexto” (TOGNINI-BONELLI, 2001, p. 87⁸). O estudo direcionado por *corpus* tem início com o próprio *corpus*, e não se baseia em formulações teóricas previamente construídas. São estudados os padrões que emergem das linhas de concordância do *corpus* com o objetivo de elaborar teorias que reflitam as evidências, seguindo as seguintes etapas: observação, hipótese, generalização, unificação. Nota-se que a hipótese não é a primeira etapa da abordagem; é formulada após a observação inicial. O foco em dados empíricos torna essa abordagem mais indutiva do que a *corpus-based*. No entanto, “não existe indução pura” (TOGNINI-BONELLI, 2001, p. 85⁹), e a interpretação dos dados continua sujeita à percepção individual do pesquisador, que “faz sua própria seleção e agrupamento dos fatos, determinados pelas suas atitudes e teorias e pela natureza da sua experiência de realidade da qual ele mesmo faz parte” (J. R. FIRTH, 1969, p. 29, *apud* TOGNINI-BONELLI, 2001, p. 85¹⁰).

Grande importância e ênfase são dadas à abordagem *corpus-driven*, por se tratar de um tipo de análise que, frequentemente, desafia teorias linguísticas baseadas na intuição individual. De acordo com Tognini-Bonelli (2001), John Sinclair chegou a denominar as formulações linguísticas baseadas somente na intuição como “crenças pré-*corpus*”. Nas palavras da própria autora, “as constatações inesperadas derivadas de evidências de *corpus* levam à conclusão que a intuição não é uma fonte absolutamente confiável de informações sobre a língua” (TOGNINI-BONELLI, 2001, p. 86¹¹).

Neste trabalho, escolheu-se analisar o vocabulário de *Star Trek* sob a ótica de ambas as abordagens, cada uma em um momento. Considerando-se os objetivos da análise, tanto uma quanto a outra forneceram resultados relevantes e distintos, constituindo dois olhares complementares e direcionados ao mesmo objeto. Aqui, entende-se *corpus-driven* como uma forma de olhar para a linguagem de *Star Trek* partindo do que aparece no *corpus* de estudo, procurando identificar UFs por um procedimento que não pressupõe conhecimento prévio da série. Já o olhar *corpus-based* partiu de frases amplamente conhecidas por fãs da série,

⁸ No original: [...] aims to derive linguistic categories systematically from the recurrent patterns and the frequency distributions that emerge from language in context.

⁹ No original: There is no such thing as pure induction.

¹⁰ No original: Each scholar makes his own selection and grouping of facts determined by his attitudes and theories and by the nature of his experience of reality of which he himself is part.

¹¹ No original: The unexpectedness of the findings derived from corpus evidence leads to the conclusion that intuition is not comprehensively reliable as a source of information about language.

buscando constatar se essas frases são ou não UFs, e descrever o seu uso. As seções seguintes serão dedicadas a demonstrar esse processo e os resultados obtidos¹².

4. Identificação de UFs em *Star Trek*: abordagem *corpus-driven*

Buscando seguir os princípios de uma abordagem direcionada pelo *corpus*, foi estudada, primeiramente, a lista de palavras e a lista de palavras-chave do *corpus* de estudo, geradas a partir do *WordSmith Tools* (WST) e expostas abaixo.

N	Word	Freq.	%	Texts	%
1	CAPTAIN	12,944	0.17	665	91.35
2	SIR	7,915	0.11	632	86.81
3	TIME	7,791	0.10	727	99.86
4	WERE	7,265	0.10	726	99.73
5	SHIP	6,392	0.09	673	92.45
6	MR	5,890	0.08	585	80.36
7	L'M	5,635	0.08	236	32.42
8	BACK	5,515	0.07	719	98.76
9	GOOD	4,823	0.06	719	98.76
10	MAKE	4,337	0.06	716	98.35
11	FIND	4,032	0.05	708	97.25
12	PEOPLE	3,875	0.05	649	89.15
13	LIFE	3,780	0.05	678	93.13
14	DOCTOR	3,379	0.05	583	80.08
15	LTS	3,245	0.04	221	30.36
16	LT	3,183	0.04	362	49.73
17	LONG	3,170	0.04	668	91.76
18	ENTERPRISE	3,143	0.04	413	56.73
19	COMMANDER	3,024	0.04	582	79.95
20	YEARS	2,913	0.04	641	88.05

Figura 1. Lista de Palavras do *corpus* de estudo (com *stoplist*).

¹² Na literatura sobre esse assunto, como visto em Tognini-Bonelli (2001), costuma-se falar primeiro sobre a abordagem *corpus-based* e, depois, sobre a *corpus-driven*. No entanto, na análise feita por este trabalho, optou-se por apresentar a abordagem *corpus-driven* primeiro e a *corpus-based* em segundo lugar, pois acredita-se que, nesse caso, essa sequência permite visualizar melhor a diferença entre os dois procedimentos.

