

Uma abordagem metodológica para a análise comparativa de comentários de viagens *on-line* de duas cidades património mundial da UNESCO – Coimbra (Portugal) e Salamanca (Espanha)

Filipa Perdigão Ribeiro^{*}

Nuno António^{**}

Marisol B. Correia^{***}

Resumo: Apresenta-se uma proposta metodológica para a análise comparativa de comentários de viagens *on-line* (CVO), ao nível do destino e da língua. Optou-se por uma abordagem exploratória das cidades universitárias e património mundial da UNESCO (Coimbra, Portugal; Salamanca, Espanha), assumindo como ponto de partida que a classificação da UNESCO e as universidades medievais são fatores (1) comparáveis e (2) fundamentais para a atração de turistas. Com base na recolha de 8.638 CVO publicados no TripAdvisor em português, espanhol e inglês (2017-2018) sobre dez locais de cada cidade, propõe-se uma metodologia mista de técnicas de análise quantitativa e qualitativa. Concluiu-se que a reputação UNESCO não parece ser relevante para os visitantes-comentadores; são, contudo, distintos os aspetos realçados, de acordo com a língua. Este estudo permite destacar as vantagens da aplicação de vários métodos de análise num estudo comparativo e demonstrar a riqueza informacional dos CVO como instrumento para (re)posicionar o destino.

Palavras-chave: Análise comparativa. eWOM. Património UNESCO. Processamento de linguagem natural. *Text mining*.

Abstract: This chapter presents a methodological approach for the comparative analysis of online travel reviews (OTR), by destination and by language. As a starting point we assume that the two cities of Salamanca (Spain) and Coimbra (Portugal), both with medieval universities and listed as UNESCO heritage, are comparable as tourist attractions. Based on the collection of 8,638 OTR published on TripAdvisor in Portuguese, Spanish and English (2017-2018) on ten sites in each city, this study applies a mixed-methods framework entailing quantitative and qualitative analysis techniques. Results show that the UNESCO's reputation does not seem to be relevant to visitors-commentators; however, the issues given emphasizes are different according to language. This study highlights the advantages of applying a mixed-method analytical approach to a comparative study and reveals the information value of OTRs as a tool to (re) position the destination.

Keywords: Comparative analysis. eWOM UNESCO heritage. Natural language processing. Text mining.

^{*} ESGHT- Universidade do Algarve & Centro de Investigação, Desenvolvimento e Inovação em Turismo – CiTUR, Portugal; *E-mail:* fperdig@ualg.pt

^{**} NOVA IMS, ISCTE-IUL & ESGHT - Universidade do Algarve; Centro de Investigação, Desenvolvimento e Inovação em Turismo – CiTUR, Portugal; *E-mail:* antonio.nuno@gmail.com

^{***} ESGHT - Universidade do Algarve & Centro de Investigação, Desenvolvimento e Inovação em Turismo – CiTUR & Centro de Investigação em Turismo, Sustentabilidade e Bem-estar - CinTurs, & CEG-IST, Instituto Superior Técnico, Universidade de Lisboa, Portugal; *E-mail:* mcorreia@ualg.pt

1 Introdução

Instituição por onde passaram ilustres figuras da nossa História, sendo também uma das mais antigas universidades do mundo [...] Infelizmente notei muitos atendentes malcriados (seriam eles estudantes?), que parecem estar detestando o trabalho. Na porta da capela de São Miguel, por exemplo, fica um atendente atrás da porta que respondia atravessado a todas as pessoas. Coimbra passou uma péssima impressão em termos de cordialidade, passando a ser um destino totalmente dispensável... (mulher, Universidade de Coimbra, 2.03.2018, Avaliação: 4).¹

O Comentário de Viagem Online (CVO) que abre este capítulo comprova o crescimento e prevalência das plataformas Web 2.0, tais como redes sociais ou aplicações móveis, e encapsula as contradições inerentes a um sistema de avaliação da experiência de visita a um local que inclui uma apreciação quantitativa (avaliação) e uma apreciação textual. No caso deste comentário, verificamos que a avaliação foi de 4 (numa escala de 1 a 5, 4 situa-se no bom / muito bom), mas o texto indica que “Coimbra passou uma péssima impressão em termos de cordialidade”. Consequentemente, os decisores que se centrem nas avaliações quantitativas terão uma visão da realidade diferente da daqueles que tenham acesso à informação textual.

Com base na recolha e exploração de dados disponíveis na rede social TripAdvisor, o objetivo geral deste trabalho é analisar o que os visitantes querem e comentam, de forma a que os responsáveis pela gestão do património possam vir a tomar decisões informadas. Como objetivo específico, apresentamos uma metodologia mista exploratória para a análise comparativa de um *corpus* de CVO (em inglês, *on-line travel reviews* – OTR) publicados pelos visitantes das cidades património da UNESCO – Coimbra, Portugal e Salamanca, Espanha, em 2017 e 2018.

Estudos apontam para os impactos significativos nas empresas, nas organizações e nos *stakeholders* do turismo patrimonial e cultural da prevalência da internet participativa, interativa e dinâmica, centrada no consumidor (TSCHEU; BUHALIS, 2016). Do ponto de vista da procura, estas tecnologias permitem ao turista aceder a mais informação e mais conhecimento e ajudam-no a controlar muitos aspetos da sua viagem (NDOU, 2011; DE ASCANIIS; CANTONI, 2017). Como consequência, as plataformas participativas transformaram de forma radical a forma como o turista percebe e se envolve com o património (EUROPEANA, 2015). Um dos aspetos fundamentais deste envolvimento é o facto de os turistas utilizarem as novas tecnologias para partilhar de forma constante as suas experiências e impressões dos

¹ Todos os Comentários de Viagens Online (CVO) reproduzidos neste capítulo reproduzem *ipsis verbis* o texto original, incluindo gralhas na ortografia.

locais que visitam, competindo desta forma com os materiais promocionais e influenciando fortemente o processo de decisão (DU CROS; MCKERCHER, 2015; HAUSMANN; WEUSTER, 2018).

Na justaposição do turismo cultural e patrimonial e da Web 2.0, há ainda muita investigação teoricamente robusta por desenvolver (HAUSMANN; WEUSTER, 2018). De facto, há uma lacuna informativa entre o que os visitantes desejam e o que vêm a encontrar nos locais. Adicionalmente, da perspetiva da gestão e do marketing é necessário compreender quem são os turistas do Património Mundial da Humanidade da UNESCO, de forma a oferecer experiências positivas e simultaneamente apoiar a conservação dos lugares (ADIE; HALL, 2016). Assim, uma das questões de base na investigação turística sobre espaços históricos urbanos é perceber o que motiva os indivíduos e famílias a visitarem as cidades patrimoniais, a partir dos seus próprios testemunhos ou comentários de viagens *on-line* (CVO), investigação que, até à data, não parece ter sido conduzida, quer em Portugal, quer internacionalmente. Por outro lado, estudos comparativos, que vão para além de contextos locais circunscritos, são essenciais para confrontar e desvendar as semelhanças do que muitas vezes “se julga serem os particularismos histórico-culturais e a singularidade dos patrimónios locais” (FORTUNA; GOMES, 2014, p. 275). Assim, conforme anunciámos acima, o objetivo geral deste estudo é salientar a importância de se conhecer em profundidade quem são os visitantes e o que eles pretendem, para que os gestores do património possam optar pelas melhores decisões, com na recolha e exploração de dados disponíveis na rede social TripAdvisor. De salientar que, em 2009, a UNESCO e o TripAdvisor lançaram uma parceria com o objetivo de mobilizar o apoio dos visitantes para preservar locais naturais e culturais inscritos na Lista do Património Mundial da UNESCO (UNESCO, 2009).

Do ponto de vista das técnicas e dos métodos de pesquisa e de análise, tem havido interesse crescente em abordagens metodológicas mistas, que incentivam a interdisciplinaridade, nos estudos em geral (WALKER; BAXTER, 2019) e na área do Turismo e da Hotelaria (MOLINA; FONT, 2016). Contudo no que respeita à investigação com grandes quantidades de dados textuais coligidos, a partir das redes sociais e de outras formas de mediatização computadorizada, parece haver um efetivo domínio dos métodos quantitativos, o que implica menor riqueza informativa e a compreensão menos holística do fenómeno em análise (MASON, 2006). Deste modo, o objetivo específico deste capítulo é apresentar uma metodologia mista exploratória para a análise comparativa de um *corpus* de CVO, publicados pelos visitantes das cidades património da UNESCO – Coimbra, Portugal e Salamanca, Espanha, em 2017 e 2018, com recurso aos métodos de *text mining*, processamento de linguagem natural e

análise de sentimentos, que permitem a abordagem quantitativa e qualitativa dos textos.

O estudo foi orientado por três perguntas de investigação principais e três perguntas secundárias:

P1 – Que instrumentos e técnicas de investigação são apropriadas para explorar CVO de sítios urbanos classificados como património mundial UNESCO?

P2 – Será a classificação património UNESCO ou património da humanidade um fator de atração?

P3 – Quais são as perceções dos visitantes?

P3.1 Que tipo de apreciação fazem?

P3.2 O que leva as pessoas a visitar Coimbra e Salamanca?

P3.3 Os visitantes comparam as duas cidades?

As duas cidades apresentam características comuns – ambas são cidades universitárias desde a fundação dos Estudos Gerais nas duas cidades no século XIII, são centros políticos e religiosos fundamentais na história dos respetivos países, sendo que a sua atratividade turística depende fundamentalmente destes fatores (GOMES, 2012) e estão classificadas como cidades património da UNESCO: Salamanca desde 1998 e Coimbra desde 2013.

