

CÓRPUS DE TEXTOS CIENTÍFICOS DE TURISMO (TEXTur): APLICAÇÃO EM UMA PESQUISA TERMINOLÓGICA

Ivanir Azevedo Delvizio¹; Mariana Nascimento Jordão²; Yanae Pereira da Silva²

¹Doutora em Linguística, Universidade Estadual Paulista (UNESP), Faculdade de Engenharia e Ciências, Rosana.

²Graduanda em Turismo, Universidade Estadual Paulista (UNESP), Faculdade de Engenharia e Ciências, Rosana.

Resumo: O TEXTur é um córpus digitalizado de textos científicos (resumos) do campo do turismo extraídos de três periódicos nacionais e três internacionais de turismo, publicados entre 2018 e 2021. Este trabalho apresenta a aplicação do córpus TEXTur na realização de uma pesquisa terminológica que tem como objetivo a elaboração de um glossário bilíngue de termos usados no âmbito científico do turismo. O trabalho tem como objetivo descrever a etapa de levantamento e categorização dos termos em português recorrentes nos textos, de forma a fornecer dados linguísticos a serem utilizados na elaboração do glossário. Os procedimentos metodológicos da pesquisa pautaram-se nos preceitos da Terminologia e Terminografia (CABRÉ, 1993; BARROS, 2004; KRIEGER; FINATTO, 2004) e da Linguística de Córpus (BERBER SARDINHA, 2004). Para extração e seleção dos termos, utilizou-se o programa de análise lexical (análise de palavras) *WordSmith Tools 6.0*. Os termos foram classificados com base nas 17 categorias temáticas do Tesouro Brasileiro de Turismo (REJOWSKY, 2018). Como resultado, foram levantados cerca de 270 termos. Os termos foram classificados em 15 das 17 categorias temáticas do tesouro, mostrando uma grande compatibilidade temática entre os termos que compoem o glossário e os termos tratados no tesouro. Como contribuição e originalidade, o levantamento de termos diferenciou-se pelo número de termos inseridos na categoria Ciência e informação em turismo> Pesquisa, já que o TEXTur é composto por resumos científicos, sendo a descrição da metodologia uma parte fundamental. Como encaminhamento, a próxima etapa prevê a busca dos termos equivalentes em inglês no conjunto de resumos extraídos dos periódicos internacionais e a extração de dados necessários à elaboração do glossário bilíngue.

Palavras-chave: Linguística de Córpus; Texto Científico; Turismo, Terminologia.

Introdução

Em trabalho apresentado anteriormente, foram descritos os fundamentos e procedimentos que alicerçaram a construção do Córpus de Textos Científicos de Turismo (TEXTur), uma coletânea digitalizada de textos científicos (resumos), extraídos de três periódicos nacionais e de três periódicos internacionais de turismo, tendo como objetivo subsidiar a escrita científica.

Como implicações da construção do TEXTur, haviam sido enumeradas suas potencialidades de aplicação como: (1) recurso didático em disciplinas de metodologia e redação científica; (2) fonte de extração e observação de dados para realização de pesquisas linguísticas de diversas naturezas (3); insumo para o desenvolvimento de ferramentas computacionais de suporte à escrita científica e (4) fonte de consulta para alunos, docentes e pesquisadores de turismo e áreas afins. Neste trabalho, apresentamos o resultado da aplicação do TEXTur na realização de uma pesquisa linguística terminológica (aplicação 2), que teve como objetivo realizar o levantamento e a sistematização dos termos recorrentes nesse conjunto de textos científicos, sendo essa uma etapa preliminar para a elaboração de um glossário bilíngue de termos do âmbito científico do turismo.

O levantamento terminológico realizado a partir do córpus TEXTur baseou-se nos fundamentos teórico-metodológicos da Terminologia/Terminografia e da Linguística de Córpus. A Terminologia se refere à disciplina científica que estuda o conjunto de termos utilizados em um domínio específico do conhecimento (BARROS, 2004). Sua unidade básica de estudo é o termo, definido como “uma unidade lexical [palavra] com um conteúdo específico dentro de um domínio específico” (BARROS, 2004, p.39). A Terminografia é a dimensão aplicada da Terminologia, referindo-se à prática de elaboração de vocabulários especializados (dicionários e glossários).