N	Key word	Freq.	%C.	Fred	Keyness
1	CAPTAIN	12,944	0.17	20,395	60,945.01
2	MR	5,890	0.08	0	47,666.33
3	L'M	5,635	0.08	0	45,602.49
4	SHIP	6,392	0.09	1,974	42,657.03
5	BACK	5,515	0.07	2,146	35,619.99
6	MAKE	4,337	0.06	1,721	27,927.61
7	DOCTOR	3,379	0.05	49	26,832.44
8	SIR	7,915	0.11	29,931	26,295.90
9	LTS	3,245	0.04	0	26,259.86
10	LT	3,183	0.04	0	25,758.11
11	SPOCK	2,586	0.03	0	20,926.74
12	WARP	2,701	0.04	293	19,949.42
13	STARFLEET	2,401	0.03	0	19,429.60
14	LF	2,246	0.03	0	18,175.25
15	L'VE	2,226	0.03	0	18,013.40
16	L'LL	2,022	0.03	0	16,362.52
17	SHE'S	1,750	0.02	0	14,161.37
18	DR	1,665	0.02	0	13,473.51
19	HOME	1,691	0.02	74	13,072.24
20	WORF	1,466	0.02	0	11,863.12

Figura 2. Lista de Palavras-Chave¹³ do *corpus* de estudo (baseada na Lista de Palavras com *stoplist*).

Essas listas iniciais foram construídas com uso de *stoplists*, com o objetivo de filtrar as ocorrências de palavras gramaticais (preposições, artigos, pronomes etc), uma vez que se pretende estudar as palavras que tem significado no contexto específico do *corpus* de estudo. As listas permitem as primeiras suposições sobre as características do vocabulário usado em *Star Trek*. Nota-se, por exemplo, que a palavra *captain* é, ao mesmo tempo, a mais frequente e a mais chave, o que condiz com o fato de que a maior parte da narrativa acontece dentro de espaçonaves cujo capitão é, sempre, um dos personagens principais.

Como o interesse deste estudo foi identificar UFs, o próximo passo foi encontrar palavras que co-ocorram frequentemente com os candidatos a termo constantes nas listas iniciais, constituindo frases. Na computação, a co-ocorrência de itens em uma cadeia de dados é denominada N-Grama. Em outras palavras, um N-Grama é uma porção de N itens ocorrendo próximos uns dos outros em um trecho de informação (CAVNAR, TRENKLE, 1994). Se a informação for linguística, os dados serão itens linguísticos como letras, sílabas ou palavras inteiras. Programas de análise lexical permitem visualizar N-Gramas formados por palavras

¹³ *Corpus* de referência usado para produzir a lista de palavras-chave: *COCA 100k Wordlist*

dentro de um *corpus*. No WST, os N-Gramas são denominados *clusters*. Um *cluster* é um conjunto de palavras que co-ocorrem com uma palavra inicial de busca, permitindo a observação de padrões de fraseologia a partir de linhas de concordância (SCOTT, 2014). Assim, para identificar UFs no *corpus* de estudo, foi usada a ferramenta de concordância¹⁴ do WST para visualizar *clusters* com as palavras-chave obtidas nas listas iniciais. A Figura 3, abaixo, demonstra alguns dos *clusters* (com duas palavras) mais frequentes com a palavra *captain*.

N	Cluster	Freq.	Set	Length
1	THE CAPTAIN	1,404		2
2	CAPTAIN KIRK	543		2
3	CAPTAIN I	380		2
4	TO CAPTAIN	334		2
5	AYE CAPTAIN	323		2
6	YOU CAPTAIN	322		2
7	CAPTAIN JANEWAY	306		2
8	CAPTAIN THE	267		2
9	CAPTAIN ARCHER	244		2
10	YES CAPTAIN	230		2
11	CAPTAIN WE	202		2
12	IS CAPTAIN	180		2
13	YOUR CAPTAIN	177		2
14	CAPTAIN SISCO	163		2
15	CAPTAIN YOU	143		2
16	TO THE CAPTAIN	142		3
17	THIS IS CAPTAIN	140		3
18	CAPTAIN AND	138		2
19	CAPTAIN OF	133		2
20	ME CAPTAIN	130		2

Figura 3. Lista de *clusters* com a palavra *captain*.

Como visto, a palavra *captain* está frequentemente associada a nomes próprios (Kirk, Janeway, Archer, Sisko), e também à palavra *aye*, entre outras. No WST, é possível expandir cada um desses *clusters*, visualizando-os em listas de linhas de concordância para observar o contexto em que ocorrem. A análise de contextos com a frase “*Aye, Captain*” permite concluir seu significado:

- 1) - I'll expect a full report in the morning.

¹⁴ No *WordSmith Tools* (e nos programas de análise lexical em geral), o levantamento de *clusters* é feito pela ferramenta de concordância.

- **Aye, captain.**

2) - Prepare to lock on to the crewmen.

- **Aye, Captain.**

3) - Lay in a course for the Tzenkethi border.

- **Aye, aye, Captain.**

4) - Scotty, lock the transporter on that small vessel. We're going to rescue that pilot.

- **Aye, aye, Captain.**

5) - Mr. Scott, lock phasers onto our coordinates.

- **Aye, captain.** All phasers locked on.