As cidades históricas e monumentais tornaram-se uma das principais atrações turísticas a partir dos anos 1980. Tanto pela importância da cultura visual, como pelo fator histórico, a sua arquitetura histórico-monumental converteu-se num elemento da identidade local e, também, no recurso que mais interesse suscita nos turistas (GOMES, 2012) e que pode ainda instigar a competição entre destinos, locais e cidades (PEIXOTO, 2001). Dada a posição geográfica de Salamanca e de Coimbra, que distam apenas 304 km (uma viagem de carro de pouco mais de três horas), e das características acima referidas, as cidades podem ser consideradas concorrentes diretas em termos de destino, uma vez que apresentam semelhanças entre os respetivos perfis e posições no mercado turístico internacional (GOMES, 2012). Assim, justifica-se a escolha destas duas cidades para desenvolver uma análise comparativa, tanto mais que alguns estudos apontam para fluxos turísticos entre as duas cidades (GOMES, 2012). Por outro lado, viagens a destinos múltiplos na Europa, embora não sendo um fenómeno novo – a *Grand tour* e o Inter-rail são exemplos deste conceito de viagem – são cada vez mais frequentes e acessíveis a segmentos significativos da população, não só europeia, mas também da Ásia e da América. Outras alterações no padrão de viagens incluem um maior índice de viagens individuais ou planeadas de forma independente – por oposição a grupos –, a imersão numa região específica e o

crescimento de viagens temáticas (LARSSON; REIBRING, 2018). Estas tendências são particularmente facilitadas pelas ferramentas de planeamento de viagem *on-line* e naturalmente pelos CVO.

Os CVO são indubitavelmente cruciais na tomada de decisão para a escolha de visita a um destino. Os consumidores, mais do que nunca, apoiam-se em estratégias de escolha através da internet e para o consumo turístico este meio é absolutamente essencial nas decisões de compra, de escolha dos destinos e de escolha do que visitar no próprio destino (DE ASCANIIS; CANTONI, 2017). Com efeito, estudos recentes da área do *marketing* têm-se dedicado a explorar a influência dos CVO nos futuros viajantes ou turistas, na perspetiva dos que viajarão e de que forma interpretam e assimilam esses comentários (SÁNCHEZ-JIMÉNEZ; CORREIA; MATOS, 2018). Menos frequentes são os estudos focados nos comentários em si. Na perspetiva do produto turístico e da mensuração, a maior parte das análises tem-se centrado nos CVO relativos a hotéis, publicados no Booking.com ou no TripAdvisor, mas não há muitos estudos dedicados aos comentários sobre os locais visitados.

Este capítulo segue a seguinte estrutura: a próxima secção é dedicada ao enquadramento do património listado da UNESCO e às duas cidades em análise. Em seguida, apresenta-se uma revisão breve sobre a Web 2.0, o eWOM (*Electronic Word of Mouth*), a rede social TripAdvisor e a sua importância para os turistas e viajantes contemporâneos. Na terceira secção apresenta-se a metodologia e os vários métodos propostos são ilustrados com alguns resultados. Por fim, apresentam-se as secções da discussão dos resultados e das conclusões.

2 Classificação de património mundial da humanidade UNESCO, Coimbra e Salamanca – Património UNESCO

O património urbano é o tipo de património mais frequente na Lista do Património Mundial, representando mais de 53% dos locais culturais inscritos somando 428 propriedades. Entre estas, 241 estão listadas como cidades históricas e 187 aparecem como propriedades em contexto urbano (UNESCO, 2016). Um dos objetivos da UNESCO é implementar experiências de visita únicas e simultaneamente criar benefícios sociais e económicos para os próprios lugares (UNESCO, 2016).

Vários estudos mostram que o património listado provoca aumento de fluxo turístico e, muitas vezes, mais apoio por parte dos governos nacionais e locais, para preservação e/ou manutenção (VANBLARCOM; KAYAHAN, 2011), embora estes apoios possam variar de acordo com os diferentes contextos nacionais. Contudo, há também outros impactos, tais como alterações no acesso aos locais ou utilizações distintas dos

espaços. Assim, torna-se necessária a avaliação sistemática dos impactos da patrimonialização UNESCO para gerir os locais, nomeadamente a construção de indicadores apropriados, uma vez que as candidaturas a Património Mundial da UNESCO, especialmente a partir dos anos 1990, tinham como objetivo principal o aumento de visitantes e de receitas do turismo (VANBLARCOM; KAYAHAN, 2011). Contudo, os estudos que medem quantitativamente os impactos económicos parecem ser inconclusivos (VANBLARCOM; KAYAHAN, 2011; KAYAHAN; VANBLARCOM, 2012). Em contraste, outros estudos salientam o papel limitado da classificação da UNESCO e defendem que o impacto de se ser membro da lista do património mundial depende de outros aspetos contextuais. Por exemplo, Prud'homme (2008) concluiu que há uma perceção exagerada dos impactos para o desenvolvimento local. Resultados do relatório da Research Consulting Ltd. & Trends Business Research Ltd. (2009) comprovam as conclusões de Prud'homme e indicam que 70% a 80% dos lugares parecem estar a fazer pouco ou nada com a designação para obterem impactos económicos significativos. Em Coimbra, notaram-se impactos imediatos no aumento das visitas à Universidade de Coimbra, mas os empresários locais não são unânimes nos benefícios auferidos (SOLDADO, 2015).

Entre os vários benefícios importantes que acompanham a classificação, destacam-se o prestígio associado à classificação, o apoio em termos financeiros e o acesso a conhecimento especializado para a preservação e conservação. Adicionalmente, o prestígio funciona como catalisador para aumentar a preocupação com a preservação do património por parte dos governos e dos cidadãos (UNESCO, 2000). O valor da marca Património Mundial da UNESCO parece ser relevante em certos países e comunidades. Em países como Portugal, por exemplo, os processos de candidatura e de posterior classificação são acompanhados de perto pelos media nacionais e revestem-se de importância significativa para a valorização da identidade cultural e nacional.

2.1 Coimbra e Salamanca

Coimbra, situada nas margens do rio Mondego e localizada na Região Centro de Portugal, tem cerca de 134.156 habitantes (PORDATA, 2018a). Números de 2017 indicam 430.076 de dormidas (PORDATA, 2018b). Os visitantes-tipo de Coimbra são maioritariamente de origem portuguesa (53,8%), seguidos de residentes em Espanha (13,6%) e em Itália (6,2%) (FORTUNA; GOMES, 2014). Em 2017, a universidade registou mais de 500.000 visitantes, dos quais 20% franceses, 10% brasileiros, seguidos dos espanhóis, italianos e portugueses (NOTÍCIAS DE COIMBRA, 2018, 4 janeiro).

Salamanca é uma cidade média, do interior peninsular e conta com 144.949 habitantes (INE, 2017a). A partir da sua integração em 1988 na lista de Património Mundial da UNESCO e da sua posterior designação como Cidade Europeia da Cultura (2002), cresceram as iniciativas para o desenvolvimento do turismo, com base no potencial turístico que se apoia no valor singular do urbanismo, da arquitetura, da história e da cultura, de heranças romana, medieval, renascentista e barroca, e do qual sobressaem a catedral, a Plaza Mayor e a Universidade (BERZOSA-PAREJA, 2018). Em 2017 as dormidas cifraram-se nos 1.103.176 (INE, 2017b).

2.2 Web 2.0 e Electronic Word of Mouth (eWOM)

A Web 2.0, assim “batizada” em 2004, refere-se à transição que ocorreu entre as páginas estáticas em HTML que surgiram com a Web 1.0, para as páginas web dinâmicas e interativas. A Web 2.0, também apelidada de Social Web ou Social Era, consiste numa internet participativa, interativa e dinâmica, centrada no consumidor, cujos conteúdos são gerados pelos utilizadores (CGU), ou seja, os utilizadores da informação são também os consumidores (O’REILLY, 2007). A participação dos consumidores, no papel de *prosumers*, isto é, como consumidores e produtores de informação ativos, está a resultar em CGU e que permitem explorar novas dimensões da comunicação. De realçar que os CGU são considerados mais confiáveis do que as fontes oficiais, porque são considerados genuínos e não voltados para os negócios (DE ASCANIIS; CANTONI, 2017).

Os comentários de viagens online ou CVO são uma forma de CGU que refletem as avaliações e os comentários sobre a experiência do próprio visitante, bem como sobre o destino propriamente dito. O CGU pode igualar o eWOM (*Electronic Word of Mouth*) e são cruciais na formação da imagem e na reputação dos destinos (DE ASCANIIS; CANTONI, 2017). Quando um destino excede as expectativas dos visitantes, é provável que estes se sintam motivados a partilhar a sua experiência positiva com outras pessoas. Por outro lado, os visitantes desapontados podem usar o eWOM negativo para espalhar as suas emoções negativas.

A influência interpessoal e a eWOM são classificadas como as fontes de informação mais relevantes quando o consumidor se encontra no processo de decisão para efetuar uma compra. Essa influência pode ser especialmente importante no setor do turismo e da hotelaria, já que os seus produtos intangíveis são difíceis de avaliar antes do seu consumo (DIAS; RIBEIRO; CORREIA, 2013). As plataformas colaborativas e interativas, *i.e.* plataformas 2.0, oferecem um meio acessível – onde as pessoas podem expressar as suas opiniões e propostas sobre qualquer assunto e onde podem tornar-se cocriadores – gerando um espaço de cocriação entre entidades e públicos

(IGLESIAS-SÁNCHEZ; CORREIA; JAMBRINO-MALDONADO, 2019). O marketing viral é especialmente persuasivo no setor turístico e hoteleiro, uma vez que os utilizadores das plataformas 2.0 estão mais dispostos a aceitar e a confiar nas informações de pessoas que são semelhantes a elas. Assim, o marketing viral permite aos CVO atingir um público mais amplo (SÁNCHEZ-JIMÉNEZ; CORREIA; MATOS, 2018).