As pesquisas terminológicas/terminográficas atuais usam como abordagem de estudo a Linguística de Córpus, segundo a qual a observação e a extração dos termos são feitas a partir de um córpus de análise, compreendido como um conjunto de dados linguísticos [textos], sistematizados segundo determinados critérios, “representativos da totalidade do uso linguístico ou de algum de seus âmbitos, que possam ser processados por computador, com a finalidade de



propiciar resultados para a descrição e análise linguística” (BERBER SARDINHA, 2004, p.18). Também é utilizado nesta pesquisa um *córpus* de referência, composto de textos de caráter genérico e de proporções maiores que serve de parâmetro de comparação para que sejam identificadas, por meio de recursos computacionais, as palavras-chave do *córpus* de análise (TAGNIN, 2010).

Para processar grandes quantidades de textos, existem alguns programas de análise lexical (análise de palavras). O programa utilizado nesta pesquisa foi o *WordSmith Tools 6.0*. (<https://www.lexically.net/wordsmith/>). Esse programa disponibiliza três ferramentas principais: *Wordlist* (Lista de palavras), *Keywords* (Palavras-chave) e *Concordance* (Linhas de concordância), que auxiliam no processo de análise do *córpus* e nas etapas de extração/identificação dos termos.

Outra etapa importante da pesquisa terminográfica é a organização dos termos em um sistema de conceitos, compreendido como um conjunto estruturado de conceitos organizados em classes e subclasses e que mantêm entre si uma série de relações (hierárquicas, associativas e de equivalência) (CABRÉ, 1993). Os tesouros especializados são ferramentas muito úteis para orientar a estruturação de um campo em uma pesquisa terminológica (CABRÉ, 1993). Assim, para fazer a classificação temática dos termos, adotou-se como guia a categorização proposta no Tesouro Brasileiro de Turismo (TBT) (<https://tesauroturismo.eca.usp.br/vocab/index.php>) (REJOWSKY, 2018). Esse tesouro contém 1702 termos do campo científico do Turismo, distribuídos em 17 categorias temáticas, subdivididas em até sete níveis de classes e subclasses, servindo para indexação de palavras-chave, recuperação da informação e controle terminológico.

Como a estruturação dos termos em um sistema de conceitos é determinada pelo *córpus* de análise e pela visão ou abordagem do pesquisador em relação ao domínio estudado (BARROS, 2004), cabem algumas considerações sobre o domínio do turismo. Como constatam Rejowsky e Barbanti (2018, p.184), a construção do conhecimento nas pesquisas na área do turismo “se nutre de contribuições vindas de diversas áreas que lhes fornecem conceitos, teorias e metodologias”. O TEXTur é composto por resumos de pesquisas na área do turismo e, por conseguinte, o conjunto terminológico dele extraído deverá apresentar, igualmente, termos referentes a conceitos, teorias e metodologias de diversas áreas. Desse modo, neste trabalho, compreende-se o turismo como um campo científico multidisciplinar (REJOWSKY; BARBANTI, 2018) e sua terminologia reflete essa multidisciplinaridade.

Metodologia

Trata-se de um estudo linguístico descritivo inserido na área de Linguística Aplicada e na linha de pesquisa Terminologia e Terminografia, tendo como abordagem a Linguística de *Córpus*.

Para a construção do TEXTur, foram selecionados três periódicos de turismo nacionais (Revista Brasileira de Pesquisa em Turismo, Revista Turismo em Análise, Turismo: visão e ação) e três internacionais (*Annals of Tourism Research*; *Tourism Management* e *Journal of Travel Research*) por meio do cruzamento entre a classificação Qualis Capes (2013-2016) e o resultado de duas pesquisas sobre ranqueamento de periódicos internacionais de turismo (PECHLANER et al., 2004; MCKERCHER; LAW; LAM, 2006 apud REJOWSKY, 2010). Dos periódicos selecionados, foram extraídos os resumos publicados em português, as respectivas traduções em inglês e os resumos publicados originalmente em inglês entre 2018 e 2021. Na etapa da pesquisa ora descrita, a extração de dados se deu a partir de um recorte do *córpus* referente ao conjunto de 317 resumos em português extraídos dos três periódicos nacionais.