Pelo contexto, podemos concluir que “*Aye, Captain*” é uma frase dita por membros da tripulação ao obedecer às ordens dadas pelo capitão. É uma *lexia* usada com frequência, de forma fixa, e sempre no mesmo contexto (após uma instrução dada pelo capitão), o que a torna uma candidata a UF.

Seguindo os mesmos passos, podemos encontrar outros possíveis candidatos a UF, como demonstrado a seguir:

6) - We must evacuate decks 20 to 24.

- **Make it so.** Shut down life support to those decks.

7) - Captain, transfer complete, we are ready to leave orbit.

- **Make it so,** Number One.

8) - Can you modify an emergency beacon to operate on UV?

- Yes, sir.

- **Make it so.**

9) - Course for Station McKinley laid in.

- **Make it so,** Number One.

10) - I recommend we launch a probe.

- **Make it so.**

Na Figura 2, a palavra *make* ocupa o 6º lugar na lista de palavras-chave, o que sugere que seu uso no vocabulário de *Star Trek* pode aparecer em contextos passíveis de análise. Com efeito, foram encontradas 99 ocorrências da frase “*Make it so*”, entre outras *lexias* com *make*. A análise de contextos como os exemplificados acima demonstra que essa frase sempre é usada

pelo capitão da nave em resposta a um membro da tripulação, orientando o cumprimento de determinada tarefa. O número de ocorrências, o contexto e a fixidez da frase indicam um possível candidato a UF.

Vejamos outros contextos:

- 11) Our **warp core** is designed to operate for up to three years before refueling.
- 12) - Mr. La Forge, prepare to initiate **warp drive**.
- There's too much interference to form a **warp field**.
- 13) - **Warp engines** are now available.
- Very good. Number One, get us to Daled as quickly as possible.
- 14) At **maximum warp**, we can be there in less than three days.
- 15) Commander Tucker believes it's safe to maintain this **warp factor**.
- 16) - Helm, give us **warp speed**.
- Warp one, sir.
- 17) I remember. The ship was destroyed by a **warp-core breach**.

Nos contextos acima, as lexias destacadas são formadas a partir da palavra *warp*, que ocupa a 12ª posição na lista de palavras-chave. O número total de ocorrências de cada uma dessas lexias está detalhado no Quadro 1. Pelos contextos, observa-se que *warp* está relacionado à velocidade da espaçonave e aos equipamentos necessários para obter essa velocidade. Pela ocorrência frequente e estrutura fixa, essas lexias podem ser consideradas candidatas a UF e, também, a termos, integrando o vocabulário específico de *Star Trek*.

Quadro 1. Lexias com a palavra *warp* e número de ocorrências.

Lexias com a palavra <i>warp</i>	Número de Ocorrências
<i>warp core</i>	260
<i>warp drive</i>	249
<i>warp engines</i>	128
<i>warp field</i>	114
<i>maximum warp</i>	105
<i>warp factor</i>	89
<i>warp speed</i>	83
<i>warp core breach</i>	44

Foram apresentados alguns exemplos de lexias formadas a partir de palavras-chave encontradas no *corpus* de estudo. Destacam-se as UFs formadas a partir da palavra *warp*, que ocorrem com grande frequência e são termos específicos da série, de grande relevância na narrativa. Seria possível continuar a análise e encontrar ainda mais lexias, usando o procedimento demonstrado ou outros métodos e ferramentas; no entanto, este trabalho não objetiva a identificação e descrição exaustiva dos fraseologismos em *Star Trek*, bastando demonstrar alguns exemplos que evidenciam a presença desse vocabulário específico.

Todas essas lexias foram encontradas a partir de uma abordagem *corpus-driven*, pela observação de *clusters* com as palavras-chave encontradas pelas ferramentas do WST. Essa análise direcionada pelo *corpus* poderia ter sido feita por qualquer pesquisador que usasse o mesmo *corpus* de estudo e as mesmas ferramentas. Ou seja, nesse caso, não seria necessário ter conhecimento prévio sobre o vocabulário da série, ou ser um fã de *Star Trek*, para encontrar as mesmas UFs ou outros exemplos.

5. Identificação de UFs em *Star Trek*: abordagem *corpus-based*

Como visto, em uma abordagem baseada em *corpus*, supõe-se que há conhecimento prévio sobre determinado objeto de estudo dentro do *corpus*, e se pretende usar a metodologia da LC para confirmar a existência desse objeto ou analisar suas características mais detalhadamente. Assim, diferentemente da abordagem anterior, foram estabelecidas previamente algumas lexias de interesse, já conhecidas dentro do universo de *Star Trek*, para um estudo aprofundado.

Certas frases enunciadas por personagens da série são amplamente conhecidas pelos fãs. Algumas delas ocorrem com grande frequência; outras ocorreram apenas uma ou algumas vezes, mas o contexto em que apareceram foi o suficiente para torná-las favoritas. Informalmente, esse tipo de frase pode ser chamado de *catchphrase*; ou seja, são frases que, devido ao seu uso peculiar, “pegaram” e passaram a ser comumente usadas e referidas. É comum encontrar artigos sobre o assunto na *internet*, como Marshall (2013) e outros.