Atualmente, a utilização generalizada das plataformas 2.0 está a provocar alterações radicais na promoção dos destinos turísticos, através da estratégia clara de incorporação de CGU. Consequentemente, os CVO representam uma fonte de informação muito valiosa para as agências de viagens e para os *Destination Management Organizations* (ou Organizações de Gestão dos Destinos). Através da análise dos comentários e avaliações dos visitantes, as Organizações de Gestão dos Destinos estão a ajustar-se, diferenciando a promoção dos destinos turísticos e adaptando-se aos atuais desafios colocados por este novo paradigma de interação e comunicação (ROQUE; FERNANDES; RAPOSO, 2012).

Com a Web 2.0 os turistas ou visitantes apresentam CVO cada vez mais documentados, verificando-se um processo *on-line* de *story-telling* que parece ajudá-los a melhorar as suas experiências. A título ilustrativo podemos referir os estudos desenvolvidos por Rahmani, Gnoth e Mather (2018) que utilizam a Linguística de Corpus (na área das TIC a designação utilizada para esta técnica está relacionada com o Processamento de Linguagem Natural e o *Text Mining*, ver abaixo) para extrair os significados “ocultos” nos CVO e identificar as suas características, aspetos explorados abaixo.

O TripAdvisor, apresentado em 2000 e sendo uma das plataformas 2.0 mais utilizadas atualmente, opera como guia de viagens *on-line* e oferece avaliações e comentários gerados pelos utilizadores sobre conteúdos relacionados com viagens. Em 2018, gerou aproximadamente 730 milhões de avaliações e opiniões de utilizadores, cobrindo mais de oito milhões de anúncios de restaurantes, hotéis, aluguer de férias e atrações; tem 490 milhões de visitantes únicos e envia aproximadamente 80 milhões de *e-mails* por semana (TRIPADVISOR, 2018a). Esta plataforma permite realizar pesquisas por hotéis, atividades, restaurantes, voos, cruzeiros, automóveis para alugar, entre outras categorias, sendo que as avaliações apresentam duas componentes, uma quantitativa (de 1 a 5) e ainda os comentários escritos pelos utilizadores. Neste estudo analisamos a categoria *atividades*, sendo dez relativas a Coimbra e dez a Salamanca. Na secção da metodologia, listamos os vários tipos de dados disponíveis na plataforma por utilizador.

O TripAdvisor, de entre os critérios para os avaliadores, menciona que as avaliações devem oferecer conselhos confiáveis de viajantes reais (TRIPADVISOR,

2019). Isto significa que os conteúdos criados pelos utilizadores devem ser experiências em “primeira mão” e dar uma contribuição substancial para a questão discutida.

3 Metodologia

3.1 Abordagem de métodos mistos

O recurso à pluralidade de métodos de diferentes áreas disciplinares (TIC, Turismo e Linguística) adapta-se bem a temas de investigação das ciências sociais que são subjetivos, complexos e multidimensionais, e que se encontram na fase exploratória inicial, como é o caso do presente estudo. Esta abordagem permite uma compreensão mais completa do fenómeno em análise, evidenciado várias perspetivas de análise e permitindo aceder à riqueza textual dos CVO de forma a retirar resultados diferentes, mas complementares (OLSEN, 2004). A complementaridade dos métodos utilizados também oferece um impacto mais prático e confirma a fiabilidade dos resultados (MOLINA; FONT, 2016). Embora haja diferentes definições de investigação com base em métodos mistos, todas concordam que esta abordagem implica a combinação de pelo menos um método quantitativo (cujo objetivo é recolher números) e um qualitativo (cujo objetivo é coligir, por exemplo, palavras) num único estudo (MOLINA; FONT, 2016). Nas secções seguintes apresentamos as várias técnicas aplicadas que têm como pressuposto o enfoque em estratégias para coligir e analisar dados quantitativos e qualitativos e os respetivos resultados.

3.2 Extração e descrição dos dados

Em função do objetivo do estudo e de modo a estabelecer comparações, decidimos selecionar para Coimbra e para Salamanca, respetivamente, os dez locais mais recomendados aos turistas por *sites* de informação turística ou por jornais digitais com secções dedicadas ao turismo.² Para efeitos de descrição e análise dos dados, utilizam-se os termos *visitante*, *turista* ou *utilizador* de forma indiferenciada.

O conjunto de dados analisados foi extraído do TripAdvisor em abril de 2019, através de um *web robot*, ou simplesmente *bot*, uma aplicação que executa tarefas de forma automática (DUNHAM; MELNICK, 2009). Com este *bot*, construído especificamente para o efeito na linguagem de programação C#, foram extraídos 8.638 CVO, relativos aos comentários publicados em 2017 e 2018. A frequência e distribuição

² Entre os vários portais consultados, apresentamos a título ilustrativo Duncan, F. (2012, 17 maio). “Salamanca, Spain: A Cultural guide”, *The Telegraph*. Disponível em: <https://www.telegraph.co.uk/travel/destinations/europe/spain/castile-and-leon/articles/Salamanca-Spain-a-cultural-city-guide/> e *Go to Portugal*, <https://gotoportugal.eu/pt/sitios-a-visitar-coimbra/>.

dos CVO extraídos por cidade e idioma pode ser vista na Tabela 1, já a frequência de cada um dos dez locais por cidade e por idioma pode ser observado na Figura 4.

Tabela 1 – Frequência e distribuição de CVO por cidade e idioma

		Inglês	Espanhol	Português	TOTAL
Coimbra	Frequência	884	465	1.643	2.992
	Distribuição (em relação ao total)	10,2%	5,3%	19,0%	34,6%
	Distribuição (em relação à cidade)	29,6%	15,5%	54,9%	100,0%
Salamanca	Frequência	1.162	3.897	587	5.646
	Distribuição (em relação ao total)	13,5%	45,1%	6,8%	65,4%
	Distribuição (em relação à cidade)	20,6%	69,0%	10,4%	100,0%
TOTAL	Frequência	2.046	4.362	2.230	8.638
	Distribuição (em relação ao total)	23,7%	50,5%	25,8%	100,0%

Fonte: Elaboração própria.

O conjunto de dados extraído e colocado num ficheiro em formato Excel é composto pelos seguintes campos:

- *City*: nome da cidade (Coimbra ou Salamanca);
- *FullText*: classificação qualitativa atribuída pelo utilizador ao local;
- *GlobalRating*: classificação quantitativa que cada local tem no TripAdvisor à data da extração. A classificação vai de 1 a 5, representado um 3 um valor médio e 5 um excelente (TRIPADVISOR, 2018b);
- *Language*: idioma em que foi escrito o comentário;
- *Location*: local de residência do utilizador (registo no TripAdvisor);
- *Name*: nome do utilizador. Para poder publicar conteúdos e comentários os utilizadores têm que criar um perfil no TripAdvisor. A este perfil deve ficar associado um nome/pseudónimo;
- *PublishDate*: data em que foi publicado o comentário;
- *ReviewRating*: classificação quantitativa atribuída pelo utilizador ao local;
- *SiteDesignation*: nome do local.

3.3 Análise dos dados

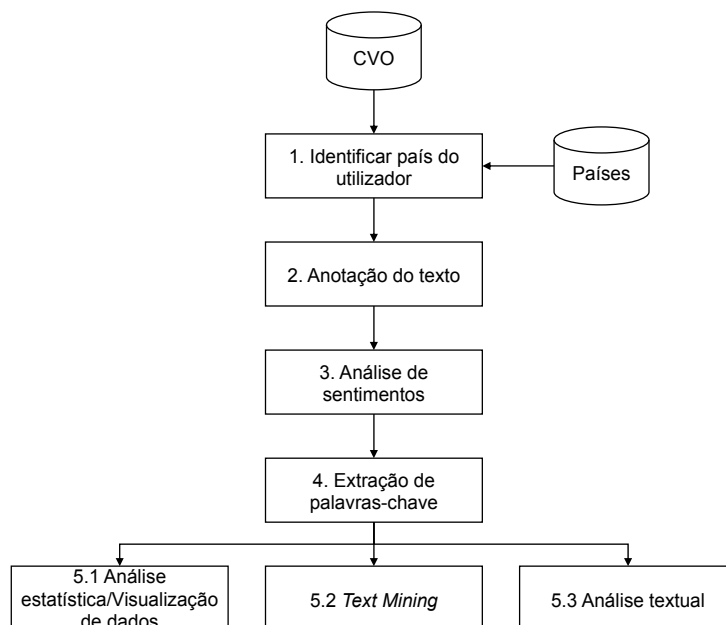
Apesar dos CVO terem duas componentes de avaliação, uma quantitativa e uma outra qualitativa (o texto escrito pelo utilizador) (EUROPEAN COMMISSION, 2014), a maioria da literatura sobre CVO incide apenas sobre a componente quantitativa. Mas, a verdade é que a componente qualitativa tem o potencial de fornecer uma visão mais rica dos CVO (DUAN; YU; CAO; LEVY, 2016). Por esse motivo, para se analisar os dados das duas componentes, decidiu-se recorrer a um conjunto diverso de métodos, oriundos de diversas áreas como o *Data Mining*, o Processamento de Linguagem

Natural, o *Text Mining* e a Análise Textual. Esta última técnica de análise é subsidiária das restantes e apoiou a interpretação dos dados.

Através da manipulação de bases de dados, aprendizagem automática, estatística e visualização de dados, entre outros métodos das ciências e tecnologias de informação, o *Data Mining* permite encontrar padrões em dados estruturados e extrair informação, transformando essa informação em conhecimento (ADEDYOIN-OLWE; GABER; STAHL, 2013; LIAO; CHU; HSIAO, 2012). Por sua vez, outros métodos como o Processamento de Linguagem Natural, o *Text Mining* e a Análise Textual permitem extrair informação e conhecimento de dados não estruturados, como é o caso da componente qualitativa dos CVO. Enquanto o *Text Mining* compreende um conjunto de técnicas que permitem caracterizar e transformar o texto, utilizando para isso as próprias palavras como unidade de análise (por exemplo, frequências, distribuição ou presença/ausência de palavras específicas), os algoritmos de Processamento de Linguagem Natural utilizam processamento sintático e/ou semântico, tendo como base regras ou métodos estatísticos para analisar, segmentar e extrair informação (FELDMAN; SANGER, 2007; MANNING; MANNING, 1999; NADKARNI; OHNO-MACHADO; CHAPMAN, 2011).

