

Imatge de Catalunya percebuda per turistes angloparlants i castellanoparlants*

Estela Marine-Roig

Universitat de Lleida. Facultat de Dret, Economia i Turisme
estela.marine@aegern.udl.cat

Esteve Mariné Gallisà

Universitat Oberta de Catalunya. Estudis d'Informàtica
emarine@uoc.edu



Rebut: octubre de 2016
Acceptat: febrer de 2017

Resum

Des de fa una dècada, s'ha multiplicat notablement el volum dels continguts en línia generats per usuaris i consumidors (UGC), juntament amb l'expansió dels mitjans de comunicació socials. En el camp de l'hoteleria i del turisme, s'ha produït un increment exponencial de les opinions en línia sobre viatges (OTR). En aquest article es proposa un mètode per analitzar les produccions paratextuals lligades a les OTR i extreure'n la imatge percebuda de la destinació segons la procedència dels turistes. El mètode s'aplica a una mostra aleatòria de 330.000 OTR allotjades a TripAdvisor, sobre Coses per Fer a Catalunya, escrites en castellà o en anglès entre 2011 i 2015. D'acord amb les opinions analitzades, pel que fa a la imatge percebuda (i transmesa) pels turistes de parla anglesa i, en especial, pels nord-americans, es pot concloure en síntesi que Catalunya és Barcelona i que Barcelona és Gaudí.

Paraules clau: imatge de la destinació; contingut generat per l'usuari; dades massives; opinions en línia sobre viatges; marques turístiques territorials

* Aquest estudi forma part d'un projecte més ampli titulat «Efectes de la transformació de les pautes de mobilitat global en l'evolució de les destinacions turístiques (MOVETUR)» finançat pel Ministeri d'Economia i Competitivitat espanyol (CSO2014-51785-R).

Resumen. *Imagen de Cataluña percibida por turistas anglohablantes y castellanohablantes*

En lo que va de década, se ha multiplicado notablemente el volumen de los contenidos en línea generados por usuarios y consumidores (UGC), junto con la expansión de los medios de comunicación sociales. En el campo de la hostelería y del turismo, se ha producido un incremento exponencial de las opiniones en línea sobre viajes (OTR). En este artículo se propone un método para analizar las producciones paratextuales ligadas a las OTR y extraer de ellas la imagen percibida del destino según la procedencia de los turistas. El método se aplica a una muestra aleatoria de 330.000 OTR alojadas en TripAdvisor, sobre Cosas para Hacer en Cataluña, escritas en castellano o inglés entre 2011 y 2015. Según las opiniones analizadas, por lo que respecta a la imagen percibida (y transmitida) por los turistas de habla inglesa y, en especial, por los norteamericanos, se puede concluir en síntesis que Cataluña es Barcelona y Barcelona es Gaudí.

Palabras clave: imagen del destino; contenido generado por el usuario; datos masivos; opiniones en línea sobre viajes; marcas turísticas territoriales

Résumé. *Image de la Catalogne perçue par les touristes anglophones et hispanophones*

Lors de la dernière décennie, le volume de contenus générés par les consommateurs ou usagers (UGC) s'est multiplié avec l'expansion des moyens de communication sociaux. Dans le domaine de l'hôtellerie et du tourisme, il y a eu une croissance exponentielle des opinions en ligne sur les voyages (OTRs). Cet article propose une méthode pour analyser les productions paratextuelles liées aux OTRs et en déduire l'image perçue de la destination selon la provenance des touristes. La méthode est appliquée à un échantillon de 330.000 OTRs recueillies sur TripAdvisor, sur les Choses à Faire en Catalogne, écrites en espagnol ou en anglais entre 2011 et 2015. Selon les opinions analysées et plus particulièrement selon celles des Nord-américains, on peut conclure en synthèse que la Catalogne se résume à Barcelone, et que Barcelone se résume à Gaudí.

Mots-clés: image de destination; contenu généré par les usagers; données massives; opinions en ligne sur les voyages; Catalogne; marques touristiques territoriales

Abstract. *Image of Catalonia perceived by English-speaking and Spanish-speaking tourists*

In the last decade, the volume of consumer and user-generated content (UGC) has increased notably alongside the expansion of social media. In the hospitality and tourism field, online travel reviews (OTRs) have grown exponentially. This article proposes a method to analyse the paratextual productions attached to OTRs and to extract from them the perceived image of the destination according to the tourists' origin. This method is applied to a random sample of 330,000 OTRs hosted in TripAdvisor about Things to Do in Catalonia written in Spanish or English from 2011 to 2015. According to the analysis of these opinions concerning the perceived (and transmitted) image among English-speaking and, especially, North American tourists, the main conclusion is that Catalonia is Barcelona and Barcelona is Gaudí.

Keywords: destination image; user-generated content; big data; online travel reviews; territorial tourism brands

circuits, excursions...) o serveis (hotels, restaurants, transports...) determinats d'una destinació. Per tenir una idea del volum i del creixement de les OTR disponibles en línia es pot citar Booking que, al gener de 2015, afirmava tenir 43 milions d'OTR d'hostes reals verificats, 75 milions al gener de 2016 i més de 106 milions a mitjans d'octubre del mateix any. En aquestes dates, TripAdvisor deia tenir més de 200 milions, 350 milions i 385 milions d'OTR, respectivament. Per analitzar aquestes quantitats d'informació, molts investigadors (Liu, Teichert, Rossi, Li, i Hu, 2017; Marine-Roig, 2017; Shafiee i Ghatari, 2016; Ylijoki i Porras, 2016; Yuan i Ho, 2015) coincideixen en la necessitat d'emprar tecnologies de processament de dades massives (*big data*).

El conjunt dels relats, opinions i valoracions continguts a les OTR permet deduir la imatge de la destinació (TDI: *tourism destination image*) percebuda pels seus visitants. La imatge percebuda, tal com es presenta i difon a través de l'allotjament web, esdevé imatge projectada i contribueix a tancar el cercle de construcció de la TDI global (Marine-Roig, 2015). Ara bé, un usuari normal es veu desbordat per la ingent quantitat d'informació disponible. Per exemple, una persona que està planificant unes vacances i vol consultar l'opinió d'anteriors visitants de la basílica de la Sagrada Família de Barcelona a TripAdvisor, el 15-10-2016 (TripAdvisor, 2016a) troba un enllaç amb un missatge associat —«*Read all 92671 reviews*»— i un altre al peu d'una fotografia de la basílica —«*See all visitor photos (49324)*»— (Llegeix totes les 92.671 ressenyes i mira totes les 49.324 fotos); i les xifres van creixent contínuament: per exemple, el 31-1-2016 hi havia 65.413 ressenyes i 27.639 fotos sobre la basílica (Marine-Roig, 2017).

En aquest context d'impossibilitat de llegir en detall els milers de ressenyes entren en joc les produccions paratextuals que permeten situar les OTR en l'espai i en el temps, així com saber quines són les atraccions més visitades i les millor valorades. El terme *paratext* va ser utilitzat per Gérard Genette per definir un conjunt de produccions (nom de l'autor o autora, títol, prefaci, il·lustracions) que acompanyen el text d'una obra literària (Genette, 1997). En un article del 2017, Marine-Roig proposa un marc per aplicar la teoria de Genette al cas de les OTR, basant-se en les interrelacions existents entre el text de l'OTR i les produccions paratextuals que l'envolten, tant les UGC com les generades per l'administrador de la web d'allotjament de les OTR (WGC: *webhost- / webmaster-generated content*). Seguint la nomenclatura de Genette, el *paratext* d'una OTR es divideix en *peritext* (títol, valoració, idioma, tema o tipus, data i localització geogràfica, seguit del perfil de l'autor, nombre de ressenyes publicades i destinacions visitades...) i *epitext* (ressenyes d'altres autors relacionades, publicitat contextual...).

En conseqüència, l'objectiu d'aquest article és proposar una metodologia per deduir la imatge percebuda pels visitants d'una destinació, situant-la en l'espai (marques turístiques territorials) i en el temps, a partir de les produccions paratextuals d'una quantitat significativa d'OTR relacionades. El mètode s'aplica a una mostra aleatòria de 210.000 OTR de TripAdvisor, sobre coses per fer (*things to do*) a Catalunya, escrites en castellà (60.000) i en anglès (150.000), entre els anys 2011 i 2015. Per explorar si la imatge percebuda varia

segons el país de procedència, es fa un zoom sobre els països més representatius de cada idioma, extraient-ne unes altres mostres aleatòries d'OTR escrites per turistes espanyols, britànics i nord-americans (40.000 de cada país).

2. Marc teòric

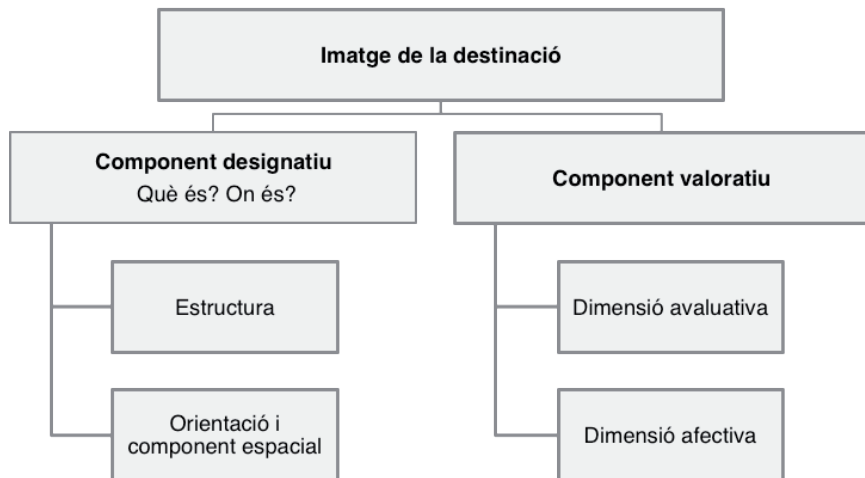
La TDI és i ha estat objecte d'estudi constant durant més de 50 anys (Chon, 1990; Gallarza et al., 2002; Li et al., 2015; Tasci et al., 2007), perquè gairebé tots els autors consideren que la imatge projectada d'una destinació és un aspecte crucial en la presa de decisions per part dels turistes que planifiquen un viatge o unes vacances. La TDI es pot definir com la suma de creences, idees i impressions que una persona té sobre una destinació (Crompton, 1979: 18) o bé com una construcció mental desenvolupada per un visitant potencial sobre la base d'unes poques impressions seleccionades enmig del torrent de totes les impressions (Fakeye i Crompton, 1991: 10). Com diuen aquests darrers autors, les imatges tenen una importància cabdal perquè transposen la representació d'una àrea dins de la ment d'un turista potencial i li donen una prepercepció de la destinació.