O interessante é que o uso lúdico dessas frases pelos fãs, de forma literal ou metafórica, costuma produzir ressignificações e novas formas de expressão. Com base nisso, as frases listadas abaixo foram selecionadas para uma análise detalhada porque, supõe-se, podem ser fraseologismos ou, ao menos, ocorrer de forma peculiar, já que são amplamente conhecidas pelos fãs e frequentemente usadas para remeter à série ou de forma ressignificada, como será

possível observar nos exemplos que serão listados. A observação de exemplos do seu contexto de uso determinará se são ou não UFs:

- a) Space, the final frontier.
- b) To boldly go where no man has gone before.
- c) Beam me up, Scotty!¹⁵

A seguir, essas frases serão analisadas tanto no *corpus* de estudo quanto em outros *corpora* de língua inglesa, buscando demonstrar seus significados tanto dentro quanto fora da narrativa fictícia da série.

No *corpus* de estudo, as frases *a* e *b* ocorrem nas séries *The Original Series*, *The Animated Series*, e *The Next Generation*. Na entrada de cada episódio dessas séries, um personagem profere a seguinte mensagem, que diz respeito à sua missão:

Space, the final frontier. These are the voyages of the starship Enterprise. Its five-year mission: to explore strange new worlds, to seek out new life and new civilizations, to boldly go where no man has gone before.

A mensagem sofre algumas alterações durante os anos de produção das diferentes séries. A frase *a* ocorre 272 vezes. A frase *b*, exatamente como escrita acima, ocorre 101 vezes. Há uma variação da frase em que a palavra *one* é usada ao invés de *man* (*To boldly go where no one has gone before*). Essa variação ocorre 175 vezes. Observa-se, ainda, que a frase com *man* ocorre apenas em episódios da série original da década de 60, enquanto a frase com *one* ocorre em episódios das séries *The Next Generation* e *The Animated Series*, que foram produzidas depois da série original. Isso indica que os produtores substituíram *man* por *one* em produções mais recentes, provavelmente em uma tentativa de tornar o vocabulário mais neutro em relação aos gêneros.

Dentro da narrativa de *Star Trek*, essas frases não são UFs. Embora apareçam com frequência e sejam fixas, seu significado não pode ser unificado como explicado por Welker (2004), e são, portanto, sintagmas livres. No entanto, a mensagem inicial da entrada de cada episódio tornou-se amplamente conhecida e referida pelos fãs, ultrapassando a narrativa e

¹⁵ Conforme será explicado mais adiante neste texto, essa frase não ocorre no *corpus* de estudo exatamente como escrita aqui. “*Beam me up, Scotty!*” é uma expressão comumente usada pelos fãs a partir da condensação do verbo frasal *beam up*, que se tornou comum a ponto de ser dicionarizado, como em Prucher (2007).

chegando à língua corrente. É possível observar o uso dessas lexias na língua inglesa em um *corpus* monitor como o COCA, que possui 450 milhões de palavras e cuja extensão vai desde 1990 até 2012 (no momento da redação deste texto), incluindo textos de cinco diferentes gêneros: língua falada, ficção, revista, jornal e texto acadêmico.

Alguns exemplos de uso da frase *the final frontier*, retirados do COCA:

18) - And I know, Dr. Herr, that you are now about to do something that would be even revolutionary stuff, which is people who are having their legs amputated above the knee. Sort of been **the final frontier** in terms of the technology until now, right?
- Yes, we're just now offering this technology for people that are amputated about the knee. So, we're just -- from the ground up, we're rebuilding people.

19) - Bill, that has a wonderful beat. You can dance to that.
- You can, but I can't.
[laughter]
- No, I just can't -- if I could dance, that would be the -- that's **the final frontier** for me. If I could do that, I probably do anything.

20) - Let's get you, Amy, away from the piano a bit. Have you sing. Now is it embarrassing to sing?
- Yes. It's **the final frontier**.

Nos exemplos acima, a frase é uma UF, porque aparece de forma metafórica com significado unificado. No primeiro exemplo, representa um estágio de desenvolvimento científico considerado uma “fronteira final”, uma barreira tecnológica a ser ultrapassada. No segundo e terceiro exemplos, representa um obstáculo pessoal de grande significância.

Embora *the final frontier* seja uma UF quando usada em contextos como os citados, resta saber se é um fraseologismo que surgiu do vocabulário de *Star Trek*, ou se existe na língua inglesa de maneira independente. Afinal, a intuição nos leva a crer que é plenamente possível usar essa frase de forma metafórica sem a influência da série.

Uma maneira de buscar evidências da influência de *Star Trek* no vocabulário da língua corrente é usar um *corpus* que permita uma análise diacrônica, possibilitando visualizar a evolução do uso de determinada frase ao longo do tempo. Considerando-se que *Star Trek* começou em 1966, o COCA não permite essa análise, pois sua amostra da língua começa no ano de 1990. Um *corpus* mais apropriado para esse tipo de consulta é o *Google Books*, que reúne milhões de livros e revistas em várias línguas, conta com mais de um trilhão de *tokens*, e tem um escopo que se inicia a partir do ano de 1500. A ferramenta *Google Ngram Viewer*

permite a visualização de N-Gramas provindos do *Google Books*. A Figura 4, abaixo, exhibe a distribuição de ocorrências de *the final frontier* nesse corpus a partir de 1900.