A análise dos dados englobou a realização de um conjunto de procedimentos sequencial, de modo a permitir a extração de informação. Este processo, representado na Figura 1, envolveu a aplicação de diversos métodos utilizando a aplicação R (R CORE TEAM, 2016).

Figura 1 – Diagrama do processamento e de análise de dados



Fonte: Elaboração própria.

O primeiro procedimento realizado foi a criação de um campo com o país de residência do utilizador que publicou cada comentário (procedimento 1 na Figura 1, Apesar de o conjunto de dados de comentários possuir o campo *Location*, este é um campo de texto livre, pelo que nem todos os utilizadores o preenchem com a designação do seu país de residência. Alguns utilizadores apenas preenchem o nome da cidade, outros complementam o nome da cidade com o nome do estado (comum em utilizadores oriundos dos Estados Unidos da América e do Canadá). Para além disso, mesmo em utilizadores que indicam o país de residência, em função do idioma utilizado para escrever o comentário, o país pode aparecer com três formas (exemplo: Espanha, España ou Spain). Assim sendo, para identificar o país de residência, o primeiro procedimento envolveu a aplicação de técnicas de Processamento de Linguagem Natural, para dividir o texto do campo *Location* em palavras. Depois, utilizando-se um outro conjunto de dados com a lista ISO-3166 (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2017) com designações de países em Português, Espanhol e Inglês, identificou-se qual o país do utilizador. Essa identificação foi feita em primeiro lugar com base na designação do país e quando tal não foi possível, por comparação contra uma lista de cidades e estados criada manualmente. Tal permitiu identificar o país em 5.675 comentários (aproximadamente 69% do total).

O procedimento seguinte (procedimento 2 na Figura 1) envolveu um conjunto de tarefas de anotação do texto como contagem de frequência de palavras, cálculo de coocorrências, entre outras, de modo a facilitar a exploração da componente

qualitativa dos comentários. Com a aplicação do pacote R “udpipe” (WIJFFELS, 2018) foi possível efetuar quatro tarefas comuns em Processamento de Linguagem Natural (MANNING; MANNING, 1999): (1) “tokenização” (dividir o texto em unidades – palavras e pontuação); (2) marcação de “partes da fala” (*part of speech tagging* – identificar a forma gramatical das palavras baseada na definição da palavra e no contexto); (3) “Lematização” – redução da palavra ao seu lema (ou seja, a forma canônica ou a forma tendencialmente mais comum da palavra); (4) Análise de dependência – análise da estrutura gramatical das frases, identificando as palavras “principais” e as relações das outras palavras com as palavras “principais”.

O terceiro procedimento (número 3 na Figura 1), envolveu a aplicação de análise de sentimentos aos textos. A análise de sentimentos (ou análise de opinião como por vezes também é chamada), é o estudo computacional das opiniões sobre entidades, indivíduos, eventos, tópicos e seus atributos; permite a quantificação de opiniões, de acordo com a sua polaridade (positiva, negativa ou neutra) (LIU; ZHANG, 2012). Este conhecimento da opinião dos utilizadores sobre a utilização de hotéis, restaurantes e atrações turísticas é considerado essencial para a criação de estratégias de serviço ao cliente e daí a sua cada vez maior utilização em turismo (GE; VAZQUEZ; GRETZEL, 2018). A análise de sentimento foi efetuada com o pacote de extensão do R, “Sentiment Analysis” (FEURRIEGEL; PROELLOCHS, 2018). Para os comentários em inglês foi utilizado o dicionário base do pacote, para espanhol o dicionário ElhPolar (SARALEGI; SAN VINCENTE, 2013) e, para português, o dicionário SentiLex-PT 02 (SILVA; CARVALHO; SARMENTO, 2012). A análise de sentimentos foi realizada por frases (tirando proveito da anotação de texto do procedimento anterior), aplicando-se o método “ruleSentimentPolarity”, o qual atribui a cada comentário um valor entre -1 e 1, sendo -1, muito negativo e 1, muito positivo. A pontuação de cada comentário foi calculada com base na média dos valores das frases que o compunham.

O quarto procedimento (número 4 na Figura 1) envolveu a aplicação de uma outra técnica de Processamento de Linguagem Natural que é a extração de palavras-chave – sequência de uma ou mais palavras que fornecem uma representação compacta do conteúdo de um documento (ROSE; ENGEL; CRAMER; COWLEY, 2010). Para extração das palavras-chave foi utilizada a implementação do algoritmo RAKE (ROSE *et al.*, 2010) existente no pacote R “udpipe” (WIJFFELS, 2018). De modo simplificado, pode-se dizer que este algoritmo seleciona as palavras-chave a partir de três passos: (1) seleção de palavras não delimitadas por “stopwords” (palavras sem carga semântica que não acrescentam significado); (2) construção de uma matriz de coocorrência de palavras; (3) cálculo de uma pontuação, de acordo com a coocorrência.

Após estes quatro procedimentos, procedeu-se à análise propriamente dita, recorrendo-se então à estatística, visualização de dados, *text mining* e análise textual. Os resultados desta análise são apresentados na secção seguinte.

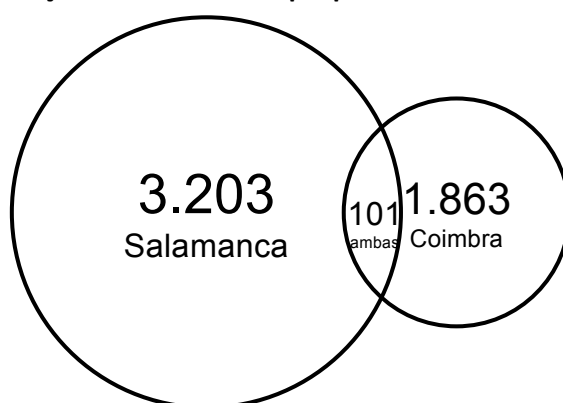
4 Resultados

4.1 Data Mining

Os 8.638 comentários foram escritos por 4.695 utilizadores, o que dá uma média de 1,7 comentários por utilizador, ou seja, houve utilizadores que publicaram mais do que um comentário. Efetivamente, apesar da mediana ser de um comentário por utilizador, o 3.º quartil é de dois comentários por utilizador, sendo que onze utilizadores publicaram dez ou mais comentários cada um. Houve mesmo um utilizador que publicou catorze comentários no período analisado, para os locais analisados. Este utilizador, residente em Espanha, registado como sendo alguém com entre 50 e 64 anos e que tinha visitado até ao momento 100 cidades, escreveu sempre em Espanhol, e tinha, à data da extração dos dados, um total de 471 comentários publicados no TripAdvisor. Dos catorze comentários que publicou sobre estas duas cidades, no período de extração, três comentários foram publicados sobre locais de Salamanca a 2/3/2018, um outro, também sobre Salamanca, foi publicado no dia 8 do mesmo mês. Depois, publicou um outro comentário sobre um local de Salamanca a 27/7/2018. Por fim, publicou nove comentários sobre locais de Coimbra nos dias 20 e 21/8/2018.

A análise dos dados por utilizador e cidade, como ilustrado na Figura 2, mostrou ainda que 101 utilizadores publicaram comentários sobre as duas cidades.

Figura 2 – Distribuição de utilizadores que publicaram comentários por cidade

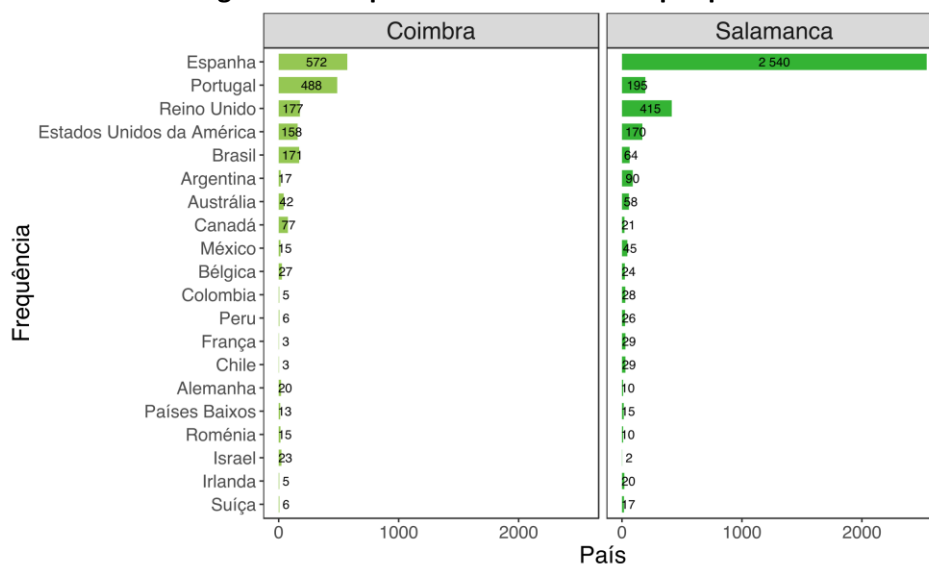


Fonte: Elaboração própria.