2.1. Components de la TDI

En el seu llibre sobre imatges de l'entorn urbà, Pocock i Hudson (1978) van proposar dos components (figura 2) per estudiar la imatge: designatiu i valoratiu. Les qualitats designatives de la imatge consisteixen en la informació bàsica sobre què és i on és, que concerneix la descripció i classificació, en un mapa mental amb propietats bàsiques com distància, orientació, localització i diversitat espacial. L'aspecte valoratiu està relacionat amb el significat lligat a, o evocat per, la forma física. Seguint aquesta classificació, Richards i Wilson (2004) relacionen l'aspecte designatiu o informatiu amb la categorització dels elements cognitius de l'entorn, i el valoratiu amb les percepcions, valors i significats, o el que se sent sobre un indret. L'aspecte valoratiu es divideix en avaluador, que concerneix la expressió d'una opinió, i afectiu, l'especificació d'una preferència. En altres paraules, l'aspecte designatiu està relacionat amb l'estructura de l'entorn i l'afectiu, amb el seu sentit o significat. En aquesta línia, Lynch (1960) afirma que cada persona té una imatge visual de la ciutat i la imatge dona a la ciutat identitat, estructura i significat.

La dicotomia d'imatge cognitiva-afectiva, procedent del camp de la geografia, l'han seguida molts autors del camp del turisme; en canvi, s'ha posat molt poca atenció en la dimensió espacial de la TDI. En el seu article, Echtner i Ritchie (1991) destaquen que la TDI consisteix en dos tipus de components principals: els basats en els atributs i els holístics com a conjunt d'imatges percebudes. Cadascun conté característiques funcionals o més tangibles i psicològiques o més abstractes. Més endavant, Gartner (1993: 193) i Basaran (2016) ja concreten que les imatges de la destinació estan formades per tres components clarament diferenciats, però jeràrquicament interrelacionats: cognitiu, afectiu

Figura 2. Components de la TDI derivats del model de Pocock i Hudson (1978)



Font: elaboració pròpia.

i conatiu. Les avaluacions perceptives/cognitives es refereixen a les creences o coneixements sobre els atributs d'una destinació, mentre que l'avaluació afectiva es refereix a sentiments o inclinació cap a ella (Baloglu i McCleary, 1999: 870). El component conatiu/volitiu afecta al comportament i a la presa de decisions en triar una destinació. La seva relació amb els altres dos components és directa, depèn de les imatges desenvolupades en l'etapa cognitiva i avaluades durant l'etapa afectiva (Gartner, 1993: 196).

La imatge cognitiva seguida de l'afectiva conformen la imatge percebuda global (Baloglu i McCleary, 1999; Beerli i Martín, 2004), influenciada per les fonts d'informació, primàries (experiència prèvia i intensitat de la visita) i secundàries (orgàniques, induïdes i autònomes), i pels factors personals (motivacions, experiència de les vacances i característiques socio-demogràfiques). Desenvolupant les teories anteriors, Marine-Roig (2015) relaciona la imatge percebuda global esmentada amb la identitat, l'autenticitat i la imatge projectada per tancar el cercle de construcció de la TDI des d'una perspectiva holística.

2.2. Fonts d'informació de la TDI

Gunn (1972) va fer una primera classificació dels agents que formen la TDI distingint els orgànics dels induïts. La diferència clau entre ambdós és el control que tenen les organitzacions de la destinació sobre la informació que es presenta. És a dir, les imatges induïdes emanen de l'àrea de la destinació i són un producte de màrqueting dels promotors de la destinació; i les imatges orgàniques provenen de fonts que no estan directament relacionades amb l'àrea de la destinació. Gartner (1993) subclassifica les fonts induïdes entre implícites

i explícites, i les orgàniques entre sol·licitades, no sol·licitades i orgàniques en sentit estricte. També introdueix els agents autònoms de formació de la imatge consistents en reportatges, documentals, pel·lícules i articles de premsa produïts independentment de la destinació. La font orgànica en sentit estricte és la informació que té una persona que ha viatjat anteriorment a la zona i es considera que és la que té la credibilitat més alta per a ella perquè està basada en l'experiència personal. Dins de les orgàniques sol·licitades està la informació desinteressada provinent d'amics i parents, que es coneix com a publicitat boca-orella (WoM: *word of mouth*) en el camp del màrqueting quan una nova va de boca en boca.

La classificació de Gartner (1993) continua sent vàlida al cap de vora un quart de segle. Només cal afegir-hi, com a font orgànica no sol·licitada, l'UGC difós a través de l'anomenat boca-orella electrònic (eWoM: *electronic word-of-mouth*). Les plataformes web que ofereixen UGC són àmpliament consultades pels turistes que cerquen pel seu compte informació sobre una destinació (Llodrà-Riera et al., 2015). L'UGC és considerada una font de confiança pels usuaris perquè dóna informació de primera mà, no sol·licitada i no esbiaixada per tercers, així com un filó inesgotable per als investigadors pel seu enorme volum i lliure accés en línia. A més de créixer espectacularment el seu volum, també en creix l'ús i la credibilitat amb el pas dels anys: Cox et al. (2009) van trobar que l'UGC era una font addicional menys creïble que les fonts oficials de la destinació; Ye et al. (2011) deien que l'UGC era una desconeguda pels investigadors i professionals de la indústria del turisme, però van demostrar que un increment de les puntuacions a les ressenyes dels viatgers suposava un creixement de les reserves en línia; Schuckert et al. (2015) ja parlen de la creixent popularitat de les OTR i del seu impacte sobre el comportament dels compradors així com sobre el costat de l'oferta; i Baka (2016) també demostra que les variacions en les puntuacions dels OTR de TripAdvisor, a través de l'eWoM, tenen un impacte en les reserves i proposa un model conceptual per gestionar la reputació en línia.

Les anteriors consideracions teòriques tenen la seva confirmació a les enquestes. Per exemple, a primers de gener de 2016, Statista (2016) va enquestar 2.010 nord-americans majors d'edat que havien viatjat per oci o per raons personals durant el 2015 i els va preguntar «Quins serveis o tecnologies d'Internet ha fet servir com a ajuda per planificar el seu viatge?». Va sortir en primer lloc l'UGC (eWoM), amb un 59,4%, i els mitjans de comunicació socials, amb un 49,8%; mentre que les opinions d'amics, companys i parents (WoM) va treure un 46,2%; i les fonts induïdes, webs d'organitzacions de màrqueting o de gestió de la destinació, un 36,4%, i les agències de viatge, un 12,3%. A les mateixes dates, Eurobarometer (2016) va fer una enquesta a 30.105 europeus de diferents grups socials i demogràfics, amb la pregunta «Quines de les següents fonts d'informació pensa que són més importants quan pren una decisió sobre els seus plans de viatge?». Va quedar en primer lloc el WoM, amb un 51%, seguit de les webs que recullen i presenten comentaris, opinions i puntuacions dels viatgers (eWoM), amb un 34%, pel davant de l'experiència

personal (33%), de les webs de proveïdors de serveis i de la destinació (17%) i de les agències de viatges i oficines de turisme (13%). S'ha de destacar que és el primer any que Eurobarometer (2016) introdueix les OTR a la macroenquesta i han quedat com la segona font d'informació amb el doble de puntuació que les webs de serveis i gestió de la destinació, i gairebé el triple que les agències de viatge i oficines de turisme.

3. Cas d'estudi: Catalunya turística

Catalunya és una destinació mediterrània, amb una història mil·lenària, una cultura i llengua pròpies, a més a més d'un important patrimoni cultural i paisatgístic. Amb més de 31 milions de visitants anuals, Catalunya és una destinació turística destacada, que ofereix nombrosos atractius a qualsevol tipus de visitant: cultura, descans, natura, turisme familiar, esportiu, de negocis... (ACT, 2016: 4).

3.1. Afluència turística

Catalunya va ser la tercera regió més turística de la Unió Europea per nombre de pernoctacions (72,7 milions), després de les Illes Canàries (94,3) i l'Île de France (77,7) durant el 2014 (Eurostat, 2016). En conjunt, els allotjaments turístics catalans van rebre 22,2 milions de viatgers (6,5 catalans; 3,1 de la resta de l'Estat, i 12,6 estrangers) i 75,3 milions de pernoctacions durant el 2015, on l'afluència va créixer en termes globals amb relació a un any abans (OdeïO, 2016).

L'any 2015, es va registrar un nou rècord d'arribada de turistes estrangers de primera destinació (17,4 milions) i despesa total (15.735,5 milions d'euros). La taula 1 recull l'evolució del nombre de turistes amb principal destinació Catalunya durant els darrers cinc anys. Altres països emissors tradicionals serien Irlanda, Països Nòrdics, Portugal i Suïssa, i altres emergents, Xina i Brasil.

Tot i el creixement del nombre total de turistes estrangers al 2015, a la taula 1 s'observa una disminució general dels mercats tradicionals, tret de França, Regne Unit i Estats Units, que registren al 2015 les millors xifres del període analitzat. Pel que fa al rànquing de despesa total l'any 2015 en milions d'euros: França (2.525,9), Regne Unit (1.408,9), Estats Units (1.179,0), Alemanya (1.017,2), Itàlia (805,4), Rússia (757,8), Països Baixos (589,2), Bèlgica (452,1) i Japó (n.d.). Relacionant aquestes dades amb les de la taula 1, els nord-americans i els russos són els que més gasten (més de 1.000 euros per cap), mentre que els francesos van en darrer lloc (menys de 500).