Figura 4. Ocorrências de *the final frontier* a partir de 1900. O eixo vertical do gráfico representa a porcentagem normalizada de ocorrências dessa frase em relação ao restante dos N-Gramas indexados pela ferramenta em um dado ano. Fonte: *Google Books Ngram Viewer*.

O gráfico mostra que já havia ocorrências da frase antes da existência de *Star Trek*. No entanto, há um visível aumento do uso da frase a partir da década de 80, cerca de 15 anos após o início da série. Esses dados podem ser interpretados como indicação de uma possível influência da série no vocabulário da língua corrente, embora de maneira inconclusiva.

A ferramenta permite, ainda, visualizar os contextos de uso da frase, representados no gráfico, a partir de exemplos do banco de dados do *Google Books*. O Google disponibiliza algumas obras completas para visualização e outras apenas em forma de fragmentos ou *snippets*. A seguir, alguns exemplos de *the final frontier* encontrados entre 1950 e 1962 (antes da existência de *Star Trek*):

- 21) Additional official luggage of the two Parties shall likewise be passed without customs inspection if it carries customs seals and if measures are taken that this luggage be handed over to the couriers only at **the final frontier** point. (1950)
- 22) This provisional, ethnic, and minimum border had few intrinsic merits, and "both the French and the Americans believed that **the final frontier** line should be farther to the east." (1962)
- 23) Men and women of science are pushing toward **the final frontier** of infectious disease, toward the day when mankind will no longer be at the mercy of such invaders. (1952)
- 24) For this immense moment in history will be the launching of the first earth satellite vehicle, man's first exploratory step in the conquest of **the final**

frontier — space itself. (1957)

Os dois primeiros exemplos evidenciam usos literais da frase, referindo-se à fronteira entre países vizinhos, o que é reforçado pelas palavras *point*, no exemplo 21, e *line*, no exemplo 22. O terceiro exemplo é uma UF, com sentido semelhante aos contextos retirados do COCA e demonstrados anteriormente (uma barreira ou obstáculo a ser ultrapassado). No último exemplo, a frase é utilizada de forma muito semelhante ao modo como aparece na série, mostrando de forma clara que a ideia do espaço como a última fronteira já existia antes de *Star Trek*.

O seguinte exemplo é de 1999, também retirado do *Google Books*:

- 25) Perhaps, contrary to popular belief, the neural network of the human brain is truly **the final frontier**, and our greatest mission is to **go to places where no one has gone before**.

Nesse exemplo, o vocabulário usado tem, claramente, influência da série: foi retirado de um texto escrito após o início de *Star Trek*, menciona uma missão, e finaliza com uma variação da frase *b*, listada anteriormente.

Considerando esses dados, é possível concluir que a frase *the final frontier* sempre existiu na língua, tanto de forma literal quanto metaforicamente. No entanto, seu uso foi, possivelmente, popularizado após a criação de *Star Trek*.

Vejam, agora, alguns exemplos da frase *b*, provindos do COCA:

- 26) Costing nearly \$100 million, this three-acre high-tech greenhouse, sealed from the outside and completely self-sufficient, will become home to eight humans, a few wide-eyed primates and a creeping, crawling, climbing cast of thousands - a Noah's Ark in the desert, **boldly going where no ecologist has gone before**.
- 27) If you want to live more dangerously, go with a brand-new 100-MHz Pentium, a Cyrix MI, or an AMD K5. You can be the first in your town **to boldly go where no computer user has gone before**.
- 28) To coin a phrase, Sister Wendy Beckett is a riddle wrapped in an enigma wrapped in a habit, a combination that's worked with British audiences, and those in America can now look forward to watching this nun who **boldly goes where no nun has gone before**.

Aqui, a frase é uma UF que tem o sentido metafórico de aventurar-se de forma ousada e trilhar caminhos previamente inexplorados. Além desse valor semântico, a frase possui uma estrutura peculiar. Os exemplos demonstram a frequente substituição da palavra *man* por outra palavra que se adeque ao contexto do que está sendo dito. A derivação, como ocorre com o verbo *go* (*going*), é esperada em qualquer UF, para adaptar a lexia à estrutura do restante da frase. Observa-se, também, que o advérbio *boldly* foi usado, sempre, antes do verbo *go*, assim como no original em *Star Trek*. Isso é uma ocorrência incomum na língua inglesa, uma vez que os advérbios costumam ser usados após os verbos. Essa construção invertida é chamada de *split infinitive*¹⁶ e, na verdade, *to boldly go where no man has gone before* é um dos exemplos mais conhecidos dessa estrutura, de acordo com a *Wikipedia*¹⁷ (ver artigo: *Split Infinitive*, em inglês) e com Edmondson (2013). Há controvérsias sobre o uso do *split infinitive*, e algumas gramáticas consideram essa uma construção incorreta, uma vez que, no latim, o infinitivo é formado por uma única palavra; portanto, separá-lo seria incorreto. Sobre a frase em *Star Trek*, Edmondson (2013) escreveu em seu *blog*:

Não há problemas com *split infinitives*, se usados com cuidado. Na frase de *Star Trek*, “*To boldly go where no man has gone before*,” a sentença não teria o mesmo efeito se fosse formulada de maneira diferente. Por exemplo: “*To go boldly where no man has gone before*,” ou “*To go where no man has gone before boldly*”. Essas últimas duas sentenças simplesmente não têm o mesmo efeito. (EDMONDSON, 2013¹⁸)

Este trabalho advoga que, se se considerar que *split infinitives* são construções que cabem no sistema da língua inglesa, não há motivo para considerá-las incorretas, o que seria uma postura excessivamente prescritiva. Em todo caso, para esta análise, a estrutura incomum tem a vantagem de tornar a frase em questão ainda mais singular.

¹⁶A construction consisting of an infinitive with an adverb or other word inserted between *to* and the verb, e.g., *she seems to really like it*. (Fonte: www.oxforddictionaries.com). Em português: Uma construção que consiste em um infinitivo com um advérbio (ou outra palavra) inserido entre o *to* e o verbo, por exemplo: *she seems to really like it*.

¹⁷ Neste caso, considera-se que a *Wikipedia* é uma referência relevante, já que *Star Trek* é uma série popular com informações amplamente disponibilizadas e discutidas no ambiente da internet.

¹⁸ No original: I think split infinitives are okay if used with caution. In the *Star Trek* statement, “*To boldly go where no man has gone before*,” the sentence would not have the same effect if it were worded differently. E.g., “*To go boldly where no man has gone before*,” or “*To go where no man has gone before boldly*”. These last two sentences just don’t have the same effect.

Um fato curioso sobre essa UF: há uma passagem na novela *The Dream-Quest of Unknown Kadath*, de H. P. Lovecraft, que contém uma frase muito parecida, o que leva alguns fãs a indagar se não seria essa uma possível fonte de inspiração:

At length, sick with longing for those glittering sunset streets and cryptical hill lanes among ancient tiled roofs, nor able sleeping or waking to drive them from his mind, Carter **resolved to go with bold entreaty whither no man had gone before**, and dare the icy deserts through the dark to where unknown Kadath, veiled in cloud and crowned with unimagined stars, holds secret and nocturnal the onyx castle of the Great Ones. (LOVECRAFT, 1943)

Independentemente da inspiração original que levou à formulação da frase em *Star Trek*, o fato é que a forma como está escrita na série é a que se tornou popular e se transformou em uma UF, em determinados contextos. A Figura 5 ilustra a distribuição de ocorrências de *where no man has gone* no *Google Books* a partir de 1900 (a ferramenta possui um limite máximo de cinco palavras por N-Grama):



Figura 5. Ocorrências de *where no man has gone* a partir de 1900. O eixo vertical do gráfico representa a porcentagem normalizada de ocorrências dessa frase em relação ao restante dos N-Gramas indexados pela ferramenta em um dado ano. Fonte: *Google Books Ngram Viewer*.

A interpretação desse gráfico pode ser feita mais diretamente do que o da frase anterior (Figura 4). Ocorrências da frase são praticamente inexistentes antes da metade da década de 60, quando *Star Trek* foi criado. O gráfico parece mostrar uma quantidade mínima de ocorrências em meados das décadas de 30 e 50. No entanto, uma análise de contextos pertencentes a essa época demonstrou que essas ocorrências aparecem erroneamente datadas no gráfico, pois se referem a um livro de 1989. Apesar dessas possíveis imperfeições da ferramenta, parece claro que se trata de uma UF que existe por influência da série. Um adendo: é importante lembrar que, apesar de sua grande extensão, o *corpus* do *Google Books* possui apenas livros e revistas em sua amostra. Um *corpus* com a mesma extensão, mas com maior variedade de gêneros

textuais, incluindo transcrições de língua falada, provavelmente forneceria evidências ainda mais consistentes.

Por fim, analisemos as características da frase *c*, “*Beam me up, Scotty*”. Apesar de ser uma frase de conhecimento geral, curiosamente, não há nenhuma ocorrência de uma construção usando exatamente essas palavras no *corpus* de estudo. Ao invés disso, aparecem exemplos como os seguintes:

29) **Beam me** home, Riker.

30) **Beam me up**, Mr. Spock.

31) Prepare to **beam up** party of four.

32) Two to **beam up**, **Scotty**.

33) Three to **beam up**, **Mr. Scott**.

Assim como as frases *a* e *b*, dentro da narrativa de *Star Trek*, essa frase tem um sentido literal. Comumente, na série, um grupo de personagens realiza missões na superfície de um planeta, enquanto sua espaçonave permanece em órbita. Ao fim da missão, os personagens na superfície usam a frase *beam up*, como nos contextos acima, para solicitar um teletransporte de volta à nave. Nesse contexto fictício, isso é feito por meio de um raio transportador (*beam*) emitido pela nave, que está acima do planeta (*up*). Na série original, “Scotty” é o engenheiro-chefe da espaçonave, e opera o equipamento do raio transportador.