Tendo em conta o país de residência dos utilizadores (quando se conseguiu identificá-lo – como indicado anteriormente), é possível constatar, conforme ilustrado

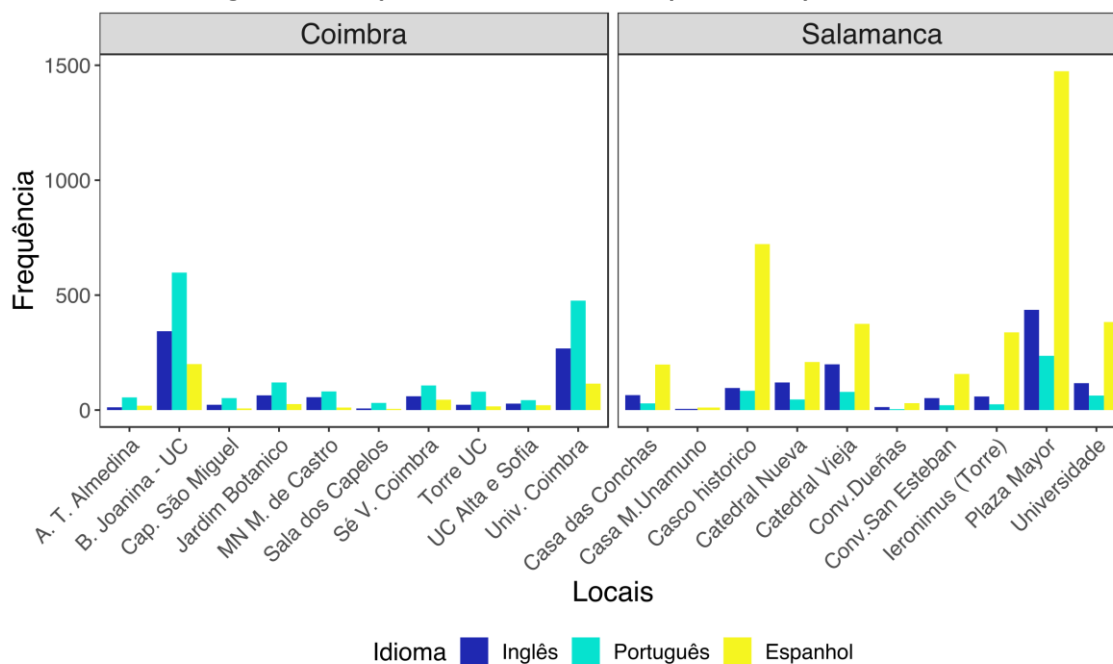
na Figura 3 (esta figura exhibe todos os países que possuam mais do que 20 comentários), que, como expectável, a maioria dos comentários é feita por utilizadores de Portugal e de Espanha. No entanto, também pela Figura 3, é possível verificar que existe uma distribuição distinta nas duas cidades. Enquanto que em Coimbra o número de comentários de estrangeiros é superior ao número dos comentários dos portugueses, sendo até que há mais comentários de utilizadores com residência em Espanha do que de utilizadores com residência em Portugal, para Salamanca os comentários de utilizadores com residência em Espanha são claramente superiores aos comentários dos utilizadores com residência noutros países. A Tabela 1 e a Figura 4, mostram que o número de comentários em espanhol (4.362) é praticamente o dobro do número de comentários em inglês e em português, respetivamente 2.046 e 2.230. Por sua vez, como apresentado na Tabela 2, existe também uma amplitude elevada na frequência de comentários por local. Destaca-se o caso de Salamanca, onde há um local apenas com dezasseis comentários, mas o mais visitado tem 2.146. Em Coimbra, apenas dois locais, a Biblioteca Joanina e a Universidade, têm mais comentários do que a média da cidade de todos os locais (299,2), respetivamente, 1.141 e 859. Em Salamanca, também só três locais têm mais comentários do que a média da cidade (564,6); são eles: a Plaza Mayor (2.146), o Casco Histórico (902) e a Catedral Vieja (653).

Figura 3 – Frequência de comentários por país



Fonte: Elaboração própria.

Figura 4 – Frequência de comentários por local e por idioma



Fonte: Elaboração própria.

Tabela 2 – Estatísticas de frequência de comentários por local

	Mínimo	1.º Quartil	Mediana	Média	3.º Quartil	Máximo
Coimbra	43,0	87,5	133,5	299,2	211,5	1.141,0
Salamanca	16,0	245,5	398,5	564,6	630,5	2.146,0

Fonte: Elaboração própria.

4.2 Processamento de linguagem natural e *text mining*

A aplicação de métodos de Processamento de Linguagem Natural e de *Text Mining*, na componente qualitativa dos comentários, permitiu também conhecer alguns resultados bastante interessantes. Por exemplo, a contagem de frequências revelou que, nos 2.992 comentários escritos sobre atrações de Coimbra, a palavra “Salamanca” foi mencionada quatro vezes. Por sua vez, nos 5.646 comentários escritos sobre atrações de Salamanca, a palavra “Coimbra” foi mencionada sete vezes. Abaixo, apresentam-se exemplos dessas referências, com indicação do sexo, cidade de residência e idade (quando disponíveis), local da visita e data do comentário:

Extrato 1

Una sorpresa esta universidad con toda la tradición intacta y una biblioteca espectacular [...] El lugar conserva todo el sabor de las universidades tradicionales. ¡No todo va a ser Oxford, Cambridge o Salamanca! No hay que perdérsela. (Homem, residente em Saragoça, Espanha, 50-64 anos, Universidade de Coimbra, 10.08.2017) [sublinhado nosso].

Extrato 2

Salamanca é uma cidade pequena e tão académica quanto a nossa cidade de Coimbra, em Portugal. Desde miúda que nutro uma paixão por esta cidade que, apesar de não tão monumental e turística, quando comparada com Madrid, é muito mais pacata, acolhedora e, de certo modo, sincera [...]. (Mulher, residente em Lisboa, Portugal, Plaza Mayor, 7.01.2018) [sublinhado nosso].

A análise de sentimentos, conforme resultados exibidos na Figura 5, mostra, tal como reconhecido em estudos anteriores sobre CVO, que existem diferenças nas componentes qualitativas dos CVO dependendo dos idiomas (tal como na componente quantitativa) (ANTONIO; ALMEIDA; NUNES; BATISTA; RIBEIRO, 2018b). Os utilizadores que escrevem comentários em português são os que, em média, produzem uma melhor avaliação quantitativa. No entanto, ao analisarmos o sentimento, reparamos que, por duas décimas, isso já não acontece em Salamanca. Aí, os utilizadores que escrevem em espanhol apresentam a melhor média. Em posição contrária, encontram-se os utilizadores que escrevem os comentários em inglês. Com exceção da avaliação quantitativa que fazem de Coimbra, os utilizadores que escrevem em inglês são os utilizadores que piores classificações médias atribuem. No entanto, estas médias devem ser analisadas com alguma prudência, pois, como mostrado por Antonio, Almeida, Nunes, Batista e Ribeiro (2018a), a análise de sentimentos permite identificar situações que mostram discrepâncias entre avaliação quantitativa e avaliação qualitativa. Vejam-se, por exemplo, dois comentários onde as duas avaliações apresentam resultados antagónicos:

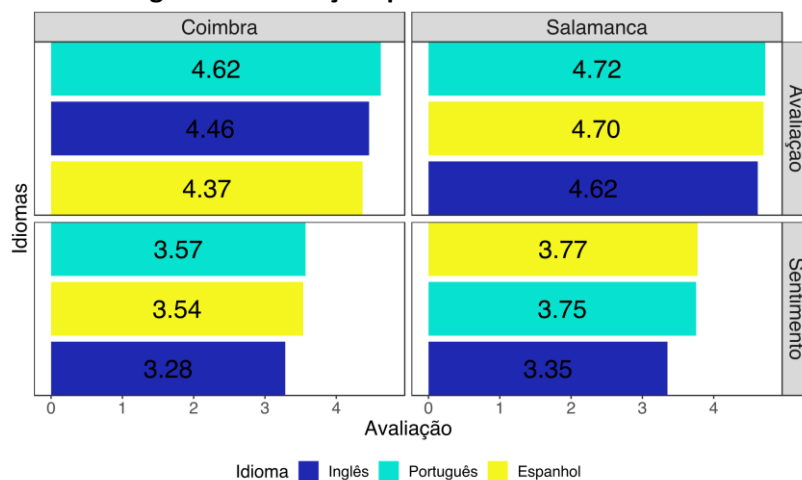
Extrato 3

Tenia muchas ganas de fotografiar por la noche el puente romano y la catedral, cuando acudí comenzaba la iluminacion nocturna, una mezcla de luz increíble (residente na Provincia de Ciudad Real, Espanha; Catedral Nueva, Salamanca, 15.05.2017. Avaliação: 1, Sentimento: 5).

Extrato 4

O trabalho da fachada da universidade é lindíssimo, mas você ficará parado ali procurando a rã (eu achei) e o astronauta (cansei de ficar de pé e procurar). É visita obrigatória (residente em Caxias do Sul, Brasil; Universidade, Salamanca. Avaliação: 5, Sentimento: 1).

Figura 5 – Avaliação quantitativa e sentimento



Fonte: Elaboração própria.

Nota: O sentimento foi normalizado de modo a estar na mesma escala da avaliação quantitativa (1 a 5).

Figura 6 para os comentários escritos em Português sobre os locais da cidade de Coimbra – permite identificar as correlações das palavras mais frequentes numa mesma frase (nomes e adjetivos), bem como o sentimento das frases onde ambas as palavras constam. No gráfico, a cor da barra junto a cada palavra indica qual a frequência da palavra no texto (clara – pouco frequente, escura – muito frequente). O gráfico representa as palavras com duas ou mais ocorrências nos comentários. Os arcos ligam as palavras que aparecem conjuntamente numa mesma frase, sendo que a largura dos arcos varia de acordo com a correlação das palavras (entra 0 e 100%, sendo que 100% significaria que as duas palavras sempre que são utilizadas aparecem na mesma frase). Neste caso, foi definido que só seriam exibidas palavras com uma correlação superior a 5%. A cor dos arcos indica o sentimento das frases onde ambas as palavras foram utilizadas (verde – sentimento positivo; azul – sentimento neutro; vermelho – sentimento negativo). Este gráfico permite alertar para relações não expectáveis e assim levar à exploração dos comentários onde essas relações aparecem. Os extratos abaixo ilustram essa correlação:

Extrato 5 “Estudantes” e “Turistas”:

[...] muito estudantes a vender souvenirs para pagarem as suas contas e muitos turistas (homem, Biblioteca Joanina, Coimbra, 11.08.2018).