3.2. Marques turístiques territorials

La marca Catalunya està formada per nou marques turístiques territorials que agrupen i promocionen els diferents patronats de turisme (ACT, 2016: 16). En

Taula 1. Turistes estrangers (milers) amb destinació principal Catalunya

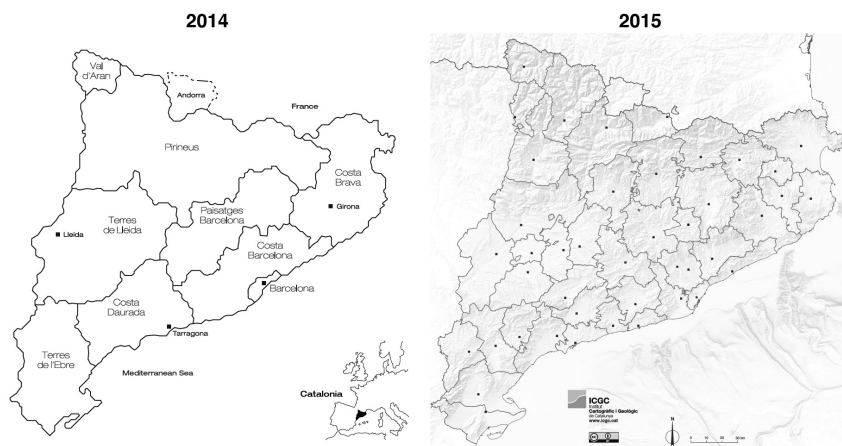
	Alemanya	Bèlgica	França	Itàlia	Japó	Països Baixos	Regne Unit	Rússia	Estats Units
2011	1.010,4	491,4	3.570,9	1.255,3	110,8	868,4	1.284,4	495,9	456,5
2012	1.162,4	515,1	3.815,6	1.168,5	161,6	691,2	1.606,1	741,4	577,4
2013	1.280,0	562,0	4.172,2	1.116,6	207,3	695,3	1.686,7	978,8	568,6
2014	1.429,4	592,3	4.602,1	1.331,2	273,4	814,3	1.781,8	833,4	512,0
2015	1.322,3	557,9	5.000,4	1.249,2	263,8	795,4	1.856,1	538,0	620,8

Font: Observatori del Treball i Model Productiu (Generalitat de Catalunya).

aquest estudi, es fan servir els següents acrònims de cinc lletres per distingir les nou marques actuals: Barcelona (Barna), Costa Barcelona (cBarc), Costa Brava (cBrav), Costa Daurada (cDaur), Paisatges Barcelona (pBarc), Pirineus (Pyren), Terres de Lleida (tLlei), Terres de l'Ebre (tEbre) i Val d'Aran (vAran). Pel que fa a anteriors denominacions de les marques, Catalunya Central (CatCe).

Les marques estan constituïdes per agrupacions de comarques contigües (figura 3) que presenten una oferta turística relativament homogènia perquè tenen uns atractius o característiques similars que les distingeixen unes de les altres. En aquest sentit, Marine-Roig i Anton Clavé (2016: 203) defineixen Catalunya com una destinació multiescalar on la imatge s'especialitza territorialment. En aquest context, l'especialització de la imatge percebuda es refereix al grau de com certs llocs estan comunicats i són percebuts a través de certs components d'activitats, imatge, atributs, sensacions o identitat que els caracteritzen i distingeixen uns dels altres com a destinacions turístiques. Des del punt de vista multiescalar, cada marca contribueix a la construcció de la imatge turística de Catalunya com a conjunt. La formació de la TDI a les

Figura 3. Marques turístiques territorials catalanes i mapa comarcal mut



Font: Agència Catalana de Turisme i Institut Cartogràfic i Geològic de Catalunya (ICGC).

destinacions multiescalars és altament complexa, ja que múltiples administracions i agents, a diferents escales geogràfiques, produeixen i gestionen les seves pròpies marques (Datzira-Masip i Poluzzi, 2014) i els visitants tendeixen a identificar-les a grans trets.

En conseqüència, una destinació de Catalunya pot estar gestionada i/o promocionada, directament o indirectament, per múltiples nivells administratius. Deixant a banda organismes internacionals com l'Organització Mundial del Turisme (UNWTO) i el European Travel Commission (ETC), hi ha un primer nivell estatal (TurEspanya), seguit del nivell autonòmic (Agència Catalana de Turisme), provincial (diputacions), comarcal (consells comarcals) i municipal (ajuntaments). Aquestes administracions poden exercir les seves competències a través d'ens de diferent estructura i composició, com patronats i consorcis (Arcarons, 2009; Datzira-Masip i Poluzzi, 2014), amb més o menys participació de les parts implicades.

La promoció oficial de les marques turístiques la fa l'Agència Catalana de Turisme (ACT) a les fires sectorials que participa. A més, l'ACT cada any publica, en sis idiomes, un dossier de premsa (ACT, 2016) amb el mapa de les marques turístiques i una descripció de cadascuna. L'ACT dedica exactament dues pàgines del dossier per descriure cada marca i incloure una fotografia representativa de la imatge de la zona al peu de cada pàgina. El dossier destaca els atributs o els atractius turístics que caracteritzen cada marca, per exemple: «Barcelona és la capital de Catalunya i una de les principals ciutats europees. Gràcies als seus atractius culturals i de lleure, en els darrers anys ha esdevingut un dels indrets més visitats de Catalunya» (p. 21); «La Costa Barcelona és un territori amb un clar compromís amb la qualitat... Aplega una alta concentració de serveis i activitats d'oci... D'entre els nombrosos atractius turístics destaquen els més de 100 quilòmetres de litoral» (p. 19); «La Costa Brava és una destinació mediterrània consolidada, els principals atractius de la qual són el clima, el paisatge, la costa, l'entorn natural i la cultura... S'identifica per l'alternança d'àmplies platges i petites cales envoltades de vegetació i penya-segats rocosos» (p. 17); «La Costa Daurada s'ha especialitzat bàsicament en el públic familiar, amb un ric i variat patrimoni cultural, i una destacada importància de l'enoturisme i la gastronomia. El litoral n'és un dels atractius principals» (p. 23); «Els Paisatges Barcelona són terres de l'interior de Catalunya, posseïdores d'un ric patrimoni natural... amb ciutats de gran tradició històrica i monumental o de mercat i monestirs emblemàtics» (p. 33); «Els Pirineus catalans ofereixen al visitant una àmplia varietat d'atractius: el paisatge amb el cim més alt de Catalunya... els pobles, que conserven un important patrimoni d'art romànic; la possibilitat de practicar nombrosos esports de muntanya i d'aventura, i les diferents propostes de turisme actiu» (p. 27); «Les Terres de Lleida privilegiades per poder gaudir tot l'any d'una gran varietat d'esports i activitats de turisme actiu, amb una cultura singular, una gastronomia autòctona i un turisme rural de qualitat» (p. 29); «Les Terres de l'Ebre agrupen les quatre comarques que travessa el riu Ebre en els prop de 130 quilòmetres de recorregut per territori català... Són terres d'una gran diversitat des del punt

Taula 2. Oferta de places d'allotjament distribuïdes per marques territorials (any 2015)

	Establiments hotelers				Càmpings		Turisme rural		Total
	Places	ADR ¹	RpAR ²		Places		Places	Places	
Barna	704	77.525	116,4	89,5	0	0	0	0	77.525
cBarc	443	58.810	65,2	41,2	45	32.904	158	1.291	93.005
cBrav	716	76.493	66,5	37,4	108	119.877	474	4.045	200.415
cDaur	284	58.369	75,5	46,8	53	63.051	209	1.723	123.143
pBrac	95	3.950	68,9	23,3	14	4.125	300	2.336	10.411
Pyren	405	15.194	67,0	18,0	101	36.828	926	6.949	58.971
tEbre	96	5.146	68,4	29,8	14	7.413	112	866	13.425
tLlei	90	4.775	49,3	16,2	4	1.575	172	1.368	7.718
vAran	108	6.429	79,7	32,7	9	3.069	16	159	9.657

1. ADR (*Average Daily Rate*): tarifa mitjana diària per habitació ocupada

2. RpAR (*Revenue per Available Room*): ingressos mitjans per habitació disponible

Font: Institut d'Estadística de Catalunya i Observatori d'Empresa i Ocupació.

de vista paisatgístic i gastronòmic» (p. 25); i «La Val d'Aran és un territori d'alta muntanya al Pirineu occidental... La seva situació i el seu relleu afavoreixen un clima, una fauna, una flora i un paisatge variats i singulars, i també una història única i singular, el màxim exponent de la qual és una llengua pròpia» (p. 31).

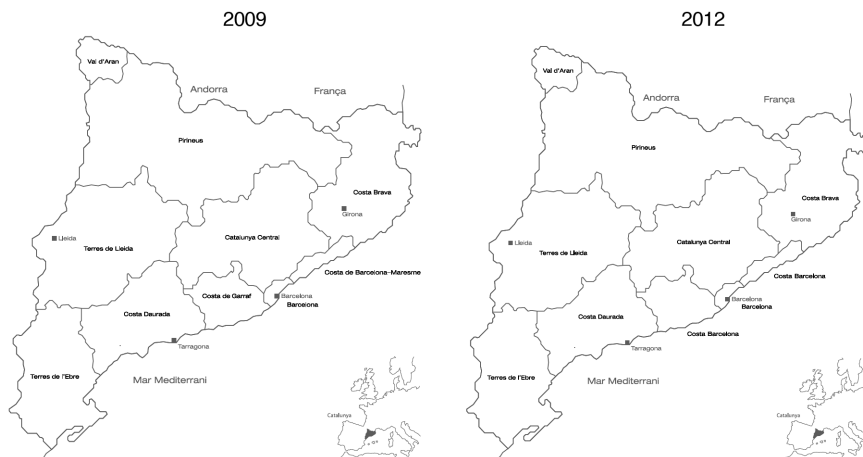
Totes les marques territorials disposen d'una gran varietat d'allotjaments turístics (establiments hotelers, càmpings, apartaments o habitatges d'ús turístic, i establiments de turisme rural), tret de Barcelona que, per les seves característiques de zona majoritàriament urbana, no disposa de càmpings ni d'establiments rurals. A la taula 2 es pot veure la distribució per marques de l'oferta d'unes 600.000 places d'allotjament. Com es pot observar, Barcelona és la que té més places en establiments hotelers i la que obté més ingressos per habitació disponible, amb força diferència sobre els altres territoris. En canvi, les tres marques costaneres la superen en el total de places d'allotjament per la disponibilitat de càmpings.