No entanto, da mesma forma como ocorreu com as frases *a* e *b*, “*Beam me up, Scotty*” pode deixar de ter um significado literal e ser resignificada quando usada na língua corrente, tornando-se uma UF, como pode ser observado nos exemplos do COCA, a seguir:

34) Hornbeck took off his jacket and sat at the table. The hot kitchen smelled of old grease and Lysol, and now, of him. A covered aluminum pot rattled on the stove, steam jetting from the lid. "**Beam me up, Scotty**," he sighed.

35) This is the science of Star Trek. No other fictional science has so thoroughly weaved its way into the fibers of our culture. If you express frustration by saying "**beam me up, Scotty**," chances are darned good that the person you're talking to will know exactly what you're talking about.

O primeiro exemplo é uma narrativa durante a qual o personagem usa a frase como uma UF, para expressar frustração com a situação. O segundo exemplo é, na verdade, um contexto explicativo, que afirma justamente que a frase é comumente usada como uma UF para expressar frustração. Outro exemplo encontrado no *Google Books* segue essa mesma linha de sentido:

36) The minute it was over, I was wishing I were in another city. No, another country. **Beam me up, Scotty.**

Nesse exemplo, a UF foi usada para expressar o desejo de sair rapidamente ou desaparecer para evitar uma situação negativa.

Pode-se dizer que essa frase é a mais facilmente identificável como uma UF originada em *Star Trek* e ressignificada na língua corrente. Com efeito, “*Beam me up, Scotty*” é uma UF tão comum, que foi dicionarizada como uma expressão “usada figurativamente para expressar incredulidade ou um desejo de partir rapidamente” (PRUCHER, 2007, p. 13¹⁹). Seguindo o mesmo raciocínio das duas figuras anteriores, a Figura 6 demonstra a distribuição de ocorrências de *beam me up scotty* no *Google Books*:



Figura 6. Ocorrências de *beam me up scotty* a partir de 1900. O eixo vertical do gráfico representa a porcentagem normalizada de ocorrências dessa frase em relação ao restante dos N-Gramas indexados pela ferramenta em um dado ano. Fonte: *Google Books Ngram Viewer*.

Nesse caso, usando como base os dados do *Google Books*, a frase (aparentemente) passou a existir como UF na língua inglesa a partir da metade da década de 70, pouco depois do início de *Star Trek*. Esse e os demais dados são evidências que indicam que a frase seria uma UF originada na série.

¹⁹ No original: **beam me up (Scotty)** used figuratively to express disbelief or a desire to depart quickly.

Esta seção demonstrou exemplos de frases de *Star Trek* que, em maior ou menor medida, ultrapassaram o vocabulário da narrativa e passaram a ser usadas de forma corrente como UFs. Destaca-se a expressão “*Beam me up, Scotty,*” que foi ressignificada a ponto de se tornar uma expressão idiomática no inglês, demonstrando a influência do universo de discurso ficcional na língua corrente.

A abordagem adotada, nesse segundo momento, foi a *corpus-based*. Partiu-se do conhecimento prévio sobre as frases analisadas, e comprovou-se a hipótese (em alguns casos, apenas parcialmente) de que essas frases podem existir como UFs no léxico da língua inglesa. Dado tempo suficiente, um linguista poderia fazer uma análise semelhante com uma abordagem *corpus-driven*. No entanto, poderia ser necessária uma grande quantidade de tempo e análise (ou ajuda do acaso) até que as três frases aqui apresentadas emergissem dos dados, ou até que o pesquisador dirigisse seu olhar para elas, sem saber previamente de sua existência. A frase “*Beam me up, Scotty,*” especialmente, poderia demorar a aparecer nos dados, uma vez que essa frase não existe, formulada exatamente dessa maneira, no *corpus* de estudo.

6. Considerações finais

Este artigo enunciou as principais teorias relacionadas aos trabalhos com o léxico com o auxílio de *corpora*. Mais especificamente, buscou-se estabelecer um paralelo entre a base teórica da Etnoterminologia e a análise do vocabulário em uma obra de ficção científica, *Star Trek*, um objetivo que foi alcançado ao se demonstrar que o universo de discurso dessa série possui cultura e vocabulário próprios.

Seguindo as premissas da TCT, foi usado um *corpus* de estudo e *corpora* da língua inglesa geral para a obtenção de exemplos pragmáticos das unidades terminológicas de interesse. Os contextos de uso encontrados nesses *corpora* permitiram a identificação de fraseologismos no vocabulário de *Star Trek*, tanto a partir de uma abordagem *corpus-driven* quanto de uma abordagem *corpus-based*. O percurso utilizado para a obtenção dos dados buscou demonstrar que as duas abordagens podem ser usadas de forma complementar, havendo espaço tanto para a observação de dados empíricos encontrados de forma concreta quanto para a intuição individual e conhecimento prévio.