Extrato 6 “Sala” e “Miguel”:

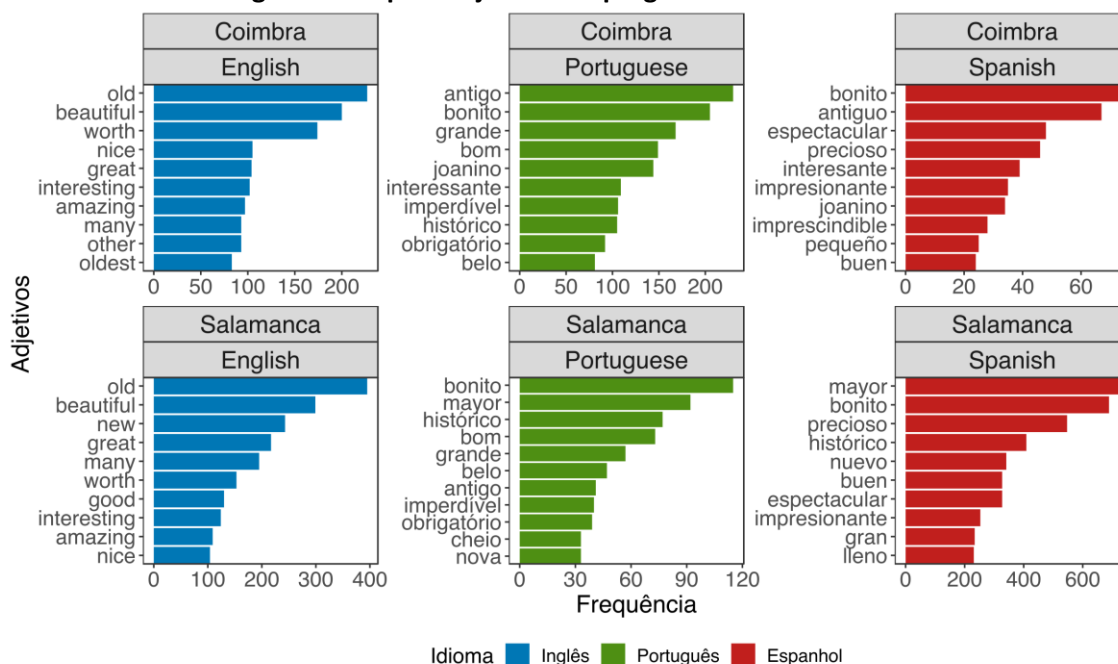
[...] achei um pouco caro, 12 euros (julho 2018) mas com ele poderá visitar as seguintes atracções: biblioteca Joanina (belíssima!), a capela de São Miguel (recentemente restaurada!), a prisão académica, a sala dos capelos, a via latina, a sala do exame privado e ainda o museu da ciência (no colégio

Tabela 3 – Frequência e distribuição de palavras e termos relacionados com património (por idioma)

Palavra/termo	Cidade	Português	Espanhol	Inglês
UNESCO	Coimbra	22 (1,3%)	1 (0,2%)	17 (1,9%)
	Salamanca	3 (0,5%)	5 (0,1%)	2 (0,2%)
património mundial, <i>patrimonio de la humanidad, world heritage</i>	Coimbra	10 (0,6%)	3 (0,6%)	10 (1,1%)
	Salamanca	2 (0,3%)	29 (0,7%)	1 (0,1%)
herança, <i>herencia, ancestry</i>	Coimbra	2 (0,1%)	0 (0,0%)	0 (0,0%)
	Salamanca	0 (0,0%)	0 (0,0%)	0 (0,0%)
edificado	Coimbra	2 (0,1%)	0 (0,0%)	0 (0,0%)
	Salamanca	0 (0,0%)	1 (0,0%)	0 (0,0%)
cultura, <i>culture</i>	Coimbra	77 (4,7%)	5 (1,1%)	10 (1,1%)
	Salamanca	15 (2,6%)	53 (1,4%)	8 (0,7%)
arquitectura, <i>arquitectura, architecture</i>	Coimbra	100 (6,1%)	7 (1,5%)	53 (6,0%)
	Salamanca	78 (13,3%)	160 (4,1%)	143 (12,3%)
tradição, <i>tradición, tradition</i>	Coimbra	39 (2,4%)	2 (0,4%)	8 (0,9%)
	Salamanca	3 (0,5%)	10 (0,3%)	5 (0,4%)
legado, <i>legacy</i>	Coimbra	1 (0,1%)	0 (0,0%)	1 (0,1%)
	Salamanca	0 (0,0%)	2 (0,1%)	0 (0,0%)
história, <i>historia, history</i>	Coimbra	367 (22,3%)	37 (8,0%)	160 (18,1%)
	Salamanca	78 (13,3%)	404 (10,4%)	101 (8,7%)

Fonte: Elaboração própria.

Figura 7 – Top de adjetivos empregues nos comentários



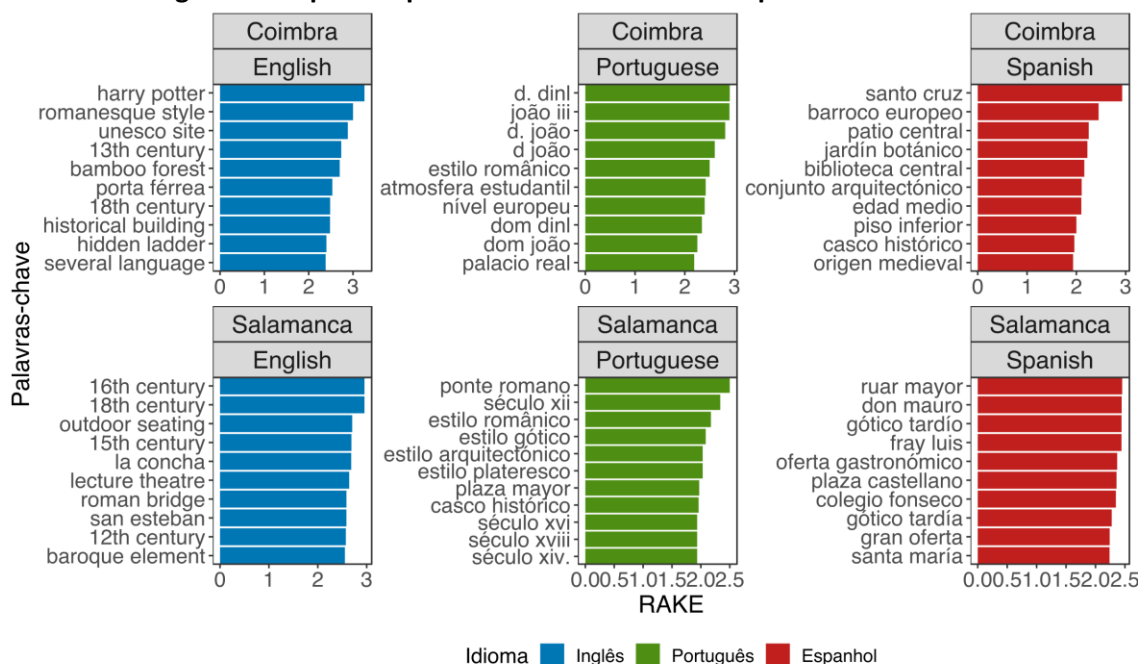
Fonte: Elaboração própria.

A extração de palavras-chave através do método RAKE também permitiu encontrar alguns resultados interessantes. Como pode ser visualizado na Figura 8, as palavras-chave diferem por idioma e por cidade. Também é possível constatar que são muitas as palavras-chave relacionadas com história e património, que vão desde várias referências a estilos arquitetónicos, nomes de reis, locais, até referências a séculos específicos.

No entanto, a maior curiosidade é o facto de a palavra-chave com maior índice RAKE, para Coimbra, em Inglês, ser “Harry Potter”. Com efeito, a posterior análise textual de alguns dos comentários parece mostrar que os utilizadores-visitantes observam a Biblioteca Joanina, as ruas de Coimbra ou o espaço da Universidade da perspetiva do universo “Harry Potter”, conforme os extratos 8 e 9 ilustram:³

³ Personagem principal ficcionada de série de sete romances infantojuvenis escritos em Língua Inglesa pela britânica J.K Rowling (1998-2007). Os romances foram também adaptados para cinema.

Figura 8 – Top 10 de palavras-chave identificadas pelo método RAKE



Fonte: Elaboração própria.

Extrato 8

Looks like part of the set of a Harry Potter movie with walls of books and small areas that you'd expect to see hidden doors and secret passageways (Biblioteca Joanina, Coimbra, 25.04.2017) [sublinhado nosso].

Extrato 9

Enjoyed our walk through the narrow streets of historical Coimbra to visit the university. Especially seeing the students in gowns (obligatory for students from the second year of studies) parading through the quadrangles and courtyards. An inspiration for J.K. Rowling and the Harry Potter novels. (residente Jerusalém, Israel; Jardim Botânico, Coimbra, 01.08.2017) [sublinhado nosso].