3.3. Estudis anteriors basats en contingut web generat pels viatgers

Les marques turístiques han evolucionat amb el temps. L'any 1985 només es considerava la perifèria (costa i Pirineus) i al 2000 es va adoptar el model global basat en l'agrupació de comarques. L'any 2009 es van segregar les marques Val d'Aran i Terres de l'Ebre de Pirineus i Costa Daurada respectivament i al 2012 es van agrupar Costa de Barcelona-Maresme i Costa del Garraf per constituir la marca Costa Barcelona (figura 4).

Amb la finalitat de poder analitzar posteriorment l'evolució temporal de la distribució territorial dels continguts web generats pels viatgers, s'han seleccionat dos treballs anteriors relacionats amb el cas d'estudi (taula 3). S'ha de fer constar que els esmentats continguts eren poc significatius abans del 2000

Figura 4. Anteriors marques turístiques territorials catalanes



Font: Agència Catalana de Turisme.

Taula 3. Distribució per marques (%) dels continguts generats pels viatgers

	Barna	cBarc	cBrav	cDaur	CatCe	Pyren	tEbre	tLlei	vAran
Estiu 2010	70,20	3,70	10,70	6,10	2,90	4,50	0,50	0,90	0,60
2001-2011	82,47	1,71	6,79	6,83	1,55	0,40	0,20	0,04	0,00

Font: González (2010) i Marine-Roig (2013).

i no estaven tan estructurats, motiu pel qual no es coneixen estudis d'anàlisi massiva d'èpoques anteriors.

- En un estudi encarregat per l'Observatori de Turisme de Catalunya, González (2010) va analitzar 11.311 entrades de blogs de viatge allotjades entre agost i octubre de 2010 a MiNube, TravBuddy, TravelBlog i TravelPod. Per compatibilitat, s'han sumat els resultats de les marques Costa del Garraf i Costa de Barcelona-Maresme, perquè al 2012 estaven refoses sota la marca Costa Barcelona sense cap modificació addicional.
- En la seva tesi doctoral, Marine-Roig (2013) va analitzar 25.003 entrades de blogs i ressenyes de viatge allotjades entre 2001 i 2011 a MyTripJournal, RealTravel, TravBuddy, TravelBlog, TravelPod, TripAdvisor i VirtualTourist.

4. Metodologia

El mètode proposat té el seu fonament en el marc teòric vist a la introducció sobre les interrelacions entre el text d'una OTR i les produccions paratextuals al

seu voltant. Pel que fa a la seva implementació, el mètode es basa en el paradigma del processament per lots de dades massives emmagatzemades. Un model conegut és el MapReduce, que permet processar en paral·lel i generar lots de dades grans en un sistema distribuït (clúster). Ara bé, en el present cas d'estudi s'han manipulat unes 500.000 pàgines web en un sol ordinador personal.

El mètode consta de tres fases: recol·lecció de la informació, mineria de dades i anàlisi de contingut. La fase de descàrrega ha suposat un temps considerable de dies per la sobrecàrrega o per les polítiques restrictives del servidor remot. S'ha de pensar que el servidor utilitzat en aquest cas (TripAdvisor.com) rep un mitjana de 350 milions de visitants únics cada mes; és a dir, un client que, per exemple, es connecta 15 dies seguits només el compten una vegada i, tot i amb això, el servidor té més d'11 milions de clients diferents cada dia que ha d'atendre en un temps raonable en línia, és a dir, quasi en temps real. Pel que fa a la fase de mineria de dades es triga unes hores i per a l'anàlisi, molt poc temps (algunes taules de freqüències es generen gairebé en temps real).

4.1. Estat de l'art

Actualment, hi ha una proliferació d'articles i d'edicions especials a les revistes científiques que tracten l'anàlisi de dades massives. En particular, algunes tenen com a objecte l'estudi de les ciutats i les destinacions intel·ligents mitjançant l'anàlisi massiu de continguts generats pels viatgers. Per exemple, Jabreel et al. (2016) fan una anàlisi semàntica de 60.000 piulades (*tweets*) relacionades amb gestors i visitants de 10 destinacions europees, inclosa Barcelona, per extreure'n els valors emocionals transmesos per ambdós. En un article sobre conceptualització de les dades massives, Ylijoki i Porras (2016) consideren les OTR com a noves fonts de dades d'accés lliure, però que requereixen un costós processament per aconseguir resultats. En canvi, Yuan i Ho (2015) posen les OTR com a exemple de dades massives que permeten extreure coneixements rellevants per a les organitzacions de màrqueting de les destinacions.

En matèria d'OTR de TripAdvisor similar al present cas d'estudi, Liu et al. (2017) presenten una recerca sobre 412.784 ressenyes per esbrinar la satisfacció dels hostes en més de 10.000 hotels xinesos que emmagatzemen en un sistema de gestió de bases de dades (DBMS: *database management system*). Aquests autors descriuen molts problemes per segmentar les OTR per idiomes, quan TripAdvisor ja les té classificades des del seu origen i cada OTR té escrit internament l'estàndard ISO_639-1 (codi de dues lletres per identificar els principals idiomes del món), el qual es pot extreure automàticament amb una simple cerca basada en una expressió (*regex*) de llenguatge regular.

4.2. Recol·lecció de la informació

Fent la cerca de *travel blogs* o *travel reviews* sobre Catalunya o Catalonia a Google, el lloc web amb més tràfic del món (Alexa.com, TopSites), aquest buscador

retorna 375.000 resultats a mitjans d'octubre de 2016; en canvi, fent la mateixa consulta per a Barcelona, en retorna 2 milions més. Això ja dona una idea del pes de la marca/destinació Barcelona en el món dels blogs de viatge i OTR de parla anglesa. El problema dels resultats que retorna Google sobre blogs o ressenyes de viatge és que moltes de les pàgines web tenen diferents estructures que no permeten mecanitzar-ne la recollida, la classificació i l'anàlisi. Aleshores, la solució està a seleccionar els llocs web que hostatgen els blogs de viatge i les OTR, perquè tenen la informació estructurada i, alguns, gairebé classificada. En base a treballs anteriors i a una actualització, s'han trobat els següents dominis relacionats amb el turisme i els viatges que tenen informació sobre el cas d'estudi: GetJealous.com (GJ), LonelyPlanet.com (LP), MiNube.com (MN), MyTripJournal.com (MT), StaTravel.com (ST), TravBuddy.com (TY), TravelBlog.org (TB), TravelJournals.net (TJ), TravellersPoint.com (TS), TravelPod.com (TP), TripAdvisor.com (TA), Venere.com (VN) i VirtualTourist.com (VT). Per seleccionar la font d'informació més idònia per al cas d'estudi, s'aplica una fórmula ponderada basada en una funció d'ordenació (B, recompte de Jean-Charles de Borda) i tres variables, popularitat (P, visites rebudes i tràfic en general), visibilitat (V, quantitat i qualitat d'enllaços entrants) i volum d'informació (S, nombre d'entrades relacionades amb el cas) que té el lloc web (Marine-Roig, 2014a):

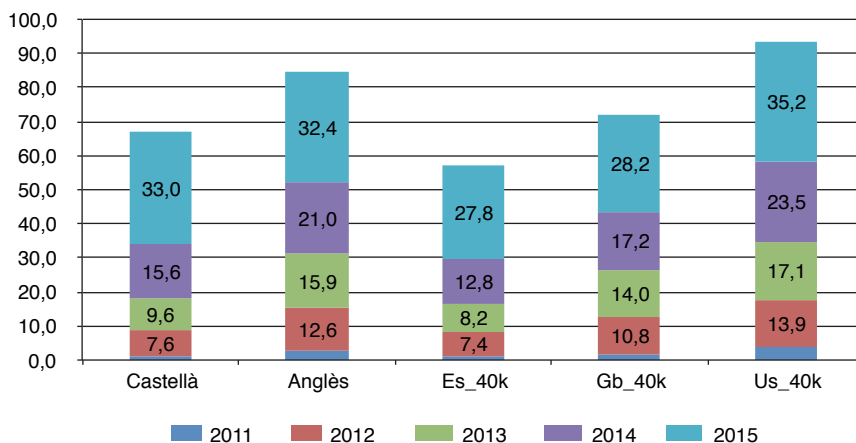
$$\text{TBRH} = 1 \times \text{B(V)} + 1 \times \text{B(P)} + 2 \times \text{B(S)}$$

La funció d'escrutini 'B' (de Borda, 1781) assigna 1 punt al darrer classificat, 2 punts al penúltim i així successivament fins a arribar al primer classificat, que té la màxima puntuació de la llista. En aquest article, s'utilitza la variant de donar zero punts a l'últim (n-1, n-2, ... 0). TBRH consisteix en una agregació ponderada de rànquings: en base a tres llistes completes (L) de webs ordenades per 'V', 'P' i 'S', la funció 'B' assigna una puntuació per a cada candidat (c), que consisteix en el nombre de candidats que van darrere 'c' a 'L'. Aquesta puntuació es multiplica pel pes de cada rànquing. Una vegada s'han sumat les puntuacions parcials, els candidats queden arranats per ordre descendent de puntuació total.