Para a extração de palavras, utilizou-se o programa *WordSmith Tools*. Por meio da ferramenta *Wordlist*, gerou-se uma lista com todas as palavras contidas nos resumos. Para isso, inseriu-se no programa uma *stop list*, ou seja, uma lista, não exaustiva, de palavras (advérbios, artigos, conjunções, preposições etc.) a serem excluídas da lista final. A diferença de resultado com o uso de uma *stop list* pode ser observada nas figuras 1 e 2. A lista geral ainda continha 7.260 entradas (Fig. 2). O próximo passo foi reduzi-la a uma lista de palavras-chave. Isso foi feito por

meio da ferramenta *Keyword*, que a comparou com uma lista de palavras extraídas de um corpúsculo de referência (com 68.000 palavras), resultando desse processo uma lista composta de 434 palavras-chave (Fig. 3). Posteriormente, as palavras-chave foram analisadas por meio da ferramenta *Concordance* (Fig. 4), que exibe as linhas em que a palavra de busca ocorre, centralizando-a entre dois segmentos de texto. Essa configuração permite observar palavras que ocorrem juntas de forma frequente (co-ocorrências), indicando se tratar de um termo complexo (formado por duas ou mais palavras), por exemplo: patrimônio ambiental, patrimônio arqueológico, patrimônio cultural, patrimônio material, patrimônio imaterial etc.

N	Word	Freq.	%
1	DE	4.506	5,67
2	#	3.593	4,52
3	E	2.505	3,15
4	A	2.411	3,03
5	DO	1.491	1,88
6	O	1.485	1,87
7	DA	1.233	1,55
8	EM	1.043	1,31
9	QUE	1.034	1,30
10	TURISMO	1.014	1,28
11	PARA	809	1,02
12	OS	792	1,00
13	SE	743	0,93
14	NO	703	0,88
15	DOS	645	0,81
16	COMO	612	0,77
17	COM	587	0,74
18	NA	555	0,70
19	UMA	511	0,64
20	AS	504	0,63

Fig. 1. Lista de palavras total (7.947 palavras)

N	Word	Freq.	%
1	TURISMO	1.014	1,28
2	PESQUISA	334	0,42
3	ANÁLISE	326	0,41
4	ESTUDO	277	0,35
5	RESULTADOS	217	0,27
6	DADOS	205	0,26
7	MEIO	177	0,22
8	OBJETIVO	174	0,22
9	ARTIGO	173	0,22
10	DESENVOLVIMENTO	166	0,21
11	TURÍSTICA	159	0,20
12	TURÍSTICO	149	0,19
13	TURÍSTICOS	134	0,17
14	HOSPITALIDADE	132	0,17
15	GESTÃO	130	0,16
16	USO	119	0,15
17	RELAÇÃO	109	0,14
18	SERVIÇOS	105	0,13
19	BASE	104	0,13
20	TRABALHO	97	0,12

Fig. 2. Lista com *stop list* (7.260 palavras)

N	Key word	Freq.	%
1	TURISMO	1.014	1,28
2	PESQUISA	334	0,42
3	ANÁLISE	326	0,41
4	ESTUDO	277	0,35
5	RESULTADOS	217	0,27
6	DADOS	205	0,26
7	MEIO	177	0,22
8	OBJETIVO	174	0,22
9	ARTIGO	173	0,22
10	DESENVOLVIMENTO	166	0,21
11	TURÍSTICA	159	0,20
12	TURÍSTICO	149	0,19
13	TURÍSTICOS	134	0,17
14	HOSPITALIDADE	132	0,17
15	GESTÃO	130	0,16
16	USO	119	0,15
17	RELAÇÃO	109	0,14
18	SERVIÇOS	105	0,13
19	BASE	104	0,13
20	TRABALHO	97	0,12

Fig. 3. Lista de 434 palavras-chave.

Concordance

objetivos comuns de proteção do **patrimônio ambiental** e geológico do
 modelo de referência para a gestão do **patrimônio arqueológico**, podendo-se
 local. [PC] Turismo. Gestão do **patrimônio arqueológico**. Modelo de
 de; ARNOLD JUNIOR, M. Gestão do **patrimônio arqueológico** e
 presente artigo analisa a relação entre **patrimônio cultural** imaterial e turismo
 Culinária típica; Slow Food; **Patrimônio Cultural** Imaterial; PNLM.
 1984-4867.v29i2p312-331 [Int] Sob o **Patrimônio Cultural** Imaterial Culinário
 cultural como ativo econômico. [PC] **Patrimônio Cultural** Imaterial; Turismo
 olhar com mais acuidade para o seu **patrimônio edificado** e sua possível
 de preservação e conservação do **patrimônio edificado**. A pesquisa tem
 um corpus unificado para a oferta do **Patrimônio Edificado** para o Turismo
 , na conservação e na valorizaçãodo **patrimônio geológico** por meio do
 . Esses locais apresentam um **patrimônio geológico** único utilizado
 uma metodologia de quantificação do **patrimônio geológico**. A pesquisa
 tradicionais locais. [PC] Geoturismo; **Patrimônio geológico** e mineiro;
 mantém uma relação intrínseca com o **patrimônio geológico** e mineiro, como
 das comunidades tradicionais com o **patrimônio geomineiro**;
 de quilombolase garimpeiros com o **patrimônio geomineiro** do Vale e suas
 mantém uma relação intrínseca com o **patrimônio geomineiro** e que a prática