5 Discussão dos resultados

As análises efetuadas utilizando os métodos quantitativos e qualitativos permitiram concluir que dos dez locais escolhidos em cada cidade, alguns foram alvo de mais comentários do que outros. Por outro lado, constatou-se que em Coimbra o número de comentários dos estrangeiros é superior aos comentários dos portugueses, enquanto que em Salamanca os comentários de utilizadores com residência em Espanha são superiores aos outros utilizadores. Estes resultados parecem indicar que Coimbra proporcionalmente recebe mais estrangeiros do que Salamanca. Poder-se-á assim colocar duas hipóteses explicativas: ou Coimbra está a promover-se adequadamente no mercado estrangeiro, ou não se está a promover bem no mercado

nacional. As hipóteses também se podem colocar para Salamanca inversamente: esta cidade está a promover-se muito bem junto do público nacional ou está a promover-se mal junto do público internacional. Verificou-se ainda que, relativamente aos estrangeiros, a maioria são oriundos de países europeus ou da América Latina.

De salientar que foi detetada uma centena de utilizadores “comuns”, ou seja, que publicaram comentários sobre Coimbra e sobre Salamanca. Estes resultados poderiam ser utilizados pelas Organizações de Gestão de Destinos e pelas entidades responsáveis pela promoção do património UNESCO para fazerem *cross-marketing* e, desta forma, potenciarem o reconhecimento das duas cidades como destinos turísticos de excelência e reconhecidas como património UNESCO.

Relativamente à pergunta de investigação P2 se seria a classificação património UNESCO ou património da humanidade um fator de atração, foi possível apurar, utilizando o método RAKE, que a palavra *UNESCO* ou palavras equivalentes, como, por exemplo, “património mundial”, ocorreram com pouquíssima frequência nos comentários, tendo, no entanto, aparecido com mais frequência as palavras “arquitetura” e “história”. Estes resultados parecem indicar que, de entre as quatro categorias de Património Mundial UNESCO baseadas nas motivações para a candidatura indicadas pela *Research Consulting Ltd. & Trends Business Research Ltd.* (2009) – celebração; designação SOS; marca para marketing/logo e catalisador para construção de um espaço – claramente a categoria de marca não está a ser trabalhada nem em Coimbra nem em Salamanca e, com base no *corpus* coligido e analisado, poder-se-á especular que não parece ser um fator de atração para as visitas. Em conclusão, a reputação UNESCO parece não estar a ser bem trabalhada.

Como já foi referido, outro aspeto interessante que a análise revelou foi a transposição do universo de ficção Harry Potter para o espaço antigo de Coimbra ou, dito de outro modo, a evocação do universo de Harry Potter, a partir da imersão no espaço de Coimbra, pelo menos para os visitantes que escreveram comentários em Inglês. Este fenómeno, comum em algumas obras de ficção e em alguns leitores, poderia ser aproveitado pelas entidades responsáveis pela promoção da cidade de Coimbra para a promoverem junto dos turistas.

Conforme indicámos na introdução, o recurso a métodos mistos em estudos sobre o turismo está em crescimento, sendo que 6% dos artigos de turismo publicados entre 1994-2005 em doze revistas científicas fundamentais adotam esta metodologia (MOLINA; FONT, 2016). No caso do presente estudo, os métodos, embora trabalhosos, revelaram-se complementares e permitiram conhecer de forma mais completa quem são os visitantes-utilizadores das duas cidades.

6 Conclusões

Não é possível compreender o turismo patrimonial moderno, sem considerar as pessoas que se apropriam – ou consomem – o património (LIGHT; PRENTICE, 1994). Nessa perspetiva, este estudo explorou a forma como as pessoas consomem o património de duas cidades classificadas pela UNESCO, recorrendo a um *corpus* coligido a partir do que as próprias pessoas escrevem sobre esse mesmo “consumo”. Com base numa abordagem mista, testou-se um conjunto de métodos e técnicas de investigação que permitem a exploração quantitativa e qualitativa de CVO na perspetiva de melhor conhecer e compreender o que os visitantes apreciam e não apreciam, desejam e querem encontrar na sua experiência. Mostrou-se de que forma as várias técnicas se podem complementar e a capacidade de estabelecer informação padronizada sobre as expectativas, perceções e apreciações dos visitantes, que são simultaneamente utilizadores do TripAdvisor. Adicionalmente, procuraram-se destacar as vantagens da aplicação dos vários métodos de análise num estudo comparativo e demonstrar a riqueza informacional dos CVO como instrumento para (re)posicionar o destino.

Embora o nosso estudo apenas contemple duas cidades, seria útil replicá-lo em outras cidades com classificação UNESCO em diferentes regiões geográficas. O alargamento do estudo permitiria aprofundar o conhecimento sobre o que os turistas/utilizadores desejam e valorizam encontrar, informação particularmente relevante para promover as cidades e para alavancar o desenvolvimento económico, o desenvolvimento de uma identidade de lugar e de promover mudanças no lugar.

De acordo com o estudo de Adie e Hall (2016), aplicado a três sítios com a classificação Património Mundial UNESCO, os turistas do Património Mundial da UNESCO são homens e mulheres em igual número, têm escolaridade muito elevada, estão empregados e viajam em grupos pequenos de duas a cinco pessoas. Estas características demográficas, embora não tendo sido testadas na nossa análise, permitem provavelmente explicar a construção textual clara e relativamente sofisticada dos CVO coligidos e as frequentes analogias e referências culturais a outros locais, aspetos que serão explorados no futuro. Dada a natureza exploratória do estudo, a abordagem de métodos múltiplos, a recolha de CVO para as duas cidades e a riqueza informacional dos dados nos três idiomas, julgamos ter provado a pertinência do nosso estudo e o potencial de investigação futura, nomeadamente de orientação mais qualitativa no que respeita à exploração das diferenças nos padrões discursivos e retóricos dos CVO nos três idiomas – Português, Espanhol e Inglês – campo que tem sido pouco explorado, pese embora a importância crescente das questões interculturais na área da comunicação mediatizada por computadores (TIAN, 2013) e ainda a análise mais completa das diferenças das apreciações entre as duas cidades.

Referências

- ADEDOYIN-OLOWE, M.; GABER, M. M.; STAHL, F. A survey of data mining techniques for social media analysis. ArXiv Preprint ArXiv:1312.4617, 2013. Disponível em: <http://arxiv.org/abs/1312.4617>. Acesso em: 1 maio 2019.
- ADIE, B.A.; HALL C. M. Who visits world heritage? A comparative analysis of three cultural sites, **Journal of Heritage Tourism**, 12(1), 67-80, 2016.
- ANTONIO, N.; ALMEIDA, A.; NUNES, L.; BATISTA, F.; RIBEIRO, R. Hotel online reviews: Creating a multi-source aggregated index, **International Journal of Contemporary Hospitality Management**, 30(12), 3574-3591, 2018a.
- ANTONIO, N., ALMEIDA, A. DE, NUNES, L., BATISTA, F.; RIBEIRO, R. Hotel online reviews: different languages, different opinions, **Information Technology & Tourism**, 18(1-4), 157-185, 2018b.
- BERZOSA-PAREJA, R. Salamanca: planificación y gestión turística de un destino patrimonial, **Polígonos: Revista de Geografía**, 30, 171-194, 2018.
- DE ASCANIIS, S.; CANTONI L. Online visit opinions about attractions of the religious heritage: An argumentative approach, **Church, Communication and Culture**, 2(2), 179-202, 2017.
- DIAS, J. A.; RIBEIRO, F. P.; CORREIA, A. Online reviews of short-term visits: exploring sense of place, **International Journal of Culture, Tourism and Hospitality Research**, 7(4), 364-374, 2013.
- DUAN, W.; YU, Y.; CAO, Q.; LEVY, S. Exploring the impact of social media on hotel service performance: A sentimental analysis approach, **Cornell Hospitality Quarterly**, 57(3), 282-296, 2016.
- DUNHAM, K.; MELNICK, J. **Malicious Bots: An Inside Look into the Cyber-criminal Underground of the Internet**, Boca Raton, FL, USA: CRC Press, 2009.
- DU CROS, H.; MCKERCHER, B. **Cultural Tourism: The Partnership between Tourism and Cultural Heritage Management** (2nd ed.), New York, NY: Routledge, 2015.
- EUROPEAN COMMISSION (Ed.). Study on online consumer reviews in the hotel sector. Final report, 2014. Disponível em: http://rpaltd.co.uk/uploads/report_files/hotel-reviews.pdf. Acesso em: 6 maio 2019.
- EUROPEANA. Moving targets: Engaging cultural tourists with collections and listings content online, 2015. Disponível em: http://pro.europeana.eu/files/Europeana_Professional/Publications/Moving%20Targets%20-%20Colour.pdf. Acesso em: 6 maio 2019.
- FELDMAN, R.; SANGER, J. **The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data**, New York, N.Y., USA: Cambridge University Press, 2007.
- FEURRIEGEL, S.; PROELLOCHS, N. Sentiment Analysis: Dictionary-based sentiment analysis (Version R Package version 1.3-2), 2018. Disponível em: <https://CRAN.R-project.org/package=SentimentAnalysis>. Acesso em: 20 abr. 2019.
- FORTUNA, C.; GOMES, C. S. Turismo, cidade e universidade: o caso de Coimbra, in Cravidão, F. & Santos, N. (eds.), **Turismo e Cultura: Destinos e Competitividade**, Coimbra: Imprensa da Universidade de Coimbra, 2014.
- GE, J.; VAZQUEZ, M.; GRETZEL, U. Sentiment analysis: a review, in M. Sigala & U. Gretzel (eds.). **Advances in Social Media for Travel, Tourism and Hospitality: New Perspectives, Practice and Cases** (pp. 243-261), New York, NY: Routledge, 2018.
- GOMES, C. S. Novas imagens para velhas cidades? Coimbra, Salamanca e o turismo nas cidades históricas, **Sociologia, Revista da Faculdade de Letras da Universidade do Porto**, 23, 37-49, 2012.