Un cop aplicada la fórmula queda en primer lloc TA, seguit molt de lluny per VT, TB i TP. En pocs anys, TA s'ha convertit en el líder del sector (Baka, 2016; Liu et al., 2017; Yoo et al., 2016) i ha absorbit 24 webs de viatges, com ST, TP i VT. A mitjans d'octubre de 2016, tenia 3 milions d'OTR i comentaris sobre Catalunya i, concretament, més de mig milió d'OTR sobre coses per fer a Barcelona capital (TripAdvisor, 2016b).

Seguidament es descarreguen les OTR sobre coses per fer a Catalunya (TripAdvisor, 2016b) (s'exclouen les OTR sobre hotels i restaurants) amb un programa copiador de webs com *HTTrack offline browser* (htrack.com) i s'elimina la informació redundant (TA inclou OTR relacionades que ja tenen la seva pròpia pàgina web). Per poder manipular els fitxers en un sistema operatiu Windows, es converteix el format UNIX / UTF-8 a PC / ANSI i les

Figura 5. Percentatge d'opinions sobre la marca Barcelona apilades per anys (2011-2015)



Font: 60.000 OTR de TA en castellà, 150.000 en anglès i 40.000 de cada país (ES, GB i US).

diverses codificacions dels caràcters especials (lletres amb accent, titlla...) es converteixen a ISO 8859-15.

A continuació, es fa una selecció de les OTR escrites en castellà o en anglès entre 2011 i 2015, però comprovant que els autors tinguin posat el país de procedència al seu perfil. Amb aquest primer filtre es descarten la meitat dels fitxers i es treu una mostra aleatòria de 150.000 OTR escrites en anglès i 60.000 més en castellà, en proporció aproximada al pes de cada idioma. També s'extreu una mostra aleatòria de 40.000 OTR de cadascun dels països més representatius d'ambdós idiomes (Espanya, Anglaterra i Estats Units d'Amèrica).

A la figura 5, es pot apreciar la distribució temporal en percentatge de les OTR sobre la marca Barcelona, que són la majoria de totes les mostres (66,8% en castellà; 84,4% en anglès; 57,2% de turistes procedents d'Espanya; 71,8%, d'Anglaterra, i 93,4%, d'Estats Units).

4.3. Minería de dades

Les OTR consisteixen en un text i uns elements paratextuals que l'envolten. Es presenten en pàgines web HTML (*HyperText Mark-up Language*), que tenen la seva estructura interna sobre la base de text pla (no codificat) delimitat per etiquetes. Les etiquetes més importants són les que indiquen el principi i el final de les seccions comunes a totes les pàgines HTML (W3Schools, 2017): Pàgina (<html>...</html>), capçalera (<head>...</head>), títol dins de la capçalera (<title>...</title>) i cos (<body>...</body>). Al seu torn, les OTR tenen la seva pròpia estructura dins de la capçalera i cos de la pàgina HTML, sense la qual l'administrador de la web no podria gestionar els milions de pàgines que té emmagatzemades el servidor o que genera sota demanda.

El text pla, en combinació amb les etiquetes, permet fer cerques simples com, per exemple, '<title>' i el buscador retorna les línies (una de cada pàgina) on hi ha el títol de l'OTR. L'acurada estructura de les OTR permet fer cerques amb més detall mitjançant expressions (*regex*) de llenguatge regular (seqüències de caràcters que formen un patró de cerca) i així poder extreure altres elements paratextuals (dates, destinacions, atraccions...). Atès que es treballa amb centenars de milers de fitxers, cal disposar de programari potent com *Replace Studio Professional*, que és una utilitat GREP (*Global Regular Expression Print*) de nova generació (Funduc.com). Els resultats de les cerques es desen en fitxers CSV (*comma separate values*), perquè són de text pla i es poden manipular amb aplicacions de full de càlcul i són compatibles amb els DBMS.

4.4. Anàlisi quantitativa

En una publicació del 2016, Kock et al. proposen mesurar la imatge amb escales multiítem, amb la pregunta «Considerant-ho tot, anar de vacances a [destinació] és» bo o dolent, positiu o negatiu, favorable o desfavorable, i val la pena o no val la pena, amb l'escala de 7 punts de Likert; i per mesurar la part afectiva «Què sent en relació amb [destinació]», simpatia o aversió, atracció o repulsió, li agrada o li desagrada, i s'hi sent còmode o incòmode. En el present estudi es tracta de fer el que es coneix com a mineria d'opinions, que ja estan donades, sobre atraccions, activitats i serveis determinats d'una destinació. L'anàlisi quantitativa dels elements paratextuals de les OTR, classificades per categories, té com a objectiu mesurar la imatge percebuda global com una addició de les imatges percebudes parcials, en les seves dimensions cognitiva i afectiva, i situades en el temps i en l'espai.

4.4.1. Categorització

L'anàlisi quantitativa de contingut textual consisteix a agrupar moltes paraules en poques categories (Stemler, 2001). Les categories són grups de paraules amb unes connotacions i/o un significat similar. Hi ha dues formes de construir-les, categorització *a priori*, basada en la capacitat deductiva de l'investigador, i categorització emergent, basada en la matriu de freqüència de paraules del text que s'ha d'analitzar (Stepchenkova, 2012).

El component afectiu de la TDI es pot avaluar des de dues fonts: les expressions que componen el títol de l'OTR i la qualificació que dona el turista a l'atracció que ha visitat o al servei que ha utilitzat.

- a) *Títol de l'OTR*. Dins de la coneguda com a mineria d'opinions, hi ha l'anàlisi de sentiments o de sensacions. El títol de l'OTR sintetitza l'opinió del seu autor i està plegat de tòpics, adjectius qualificatius i recomanacions que permeten deduir el component afectiu de la TDI. Aleshores, es tracta de construir una categoria dicotòmica sobre bones i males sensacions. Per exemple, «admirable» i «fantàstic» són adjectius positius, i «no t'ho perdís» o «s'ha de veure», recomanacions positives. Per una altra banda, «decebe-

- dor» i «avorrit» són adjectius negatius, i «compte amb els carteristes!» i «no val la pena», recomanacions negatives.
- b) *Qualificació de les OTR*. Els usuaris de TA qualifiquen les atraccions, les activitats o els serveis mitjançant un codi d'una a cinc estrelles, sent 5* la màxima nota i 1*, la mínima. Per facilitar les comparacions, les qualificacions es poden convertir a una escala decimal seguint la nomenclatura de TA: *Excellent* (10), *Good* (7,5), *Average* (5), *Poor* (2,5) i *Terrible* (0).

La part WGC del títol de les OTR té el nom de l'atracció, l'activitat o el servei i de la destinació, i l'enllaç associat amb el títol té els codis corresponents i altres dades relacionades. Per poder classificar i comptabilitzar les OTR, cal construir dues llistes o categories:

- a) *Destinacions*. La categoria territorial està formada per les nou marques turístiques, i dins de cadascuna hi ha els noms i els codis (si en tenen) de la marca i de les seves comarques, municipis i altres entitats considerades destinacions, com La Pineda (urbanització a cavall entre Vila-seca i Salou) o Baqueira Beret (centre de vacances als Pirineus).
- b) *Atraccions*. La categoria de factors d'atracció conté els noms i els codis de totes les atraccions, activitats i serveis associats a la marca on estan ubicats. Les OTR sobre serveis d'àmbit provincial resulten problemàtics de classificar, o bé s'adjudiquen a la marca de la capital de la província o es descarten.

4.4.2. *Analitzador sintàctic*

El programa analitzador dels textos, que genera les taules de freqüències, necessita saber quins caràcters són separadors de paraules, les paraules compostes i les paraules no significatives. Normalment, els programes fan servir els espais en blanc, les comes, els punts, etc., com a separadors de paraules. En aquest cas i per a més seguretat, s'han inclòs tots els caràcters que no són lletres als alfabet català, castellà i anglès. Les paraules compostes són grups de dues o més paraules que tenen un significat com a conjunt diferent del de cadascuna d'elles; per exemple, «Sagrada Família» i «No t'ho perdís». La llista negra consta de les paraules o grups de paraules que no interessin per al tipus d'anàlisi que s'està fent com la majoria d'adverbis, conjuncions, determinants, preposicions i pronoms.

L'analitzador genera taules de freqüències amb tres columnes: paraula clau, que pot ser composta; nombre d'ocurrències, i percentatge sobre el total de paraules (incloses les no significatives). L'algorisme carrega d'entrada les llistes de paraules compostes i no significatives, i crea la taula dels resultats. Un primer bucle cerca, compta, desa i suprimeix totes les paraules compostes que hi hagi al text. Un segon bucle fa el recorregut de totes les paraules que resten al text, comprova que no estiguin a la llista negra i, si estan a la taula dels resultats, incrementa la seva freqüència; si no hi estan, les dona d'alta amb freqüència u.

5. Resultats i discussió

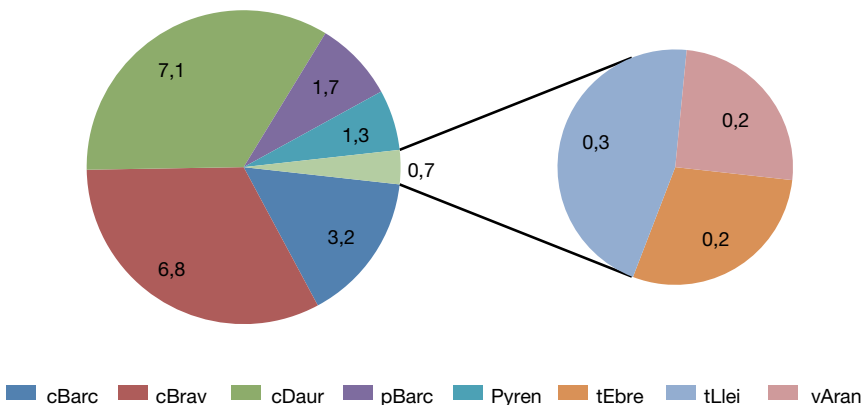
Seguint la tònica general de creixement del volum de l'UGC i de l'increment de la quantitat d'OTR vista anteriorment, la distribució temporal de les mostres aleatòries posa en evidència que, en cinc anys, les OTR en anglès s'han multiplicat per 12 i els escrits en castellà, per 30. A la figura 5, s'hi pot observar com creix el nombre d'OTR sobre Barcelona any rere any en tots els casos i passa el mateix en altres marques turístiques.