Fig. 4. Exemplos de linhas de concordância da palavra “patrimônio”

Na análise, também foram considerados alguns indicadores, observando-se: o grau de frequência da palavra, se ela designava um conceito especializado da área, se existiam contextos que a definissem ou explicassem, co-ocorrências, registro em dicionários/glossários/tesauros (BARROS, 2004; KRIEGER; FINATTO, 2004; CABRÉ), constituindo uma análise não só quantitativa, mas qualitativa. Importante ressaltar que foram consideradas as variações



denominativas e as formas abreviadas. Por fim, os termos foram inseridos em um quadro e agrupados segundo as categorias temáticas do TBT (REJOWSKY, 2018).

Resultados e discussões

Por meio dos procedimentos metodológicos descritos, foram levantados cerca de 270 termos, distribuídos em 15 das 17 categorias temáticas utilizadas no TBT: Alimentos e bebidas (05); Ciência e informação em turismo (65); Economia e turismo (72); Educação e formação em turismo (09); Eventos turísticos (01); Hospitalidade no turismo (07); Lazer e turismo (05); Organismos de Turismo (01); Patrimônio Turístico (32); Planejamento Turístico (13); Política de Turismo (09); Serviços turísticos (17); Sociedade e Turismo (18); TIC em Turismo (09); Transportes no turismo (04). Não foram selecionados termos que se enquadrassem em: Comunicação turística e Legislação turística. Os dados coletados foram organizados em um quadro, cujo link de acesso será disponibilizado durante a comunicação oral. Para ilustrar, seguem alguns recortes:

154.	acolhimento	Hospitalidade no Tur.
155.	acessibilidade	Hospitalidade no Tur.>Acolhimento
156.	hospitalidade	Hospitalidade no Tur.
157.	comensalidade	Hospitalidade no Tur.>Componentes da h.>Relações Interpessoais
158.	hospitalidade no meio urbano, hospitalidade urbana	Hospitalidade no Tur.>Componentes da h.>Relações Interpessoais>Urbanidade
159.	hóspede	Hospitalidade no Tur.>Figuras da h.
160.	anfitrião	Hospitalidade no Tur.>Figuras da h.

Fig. 5. Termos inseridos na categoria “Hospitalidade no Turismo”

Na primeira coluna, consta a numeração do termo. Na segunda, constam os termos em português selecionados do cópulus, seguidos de suas variações e/ou formas abreviadas separadas por vírgulas. Inseriu-se em primeiro lugar o termo recomendado pelo tesauro como de uso preferencial (p. ex.: destino turístico/destinação turística) ou, não havendo recomendação, o termo mais frequente no cópulus (p. ex.: turismo de natureza/turismo na natureza; turismo de experiência/turismo experiencial). Na segunda coluna, foram incluídos e destacados em azul os termos preferenciais indicados pelo Tesauro, mas não encontrados no cópulus (p. ex.: **hospitalidade no meio urbano**/hospitalidade urbana). Em relação à categorização dos termos, todos os termos levantados puderam ser classificados dentro das 17 categorias previstas no TBT. Dos 267 termos levantados, 112 constavam no tesauro (nesses casos, deixamos a categoria destacada em azul) (Fig. 5). Embora parte dos termos não constasse no tesauro, foi possível classificá-los seguindo-se sua lógica de categorias/subcategorias e níveis/subníveis. Nesses casos, não usamos destaque em azul (Fig. 6). Os termos encontrados no TEXTur ocuparam categorias tão variadas quanto àquelas previstas no tesauro. Essa compatibilidade entre o conjunto terminológico levantado e a proposta de estruturação dos termos do tesauro corroboraram a percepção de Rejowsky e Barbanti (2018, p.193) sobre “a amplitude e a complexidade do campo do turismo em suas variadas interfaces disciplinares” e a potencialidade do tesauro como “instrumento de apoio às pesquisas em turismo”, no caso, uma pesquisa linguística. Dos termos extraídos do TEXTur e não encontrados no tesauro, destacaram-se aqueles inseridos na (sub)categoria “Ciência e informação em turismo” e “Pesquisa”.