- HAUSMANN, A.; WEUSTER L. Possible marketing tools for heritage tourism: the potential of implementing information and communication technology, **Journal of Heritage Tourism**, 13(3), 273-284, 2018.
- IGLESIAS-SÁNCHEZ, P. P.; CORREIA, M. B.; JAMBRINO-MALDONADO, C. Challenges of open innovation in the tourism sector, **Tourism Planning & Development**, 16(1), 22-42, 2019.
- INE. Cifras oficiales de población de los municipios españoles, 2017a. Disponível em: <https://www.ine.es/jaxiT3/Datos.htm?t=2911>. Acesso em: 11 maio 2019.
- INE. Encuesta de Ocupación Hotelera. Número de viajeros por puntos turísticos, Salamanca (datos de 2010 a 2017), 2017b. Disponível em: <https://www.ine.es/jaxiT3/Tabla.htm?t=2078&L=0>. Acesso em: 11 maio 2019.
- INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO country codes, 2017. Disponível em: <https://www.iso.org/obp/ui/#search>. Acesso em: 13 fev. 2017.
- KAYAHAN, B.; VANBLARCOM, B. Cost benefit analysis of UNESCO world heritage site designation in Nova Scotia, **Review of Economic Analysis**, 4, 247-273, 2012.
- LARSSON, T.; REIBRING, A. Tracking Multi-Destination Travel in Europe from Long-Haul Source Markets – US, China, Japan and India, European Travel Commission Handbook, Market Intelligence Report, Brussels: ETC, 2018. Disponível em: http://www.etc-corporate.org/?page=report&report_id=122. Acesso em: 11 maio 2019.
- LIAO, S. H.; CHU, P. H.; HSIAO, P. Y. Data mining techniques and applications – A decade review from 2000 to 2011, **Expert Systems with Applications**, 39(12), 11303-11311, 2012.
- LIGHT, D.; PRENTICE, R. C. Who consumes the heritage product? Implications for European heritage tourism, in G. J. Ashworth & P. J. Larkham (eds.), *Building a New Heritage: Tourism, Culture and Identity in the New Europe*, p. 90-116, Routledge: London, 1994.
- LIU, B.; ZHANG, L. A survey of opinion mining and sentiment analysis. In C. C. Aggarwal & C. X. Zhai (eds.), **Mining Text Data**, p. 415-463, 2012.
- MANNING, C. D.; MANNING, H. S. **Foundations of Statistical Natural Language Processing** (2nd ed.), London, UK: The MIT Press, 1999.
- MASON, J. Mixing methods in a qualitative driven way, **Qualitative Research**, 6(1), 9-25, 2006.
- MOLINA, J.; FONT, X. Mixed methods in sustainable tourism research: an analysis of prevalence, designs and application in JOST (2005-2014), **Journal of Sustainable Tourism**, 24(4), 549-573, 2016.
- NADKARNI, P. M.; OHNO-MACHADO, L.; CHAPMAN, W. W. Natural language processing: an introduction, **Journal of the American Medical Informatics Association: JAMIA**, 18(5), 544-551, 2011.
- NDOU, V. New approaches for managing tourism complexity: Implications and insights, in *Digital Culture and e-Tourism: Technologies, Applications and Management Approaches* (p. 123-138), Hershey, NY: IGI Global, 2011.
- Notícias de Coimbra (2018, 4 janeiro). 500 mil pessoas visitaram a Universidade de Coimbra em 2017. Disponível em: <https://www.noticiasdecoimbra.pt/500-mil-pessoas-visitaram-universidade-coimbra-2017/>. Acesso em: 7 maio 2019.
- OLSEN, W. Triangulation in Social Research: Qualitative and Quantitative Methods can really be mixed. In: M. HOLBORN & HARALAMBOS. **Developments in Sociology**. Ormskirk: Causeway Press, 2004.
- O'REILLY, T. What is Web 2.0: Design patterns and business models for the next generation of software. **International Journal of Digital Economics**, 65, 17-37, 2007.

- PEIXOTO, P. As cidades e os processos de patrimonialização: a corrida ao estatuto de património mundial e a identidade simbólica das cidades. *In*: PINHEIRO, M. BAPTISTA, L. V. & VAZ, M.^a J. **Cidade e Metrópole: Centralidades e Marginalidades**, p. 171-179, Oeiras: Celta Editora, 2001.
- PORDATA. População residente, estimativas a 31/12, 2018a. Disponível em: <https://www.pordata.pt/Municipios/Popula%C3%A7%C3%A3o+residente++estimativas+a+31+de+Dezembro-120>. Acesso em: 6 maio 2019.
- PORDATA. Hóspedes nos alojamentos turísticos: total e por tipo de alojamento, 2018b. Disponível em: <https://www.pordata.pt/Municipios/H%C3%B3spedes+nos+alojamentos+tur%C3%ADsticos+total+e+por+tipo+de+alojamento-750>. Acesso em: 8 maio 2019.
- PRUD'HOMME, R. The socio-economic impacts of inclusion of a site on the World Heritage list: Three studies, 2008. Disponível em: <http://www.prudhomme.com/resources/2008+Impact+Liste+Patrimoine.pdf>. Acesso em: 2 maio 2019.
- R CORE TEAM. R: A language and environment for statistical computing, 2016. Disponível em: <https://www.R-project.org/>. Acesso em: 10 abr. 2019.
- RAHMANI, K.; GNOTH, J.; MATHER, D. Tourists' Participation on Web 2.0: A Corpus Linguistic Analysis of Experiences, **Journal of Travel Research**, 57(8), 1108-1120, 2018.
- RESEARCH CONSULTING LTD. & TRENDS BUSINESS RESEARCH LTD. World Heritage Status: Is There Opportunity for Economic Gain?, 2009. Disponível em: <http://www.lakesWH.co.uk/documents/WHTheEconomicGainFinalReport.pdf>. Acesso em: 9 maio 2019.
- ROQUE, V.; FERNANDES, G.; RAPOSO, R. Identificação dos Media Sociais utilizados pelas organizações de gestão de destinos: o caso de estudo do destino turístico Serra da Estrela, **Revista Turismo & Desenvolvimento**, 17/18, 225-234, 2012.
- ROSE, S.; ENGEL, D., CRAMER, N.; COWLEY, W. Automatic keyword extraction from individual documents. *In*: M. W. Berry & J. Kogan (eds.), **Text Mining**, p. 1-20, 2010.
- SÁNCHEZ-JIMÉNEZ, M. A., CORREIA, M. B.; MATOS, N. Análisis de las redes sociales como instrumento de comunicación en los destinos turísticos del Algarve, **Revista GEINTEC-Gestão, Inovação e Tecnologias**, 8(3), 4476-4494, 2018.
- SARALEGI, X.; SAN VICENTE, I. Elhuyar at TASS 2013. *In*: A. D. Esteban, I. A. Loinaz & J. V. Román (eds.), Proceedings of XXIX Congreso de la Sociedad Española de Procesamiento de Lenguaje Natural, p. 143–150, Madrid, Spain: El Congreso Español de Informática, 2013.
- SILVA, M. J.; CARVALHO, P.; SARMENTO, L. Building a sentiment lexicon for social judgement mining. *In*: H. Caseli, A. Villavicencio, A. Teixeira & F. Perdigão (eds.), Computational processing of the Portuguese Language, p. 218–228, 2012.
- SOLDADO, C. Há um “efeito UNESCO” em Coimbra?, Público (2015, 22 junho). Disponível em: <https://www.publico.pt/2015/06/22/p3/noticia/ha-um-efeito-UNESCO-em-coimbra-1823585>. Acesso em: 2 dez. 2019.
- TIAN, Y. Engagement in online hotel reviews: A comparative study, **Discourse, Context and Media**, 2, 184-191, 2013.
- TRIPADVISOR. 35 Amazing TripAdvisor Statistics and Facts (2019) | By the Numbers. DMR – Business Statistics, 2018a. Disponível em: <https://expandedramblings.com/index.php/tripadvisor-statistics/>. Acesso em: 17 abr. 2019.
- TRIPADVISOR. All about your TripAdvisor bubble rating, 2018b. Disponível em: <https://www.tripadvisor.com/TripAdvisorInsights/w810>. Acesso em: 20 abr. 2019.

- TRIPADVISOR. About TripAdvisor, 2019. Disponível em: http://www.tripadvisor.com/pages/about_us.html. Acesso em: 17 abr. 2019.
- TSCHEU, F.; BUHALIS, D. Augmented reality at cultural heritage sites. In: A. Inversini & R. Schegg (eds.), *Information and Communication Technologies in Tourism* (p. 607-619), Cham, Switzerland: Springer International, 2016.
- UNESCO. World Heritage Information Kit: Benefits of Ratification, 2000. Disponível em: <http://www.UNESCO.org/whc/nwhc/pages/doc/main.htm>. Acesso em: 5 abr. 2019.
- UNESCO. UNESCO and online travel giant TripAdvisor launch World Heritage partnership, 2009. Disponível em: <https://whc.UNESCO.org/en/news/561>. Acesso em: 3 maio 2019.
- UNESCO. Report on the World Heritage Thematic Programmes, WHC/16/40.COM/5D, 2016. Disponível em: <http://whc.UNESCO.org/archive/2016/whc16-40com-5D-en.pdf>. Acesso em: 8 maio 2019.
- VANBLARCOM; B.; KAYAHAN, C. Assessing the economic impact of a UNESCO World Heritage designation, *Journal of Heritage Tourism*, 6(2), 143-164, 2011.
- WALKER, C.; BAXTER, J. Method sequence and dominance in mixed methods research: a case study of the social acceptance of wind energy literature, *International Journal of Qualitative Methods*, 18, 1-14, 2019.
- WIJFFELS, J. “udpipe”: Tokenization, Parts of Speech Tagging, Lemmatization and Dependency Parsing with the “UDPipe” “NLP” Toolkit (Version R package version 0.8), 2018. Disponível em: <https://CRAN.R-project.org/package=udpipe>. Acesso em: 6 abr. 2019.