5.1. Component espacial de la TDI

Un cop filtrats els resultats de la fase de recol·lecció amb la categoria de correspondència entre destinacions i marques territorials, es posa en evidència que la majoria d'OTR fan referència a la marca Barcelona (figura 5), seguida de lluny per les marques Costa Brava i Costa Daurada (figura 6).

La figura 6 mostra el detall de la distribució territorial en percentatge de les OTR de TA que no corresponen a la marca Barcelona. En relació amb ella, només tres marques costaneres —cBarc (3,2), cBrav (6,8) i cDaur (7,1)— tenen un pes mínimament significatiu. A l'altre extrem, estan tEbre, tLlei i vAran que, entre les tres, no arriben a l'1% de les OTR de TA sobre Catalunya. Si es redueixen els resultats a la mostra de 150.000 OTR escrites en anglès, resulta que Barcelona té un 84,36% de les OTR, mentre que les altres marques es reparteixen la quota restant: cBarc (1,77%), cBrav (5,40%), cDaur (6,48%), pBarc (1,48%), Pyren (0,32%), tEbre (0,06%), tLlei (0,08%) i vAran (0,05%). I fent un zoom sobre la mostra de les 40.000 OTR escrites per turistes nord-americans, els resultats mostren una preponderància absoluta de la marca Barcelona: Barna (93,39%), cBarc (1,03%), cBrav (2,86%), cDaur (0,48%), pBarc (2,03%), Pyren (0,14%), tEbre (0,01%), tLlei (0,04%) i vAran (0,02%).

Figura 6. Percentatge d'opinions sobre les marques de fora de Barcelona (2011-2015)



Taula 4. Rànquing de 25 atraccions o serveis per país de procedència i qualificació donada

Espanya			Anglaterra			Estats Units		
Atracció/Servei	OTR	Nota	Atracció/Servei	OTR	Nota	Atracció/Servei	OTR	Nota
Sagrada Família	2.354	9,14	Sagrada Família	5.252	9,18	Sagrada Família	7.264	9,50
Port Aventura	1.490	7,40	Parc Güell	2.494	8,22	Parc Güell	2.797	8,66
Parc Güell	1.088	8,28	Camp Nou	1.892	8,74	Barri Gòtic	1.831	9,17
Casa Batlló	878	8,92	Port Aventura	1.838	8,28	Casa Batlló	1.750	9,25
Barri Gòtic	869	8,73	Casa Batlló	1.487	9,13	Palau Música	1.127	9,43
Sta. Maria Mar	803	9,36	Barri Gòtic	1.368	9,03	Les Rambles	1.103	7,21
Palau de la Música	802	9,34	Font Màgica	1.304	8,65	La Pedrera	978	8,82
Font Màgica	736	8,72	Les Rambles	1.009	6,90	Barna Day Tours	956	9,82
Camp Nou	730	8,57	Suntransfers	954	9,71	Font Màgica	756	8,46
Passeig de Gràcia	588	8,78	La Pedrera	702	8,57	Museu Picasso	681	8,20
CosmoCaixa	472	8,84	House of Illusion	676	8,84	Catedral Barna	681	8,77
Les Rambles	444	7,36	Catedral Barna	634	8,53	Montserrat	665	9,19
La Pedrera	374	8,64	Font Lluminosa	521	8,63	Trails Day Tours	639	9,80
Sandeman Tours	314	9,50	Palau Música	489	9,16	Palau Güell	587	8,86
Montserrat	308	8,84	Palau Güell	411	8,83	Passeig Gràcia	523	8,81
El Born	295	8,67	Parc Montjuïc	389	8,38	Runner Tours	520	9,75
Teatre Dalí	282	8,61	Bus Turístic	386	8,14	Camp Nou	510	8,81
Mercat de la Boqueria	266	8,20	Lloret Beach	373	6,93	Mercat Boqueria	464	9,08
Ciutat Vella	247	8,68	Costa Caribe	366	7,55	Sta. Maria Mar	460	8,95
Sant Pau	241	9,19	Runner Tours	333	9,80	Museu MNAC	422	9,03
Parc de Montjuïc	228	8,43	Sta. Maria Mar	331	8,71	Tablao Cordobés	405	8,83
Museu MNAC	224	8,93	Passeig de Gràcia	319	8,60	Sandeman Tours	388	9,57
Tibidabo	217	8,17	Aerobús	303	9,50	Montserrat Tour	363	9,18
Palau Güell	212	8,96	Zoo Barcelona	283	7,37	Parc Montjuïc	360	8,50
Catedral de Barna	205	8,22	Montserrat	281	9,05	Tire Bike Tours	320	9,48

Font: 120.000 OTR (40.000 de cada país) de TripAdvisor sobre coses per fer a Catalunya.

Comparant aquests resultats amb els de la taula 3, es pot comprovar que la distribució territorial de les OTR ha variat poc en els darrers cinc anys, tret de la deguda a la reestructuració de les marques (figures 2 i 3). El temps està donant la raó a Ortega (2011), que va escriure l'article «Catalunya: deu marques turístiques o només Barcelona?», en què va posar com a subtítol “La diversitat del sistema actual és poc útil per a la projecció internacional” i, com a compendi, “La promoció de les deu marques turístiques catalanes és una missió complicada per la poca identificació dels receptors a fora de Catalunya, i els experts aposten per exprimir el referent de Barcelona”, perquè tres marques s’han reduït a dues i s’han posat sota el paraigua de la marca Barcelona (Costa de Barcelona-Maresme, Costa de Garraf i Catalunya Central, que s’han remodelat i han passat a denominar-se Costa Barcelona i Paisatges Barcelona). Encara que la política de màrqueting sigui poc efectiva, les marques actuals responen a una especialització de la imatge en una destinació multiescalar

(Marine-Roig i Anton Clavé, 2015, 2016); però aquests autors no van poder comprovar tEbre, tLlei i vAran per insuficiència d'informació generada pels viatgers.

5.2. *Component cognitiu de la TDI*

La taula 4 mostra la quantitat d'OTR de cada atracció o servei per país de procedència (Espanya, Anglaterra i Estats Units) dels viatgers i la nota mitjana que han obtingut. Les obres mestres de l'arquitecte Antoni Gaudí hi estan destacades amb lletra negreta. A primera vista, s'observen dues coincidències a les tres columnes: la basílica de la Sagrada Família és l'atracció més visitada, amb molta diferència, i Les Rambles la pitjor valorada, així com que l'única atracció de fora de Barcelona que han visitat tots tres és el monestir de Montserrat.

A les tres columnes de la taula 4, destaquen notablement les atraccions de patrimoni tangible de Barcelona i, en especial, les cinc obres mestres de Gaudí. Entre els 25 primers classificats, pocs n'hi ha que siguin de fora de Barcelona. Tret del monestir de Montserrat esmentat abans, espanyols i anglesos: Port Aventura (cDaur); espanyols: Teatre-Museu Dalí (cBrav); anglesos: Suntransfers (cBrav), House of Illusion (cDaur), Font Lluminosa (cDaur), Lloret Beach (cBrav) i Costa Caribe (cDaur); i nord-americans: no s'han mogut de Barcelona.

5.3. *Component afectiu de la TDI*

Per una banda, la taula 4 mostra la mitjana de les qualificacions atorgades pels turistes a les atraccions i als serveis. Totes les notes són altes, dada que demostra la satisfacció general en els tres casos (espanyols, anglesos i americans). Ara bé, alguns autors (Ayeh et al., 2013; Kusumasondjaja et al., 2012) han plantejat dubtes sobre la credibilitat d'algunes OTR; per exemple, poden sorgir malfiances quan coincideix que totes les empreses privades de serveis, que es dediquen a organitzar excursions i recorreguts turístics (*tours*), tenen unes notes altíssimes a les tres columnes de la taula 4, fins i tot pel damunt de les obres mestres de Gaudí (en negreta), que són Patrimoni de la Humanitat (UNESCO, 2005).

Per una altra banda, la taula 5 mostra les 20 paraules més freqüents a la part d'UGC del títol de les OTR. La gran majoria són adjectius qualificatius positius. Aquests adjectius tenen força més pes en el cas dels visitants anglesos i americans.

5.4. *Relacions entre els components de la TDI*

Fent la comparativa entre les tres mostres de 40.000 OTR, coincideix que els americans són els que fan més turisme per Barcelona (figura 5: 93%), els que més visiten les obres de Gaudí (taula 4), els que més bones qualificacions donen a les seves obres (taula 4) i els que més citen l'arquitecte (taula 5). Això pot estar influït pel fet que una bona part dels turistes americans són passatgers de

Taula 5. Rànquing de les 20 paraules més freqüents per país de procedència

Espanya			Anglaterra			Estats Units		
total UGC: 210035	úniques: 11.013	%	total UGC: 221.961	úniques: 7239	%	total UGC: 245.653	úniques: 7156	%
visita	1.608	0,77	great	4.138	1,86	barcelona	4.512	1,84
barcelona	1.592	0,76	amazing	2.152	0,97	great	4.101	1,67
mejor	1.058	0,50	barcelona	2.007	0,90	tour	3.506	1,43
excelente	985	0,47	beautiful	1.747	0,79	beautiful	2.143	0,87
lugar	964	0,46	visit	1.512	0,68	amazing	1.997	0,81
impresionante	941	0,45	worth	1.497	0,67	gaudí	1.626	0,66
paseo	920	0,44	good	1.484	0,67	best	1.422	0,58
bonito	871	0,41	tour	1.281	0,58	experience	1.358	0,55
experiencia	845	0,40	place	1.237	0,56	must see	1.346	0,55
genial	831	0,40	fantastic	1.187	0,53	fun	1.229	0,50
interesante	816	0,39	excellent	1.144	0,52	way	1.163	0,47
espectacular	813	0,39	best	1.107	0,50	worth	1.163	0,47
imprescindible	760	0,36	lovely	1.060	0,48	place	1.046	0,43
buen	755	0,36	fun	1.022	0,46	wonderful	996	0,41
parque	755	0,36	nice	904	0,41	visit	949	0,39
gran	729	0,35	stunning	861	0,39	nice	849	0,35
vistas	717	0,34	way	829	0,37	trip	840	0,34
precioso	683	0,33	views	825	0,37	excellent	801	0,33
recomendable	663	0,32	gaudí	806	0,36	city	743	0,30
buena	655	0,31	beach	774	0,35	interesting	711	0,29

Font: 120.000 OTR (40.000 de cada país) de TripAdvisor sobre coses per fer a Catalunya.

creuers que fan escala o que tenen el port base a Barcelona (Garay Tamajón, 2015). A l'altre extrem, estan els turistes espanyols, que són els que menys estan a Barcelona (figura 5: 57%), per tant, els que menys visiten les obres de Gaudí (taula 4), menys citen l'arquitecte (taula 5) i, potser per això, són els que donen menys pes a les percepcions positives (taula 5).