11.	amostra não probabilística, amostragem não probabilística	Ciência e Info. em Tur.> Pesquisa
12.	análise bibliométrica	Ciência e Info. em Tur.> Pesquisa
13.	análise comparativa, estudo comparativo	Ciência e Info. em Tur.> Pesquisa
14.	análise de conteúdo	Ciência e Info. em Tur.> Pesquisa
15.	análise de dados, análise dos dados	Ciência e Info. em Tur.> Pesquisa
16.	análise descritiva	Ciência e Info. em Tur.> Pesquisa
17.	análise do discurso	Ciência e Info. em Tur.> Pesquisa

Fig. 6. Recorte do grupo de termos inseridos na subcategoria “Pesquisa”.

Dos 65 termos da categoria “Ciência e Informação em Turismo”, apenas três constavam no tesouro (metodologia, tese, dissertação). Os demais referiam-se à metodologia de pesquisa. E foram abundantes pela própria característica do tipo de texto que compõe o TEXTur (resumo científico) e sua finalidade (escrita científica). Um dos principais movimentos retóricos do resumo científico é a descrição da metodologia e, portanto, isso se refletiu na terminologia extraída. A maior divergência em relação ao TBT referiu-se a termos que designavam segmentos turísticos. No TBT, alguns estão dispersos entre as categorias a que estão diretamente associados (turismo gastronômico e turismo de bebidas em “Alimentos e Bebidas”; turismo cultural em “Patrimônio turístico > Patrimônio cultural”) e outros estão agrupados (turismo de saúde, turismo de sol e praia, turismo de negócios e eventos e outros em “Economia > Marketing Turístico > Segmentação Turística”). Para padronizar, inserimos todos em “Segmentação Turística”. Esses foram os principais apontamentos em relação aos dados obtidos do processo de levantamento e categorização dos termos contidos no TEXTur.

Considerações Finais

Retomando o processo de construção do TEXTur e suas aplicações, neste trabalho descreveu-se a etapa de levantamento e categorização dos termos em português contidos nos resumos científicos extraídos de três periódicos nacionais de turismo. Como resultado, foram selecionados cerca de 270 termos, que puderam ser classificados em 15 das 17 categorias temáticas do TBT, demonstrando uma cobertura temática compatível entre os dois estudos e confirmando a “riqueza e abrangência do campo científico do Turismo” (REJOWSKY; BARBANTI, 2018). Como contribuição, o levantamento diferenciou-se por contemplar termos relativos a procedimentos metodológicos utilizados em pesquisas do campo do turismo. Em relação às limitações, registre-se que o levantamento de termos do TEXTur limitou-se aos resumos de três revistas nacionais no período de 2018 a 2021. Como o TEXTur é um corpus aberto, prevendo a inserção de novos textos, aponta-se a necessidade de sua ampliação com textos a partir de 2022 em diante e da realização de novos levantamentos, acompanhando a produção científica e a evolução terminológica da área. Por fim, reitera-se que o levantamento descrito é uma etapa preliminar à elaboração de um glossário bilíngue de termos do campo científico do turismo voltado à escrita científica.

Referências

- BARROS, L. A. *Curso Básico de Terminologia*. São Paulo: Edusp, 2004.
- BERBER SARDINHA, T. *Linguística de Corpus*. São Paulo: Manole, 2004.
- CABRÉ, M. T. *La terminología: teoría, metodología, aplicaciones*. Barcelona: Empúries, 1993.
- KRIEGER, M. da G.; FINATTO, M. J. B. *Introdução à Terminologia: teoria e prática*. São Paulo: Contexto, 2004.
- REJOWSKY, M. *Produção Científica em Turismo: análise de estudos referenciais no exterior e no Brasil*. *Turismo em análise*, v. 21, n. 2, p. 224-246, 2010.
- REJOWSKY, M. *Tesouro Brasileiro de Turismo*. São Paulo: ECA-USP, 2018.
- REJOWSKY, M.; BARBANTI C. H. *Construção de um Tesouro Brasileiro de Turismo*. *Turismo em Análise*, v. 29, n. 2, p. 182-195, 2018.
- TAGNIN, S. E. O. *Glossário de linguística de corpus*. In: VIANA, V.; TAGNIN, S. E. O. (org.). *Corpora no ensino de línguas estrangeiras*. São Paulo: HUB Editorial, 2010.