Neste ou em qualquer outro trabalho que faça uso de *corpora*, o caminho metodológico da Linguística de *Corpus* exige o manuseio de dados tanto de forma quantitativa, considerando

informações sobre o número de ocorrências e co-ocorrências de padrões lexicogramaticais, quanto qualitativa, examinando os dados e propondo interpretações.

Em suma, *Star Trek* provou ser um terreno fértil para os estudos do léxico, de forma que se propõe que mais estudos como este sejam feitos, a partir de universos de discurso de outras obras de ficção científica, uma vez que, pelo menos no caso dos *corpora* compostos por legendas de filmes e séries, as informações podem ser obtidas com relativa facilidade e de forma gratuita, e contrastadas com outros *corpora* de grande representatividade e também disponibilizados gratuitamente para consulta *on-line*. O linguista moderno possui ao seu dispor um número cada vez maior de ferramentas para a análise e descrição do seu objeto de estudo, a língua, garantindo vida longa e próspera aos estudos linguísticos.

Referências Bibliográficas

BANG, M.; FROMM, G. Terminologia em série: House M. D. In: **EntreLetras**, v. 4, n. 2. Araguaína: UFT, 2013. Disponível em: www.uft.edu.br/pgletras/revista/capitulos/8_terminologia_em_s%C3%A9rie.pdf. Acesso em 10 de dez. 2013.

BARBOSA, M. A. Para uma Etno-Terminologia: Recortes Epistemológicos. In: **Ciência e Cultura**, Campinas, v. 58, n. 2, p. 49-51, 2006.

BIDERMAN, M. T. C. Unidades complexas do léxico. In: Rio-Torto, G.; Figueiredo, O.M.; Silva, F. (Org.). **Estudos em Homenagem ao Professor Doutor Mário Vilela**. 1ªed.Porto, Portugal: Faculdade de Letras da Universidade do Porto, v. 2,p. 747-757, 2005.

CABRÉ, M. T. **La Terminologia**: teoría, metodología, aplicaciones. Barcelona: Antartida/Empuries, 1993.

CAVNAR, W. B., TRENKLE, J. M. **N-gram-based text categorization**. Michigan: Environmental Research Institute of Michigan, 1994.

DAVIES, M. **The Corpus of Contemporary American English**: 450 million words, 1990-present. 2008. Disponível em:<http://corpus.byu.edu/coca>. Acesso em 25 de ago. 2014.

EDMONDSON, R. **What is a Split Infinitive**: Grammar Guide.2013. Disponível em: http://hubpages.com/hub/Grammar_Mishaps__To_Boldly_Go_Where_No_Man_Has_Gone_Before. Acesso em 31 de ago. 2014

LOVECRAFT, H. P. The Dream-Quest of Unknown Kadath. In: **Beyond the Wall of Sleep**, Arkham House, 1943. Disponível em: http://en.wikisource.org/wiki/The_Dream-Quest_of_Unknown_Kadath. Acessoem: 01 de set. 2014.

MICHAEL, J. B., et al. Quantitative Analysis of Culture Using Millions of Digitized Books. **Science**, 2010.

MARSHALL, R. **The Origins of 11 Famous Star Trek Lines**. 2013. Disponível em: <http://mentalfloss.com/article/50607/origins-11-famous-star-trek-lines>. Acesso em 01 de set. 2014.

OLIVEIRA, L. H. M.; MÜLLER, A. F. A. Terminologia e a Utilização de Ferramentas Computacionais de Análise de *Corpus*. In: TAGNIN, S. E. O.; BEVILACQUA, C. (Org.) **Corpora na Terminologia**. São Paulo: Hub Editorial, p. 47-61, 2013.

PRUCHER, J. **Brave New Words: The Oxford Dictionary of Science Fiction**. Oxford: Oxford University Press, 2007.

SALES, R. Teoria Comunicativa da Terminologia (TCT) como Aporte Teórico para a Representação do Conhecimento Especializado. In: **VIII ENANCIB – Encontro Nacional de Pesquisa em Ciência da Informação**. Salvador, 2007. Disponível em www.enancib.ppgci.ufba.br/artigos/GT2--036.pdf. Acesso em 27 de fev. 2014.

SCOTT, M. **WordSmith Tools**. Versão 5. Oxford: Oxford University Press, 2008.

_____. **WordSmith Tools Help**. Liverpool: Lexical Analysis Software, 2014.

SILVA, M. B. Uma palavra só não basta: um estudo teórico sobre as unidades fraseológicas. In: **Revista de Letras**, Universidade Federal do Ceará, v. 1/2m n. 28, p. 11–20, 2006.

TAGNIN, S. E. O.; BEVILACQUA, C. **Corpora na Terminologia** (Introdução do livro). São Paulo: Hub Editorial, 2013.

TOGNINI-BONELLI, E. **Corpus Linguistics at Work**. Amsterdam e Philadelphia: John Benjamins, 2001. **crossref** <http://dx.doi.org/10.1075/scl.6>

WELKER, H. A. **Dicionários** - uma pequena introdução à lexicografia. 2. ed. Brasília: Thesaurus, 2004.

Artigo recebido em: 03.09.2014

Artigo aprovado em: 19.11.2014