Pel que fa a les experiències fora de Barcelona, els anglesos són els que, de les 40.000 OTR, en tenen més sobre la Costa Daurada (6.269) i, concretament, sobre Salou (taula 4): Port Aventura (1.838), House of Illusion (676), Font Lluminosa (521) i Costa Caribe (366), així com són els que més citen la paraula clau *platja* (taula 5, *beach*: 774).

6. Consideracions finals

Segons les enquestes estudiades, fetes a viatgers americans i europeus, les OTR tenen una gran influència en la projecció de la TDI, molt pel damunt d'altres fonts d'informació com les webs de proveïdors de serveis i de la destinació, i de les agències de viatges i oficines de turisme. En aquest estudi, s'ha comprovat que els elements paratextuals que envolten les OTR són útils per situar-les en

el temps i en l'espai, així com per classificar-les per temes, idiomes i país de procedència de l'autor. Això permet fer estudis de tendències temporals i situar-les en el context de les marques turístiques territorials, així com segmentar les OTR en base als altres ítems esmentats, en particular o en combinació.

El volum de dades massives analitzades representa la majoria d'OTR de TripAdvisor que reuneixen les condicions plantejades pel cas d'estudi (opinions sobre coses per fer a Catalunya, escrites en castellà o en anglès, entre 2011 i 2015, per turistes que tenen posat el país de procedència al seu perfil) i, per tant, els resultats obtinguts són representatius i fiables. El fet de tenir correctament emplenat el perfil dona un plus de credibilitat a l'autor (Kusumasondjaja et al., 2012).

Pel que fa a la distribució temporal, s'ha observat un creixement espectacular de la quantitat d'OTR. En relació amb la distribució territorial, s'ha comprovat que les polítiques de promoció i de projecció internacional de les marques de l'interior del país no donen resultat. En tot el període estudiat i en èpoques anteriors, hi ha una incongruència palmària entre l'oferta i la demanda. Tal com s'ha dit abans, l'ACT (2016) dedica el mateix espai del seu dossier per promocionar internacionalment la marca Barcelona que per a cadascuna de les altres vuit marques, mentre que els turistes estrangers dediquen més del 80% de les seves OTR a la metròpolis i, en el cas dels nord-americans, més del 90%. Això no treu que les marques estiguin ben dissenyades des del punt de vista d'especialització de la imatge en una destinació multiescalar.

L'anàlisi del component cognitiu de la TDI ha fet palès que la imatge es concentra en el patrimoni tangible de Barcelona i, molt especialment, en les obres mestres de l'arquitecte Antoni Gaudí. En el cas dels turistes espanyols i anglesos, també té molt de pes Salou (cDaur) gràcies al parc temàtic *PortAventura World*. L'estudi del component afectiu de la TDI ha permès comprovar que, per part dels turistes, hi ha una gran satisfacció, reflectida en les notes excel·lents que atorguen a les atraccions visitades i per l'ús massiu d'adjectius qualificatius positius als títols de les OTR. En conclusió, la imatge global de Catalunya percebuda pels turistes i projectada a través de les OTR, com a suma dels seus components cognitiu i afectiu, és molt positiva, però està centrada en la marca Barcelona i, en particular, en l'obra de Gaudí.

6.1. *Implicacions acadèmiques*

Des del punt de vista teòric, es considera el component espacial de la TDI, que està molt poc estudiat en el camp del turisme, diferenciant-lo del component cognitiu que respon més a la pregunta «Què és?» que no pas a «On és?», i ambdues qüestions són rellevants en la composició de la imatge percebuda. En aquest sentit, s'aporten dades objectives, corresponents al quinquenni 2011-2015, a la polèmica sobre denominació i àmbit geogràfic de les marques turístiques encetada l'any 2011 per la directora general de Turisme, en haver constatat que de totes les marques territorials només funcionaven a l'estranger

Catalunya, Barcelona, Costa Brava i Costa Daurada (Justícia i Benvenuty, 2011; Ortega, 2011).

La metodologia proposada permet manipular dades massives amb opinions lliurement expressades per visitants de la destinació. Aleshores, en no tenir gairebé limitació de volum, es pot treballar amb mostres amb un alt grau de representativitat de la població analitzada i, per tant, se'n poden extreure coneixements útils.

6.2. Implicacions gerencials

Com a font d'informació espontània procedent del costat de la demanda, les OTR es consideren fiables i tenen un gran interès per a les organitzacions de màrqueting de la indústria turística i per als investigadors del ram, perquè l'anàlisi de les opinions, preferències i valoracions escrites desinteressadament pels turistes pot servir per optimitzar algun producte o servei de la cadena de proveïment. El mètode proposat és força econòmic en comparació amb les enquestes. Un cop adquirits o implementats els programes informàtics i construïdes les llistes de paraules compostes i no significatives, així com de paraules o frases positives i negatives, es poden obtenir resultats gairebé sense cost.

6.3. Limitacions

Atès que una part de la fiabilitat del mètode proposat ve donada per l'anàlisi de dades massives, no es considera convenient aplicar-lo a marques turístiques com Terres de l'Ebre, Terres de Lleida i Val d'Aran, perquè la quantitat d'OTR no arriba ni a l'1% de les mostres i els resultats no serien homologables. Per exemple, sumades les OTR de les tres marques esmentades, no arriben ni a una mil·lèsima part de la mostra de les nou marques en el cas analitzat dels turistes nord-americans. Per una altra banda, encara que es parli d'intel·ligència artificial, la mineria d'opinions no és una ciència exacta. Per exemple, és pràcticament impossible que un programa informàtic capti la ironia o alguns girs del llenguatge en el cas del component afectiu de la imatge.

El procediment que s'ha fet servir per extreure les mostres té una gran fiabilitat perquè genera números aleatoris entre zero i u, amb 15 posicions decimals, amb el que resulta pràcticament impossible que surtin números repetits. Ara bé, l'aleatorietat no garanteix que la mostra sigui totalment representativa de la població (dates, destinacions, atraccions, etc.).

6.4. Noves línies de treball

Per analitzar més qualitativament la imatge percebuda, sense haver de fer o de replicar costoses enquestes, cal disposar de categories de paraules o frases representatives de les opinions i recomanacions positives o negatives dels visitants. En principi, hi ha molta producció científica relativa a l'idioma anglès (Gindl et al., 2010) i no tanta en altres idiomes com el castellà. S'ha de tenir present

que les categories dicotòmiques poden tenir un cert grau d'indeterminació perquè, per exemple, una destinació relaxant (sentiment positiu) per a una persona, pot ser avorrida (sentiment negatiu) per a una altra. El repte és construir categories (lèxics) que permetin fer mineria d'opinions escrites per extreure'n la imatge percebuda (i projectada) pels visitants de la destinació.

Referències bibliogràfiques

- ACT. (2016). *Dossier de premsa 2016: Turisme*. Barcelona, Catalonia: Agència Catalana de Turisme.
- ARCARONS, R. (2009). «Organització espacial de l'Administració Turística». En F. LOPEZ PALOMEQUE (ed.). *Atles del turisme a Catalunya*. Barcelona: Generalitat de Catalunya.
- AYEH, J. K.; AU, N. i LAW, R. (2013). «Do we believe in TripAdvisor?. Examining credibility perceptions and online travelers' attitude toward using user-generated content». *Journal of Travel Research*, 52 (4), 437-452.
<<https://doi.org/10.1177/0047287512475217>>
- BAKA, V. (2016). «The becoming of user-generated reviews: Looking at the past to understand the future of managing reputation in the travel sector». *Tourism Management*, 53, 148-162.
<<https://doi.org/10.1016/j.tourman.2015.09.004>>
- BALOGLU, S. i MCCLEARY, K. W. (1999). «A model of destination image formation». *Annals of Tourism Research*, 26 (4), 868-897.
<[https://doi.org/10.1016/S0160-7383\(99\)00030-4](https://doi.org/10.1016/S0160-7383(99)00030-4)>
- BASARAN, U. (2016). «Examining the relationships of cognitive, affective, and conative destination image: A research on Safranbolu, Turkey». *International Business Research*, 9 (5), 164-179.
<<https://doi.org/10.5539/ibr.v9n5p164>>
- BEERLI, A. i MARTÍN, J. D. (2004). «Factors influencing destination image». *Annals of Tourism Research*, 31 (3), 657-681.
<<https://doi.org/10.1016/j.annals.2004.01.010>>
- CHON, K.-S. (1990). «The role of destination image in tourism: A review and discussion». *The Tourist Review*, 45 (2), 2-9.
<<https://doi.org/10.1108/eb058040>>
- COX, C.; BURGESS, S.; SELBITTO, C. i BUULTJENS, J. (2009). «The role of user-generated content in tourists' travel planning behavior». *Journal of Hospitality Marketing and Management*, 18 (8), 743-764.
<<https://doi.org/10.1080/19368620903235753>>
- CROMPTON, J. L. (1979). «An assessment of the image of Mexico as a vacation destination and the influence of geographical location upon that image». *Journal of Travel Research*, 17, 18-23.
<<https://doi.org/10.1177/004728757901700404>>
- DATZIRA-MASIP, J. i POLUZZI, A. (2014). «Brand architecture management: The case of four tourist destinations in Catalonia». *Journal of Destination Marketing and Management*, 3 (1), 48-58.
<<https://doi.org/10.1016/j.jdmm.2013.12.006>>
- DE BORDA, J. C. (1781). «Mémoire sur les élections au scrutin». A: *Mémoire de l'Académie Royal*. Paris: Histoire de l'Académie des Sciences, 657-665.

- ECHTNER, C. M. i RITCHIE, J. R. B. (1991). «The meaning and measurement of destination image». *Journal of Tourism Studies*, 2 (2), 2-12.
- EUROBAROMETER (2016). *Flash Eurobarometer 432: Preferences of Europeans towards tourism*. Brussel·les (Bèlgica).
- EUROSTAT (2016). «Tourism». A: M. KOTZEVA (ed.). *Eurostat regional yearbook 2016*. Luxemburg: Publications Office of the European Union, 177-192.
- FAKEYE, P. C. i CROMPTON, J. L. (1991). «Image differences between prospective, first-time, and repeat visitors to the Lower Rio Grande Valley». *Journal of Travel Research*, 30 (2), 10-16.
<<https://doi.org/10.1177/004728759103000202>>
- GALLARZA, M. G.; GIL SAURA, I. i CALDERÓN GARCÍA, H. (2002). «Destination image. Towards a conceptual framework». *Annals of Tourism Research*, 29 (1), 56-78.
- GARAY TAMAJÓN, L. A. (2015). «Luces y sombras del turismo de cruceros: el caso de Barcelona». *Documents d'Anàlisi Geogràfica*, 61 (3), 563-580.
<<https://doi.org/10.5565/rev/dag.259>>
- GARTNER, W. C. (1993). «Image formation process». *Journal of Travel and Tourism Marketing*, 2 (2-3), 191-215.
<https://doi.org/10.1300/J073v02n02_12>
- GENETTE, G. (1997). *Paratexts: Thresholds of interpretation*. Nova York, USA: Cambridge University Press.
- GINDL, S.; WEICHELSEBRAUN, A. i SCHARL, A. (2010). «Cross-domain contextualization of sentiment lexicons». A: H. COELHO, R. STUDER i M. WOOLDRIDGE (ed.). *19th European Conference on Artificial Intelligence*. Washington, USA: IOS Press, 771-776.
<<https://doi.org/10.3233/978-1-60750-606-5-771>>
- GONZÁLEZ, R. (2010). *Estudi exploratori de les marques turístiques de Catalunya a la Web 2.0*. Barcelona: Vivencial Value (no publicat).
- GUNN, C. A. (1972). *Vacationscape: Designing tourist regions*. Austin, TX: Bureau of Business Research, University of Texas.
- JABREEL, M.; MORENO, A. i HUERTAS, A. (2016). «Semantic comparison of the emotional values communicated by destinations and tourists on social media». *Journal of Destination Marketing and Management*, 6 (3), 170-183.
<<https://doi.org/10.1016/j.jdmm.2016.03.004>>
- JUSTÍCIA, A. i BENVENUTY, L. (2011). «Catalunya duda de la idoneidad de tener diez marcas turísticas». *La Vanguardia*. Barcelona. <<http://www.lavanguardia.com/ocio/20110531/54162863546/catalunya-duda-de-la-idoneidad-de-tener-diez-marcas-turisticas.html>>.
- KAPLAN, A. M. i HAENLEIN, M. (2010). «Users of the world, unite! The challenges and opportunities of Social Media». *Business Horizons*, 53 (1), 59-68.
<<https://doi.org/10.1016/j.bushor.2009.09.003>>
- KOCK, F.; JOSIASSEN, A. i ASSAF, A. G. (2016). «Advancing destination image: The destination content model». *Annals of Tourism Research*, 61, 28-44.
<<https://doi.org/10.1016/j.annals.2016.07.003>>
- KUSUMASONDJAJA, S.; SHANKA, T. i MARCHEGIANI, C. (2012). «Credibility of online reviews and initial trust: The roles of reviewer's identity and review valence». *Journal of Vacation Marketing*, 18 (3), 185-195.
<<https://doi.org/10.1177/1356766712449365>>
- LI, J.; ALI, F. i KIM, W. G. (2015). «Reexamination of the role of destination image in tourism: An updated literature review». *E-Review of Tourism Research*, 12 (3-4), 191-209.

- LIU, Y.; TEICHERT, T.; ROSSI, M.; LI, H. i HU, F. (2017). «Big data for big insights: Investigating language-specific drivers of hotel satisfaction with 412,784 user-generated reviews». *Tourism Management*, 59, 554-563.
<<https://doi.org/10.1016/j.tourman.2016.08.012>>
- LLODRÀ-RIERA, I.; MARTÍNEZ-RUIZ, M. P.; JIMÉNEZ-ZARCO, A. I. i IZQUIERDO-YUSTA, A. (2015). «A multidimensional analysis of the information sources construct and its relevance for destination image formation». *Tourism Management*, 48, 319-328.
<<https://doi.org/10.1016/j.tourman.2014.11.012>>
- LYNCH, K. (1960). *The image of the city*. Cambridge, MA: The MIT Press.
- MARINE-ROIG, E. (2013). *From the projected to the transmitted image: the 2.0 construction of tourist destination image and identity in Catalonia*. Tesi doctoral. <<http://hdl.handle.net/10803/135006>>
- (2014). «A webometric analysis of travel blogs and review hosting: The case of Catalonia». *Journal of Travel and Tourism Marketing*, 31 (3), 381-396.
<<https://doi.org/10.1080/10548408.2013.877413>>
- (2015). «Identity and authenticity in destination image construction». *Anatolia: An International Journal of Tourism and Hospitality Research*, 26 (4), 574-587.
<<https://doi.org/10.1080/13032917.2015.1040814>>
- (2017). «Online travel reviews: A massive paratextual analysis». A: Z. XIANG i D. R. FESENMAIER (eds.). *Analytics in Smart Tourism design: Concepts and methods*. Heidelberg: Springer, 179-202.
<https://doi.org/10.1007/978-3-319-44263-1_11>
- MARINE-ROIG, E. i ANTON CLAVÉ, S. (2015). «Tourism analytics with massive user-generated content: A case study of Barcelona». *Journal of Destination Marketing and Management*, 4 (3), 162-172.
<<https://doi.org/10.1016/j.jdmm.2015.06.004>>
- (2016). «Perceived image specialisation in multiscale tourism destinations». *Journal of Destination Marketing and Management*, 5 (3), 202-213.
<<https://doi.org/10.1016/j.jdmm.2015.12.007>>
- O'REILLY, T. (2007). «What is Web 2.0: Design patterns and business models for the next generation of software». *International Journal of Digital Economics*, 65, 17-37.
<<http://mpa.ub.uni-muenchen.de/4580/>>
- OdEiO. (2016). *Balanç turístic anual 2015*. Barcelona: Observatori d'Empresa i Ocupació.
- ORTEGA, M. (2011). «Catalunya: deu marques turístiques o només Barcelona?». *Ara.cat*. Barcelona. <http://www.ara.cat/cronica/Catalunya-deu-marques-nomes-Barcelona_0_458954128.html>.
- POCOCK, D. i HUDSON, R. (1978). *Images of the urban environment*. Londres: MacMillan.
- RICHARDS, G. i WILSON, J. (2004). «The use of cultural events in city promotion: Rotterdam cultural capital of Europe 2001». *Urban Studies*, 41 (10), 1931-1951.
<<https://doi.org/10.1080/10295390208718729>>
- SCHUCKERT, M.; LIU, X. i LAW, R. (2015). «Hospitality and tourism online reviews: Recent trends and future directions». *Journal of Travel and Tourism Marketing*, 32 (5), 608-621.
<<https://doi.org/10.1080/10548408.2014.933154>>
- SHAFIEE, S. i GHATARI, A. R. (2016). «Big data in tourism industry». A: *10th International Conference on e-Commerce with focus on e-Tourism*. IEEE, 1-7.
<<https://doi.org/10.1109/ECDC.2016.7492979>>

- STATISTA (2016). *Leisure travel: planning methods US statistic*. Hamburg: Statiste GmbH.
- STEMLER, S. (2001). «An overview of content analysis». *Practical Assessment, Research and Evaluation*, 7 (17). <<http://pareonline.net/getvn.asp?v=7&n=17>>
- STEPCHENKOVA, S. (2012). «Content analysis». A: L. DWYER, A. GILL i N. SEETARAM (eds.). *Handbook of research methods in Tourism: Quantitative and qualitative approaches*. Cheltenham, UK: Edward Elgar Publishing, 443-458.
- TASCI, A. D. A.; GARTNER, W. C. i TAMER CAVUSGIL, S. (2007). «Conceptualization and operationalization of destination image». *Journal of Hospitality and Tourism Research*, 31 (2), 194-223.
<<https://doi.org/10.1177/1096348006297290>>
- TRIPADVISOR (2016a). «Basilica of the Sagrada Família». <https://www.tripadvisor.com/Attraction_Review-g187497-d190166-Reviews-Basilica_of_the_Sagrada_Familia-Barcelona_Catalonia.html>
- (2016b). «Things to Do in Catalonia». <<https://www.tripadvisor.com/Attractions-g187496-Activities-Catalonia.html>>
- UNESCO (2005). «Works of Antoni Gaudí». Extret el 5 de maig de 2016. <<http://whc.unesco.org/en/list/320>>
- W3SCHOOLS. (2016). *HTML Tutorial*. <<http://www.w3schools.com/html/>>
- YE, Q.; LAW, R.; GU, B. i CHEN, W. (2011). «The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings». *Computers in Human Behavior*, 27 (2), 634-639.
<<https://doi.org/10.1016/j.chb.2010.04.014>>
- YLIJOKI, O. i PORRAS, J. (2016). «Conceptualizing big data: Analysis of case studies». *Intelligent Systems in Accounting, Finance and Management*, 23 (4), 295-310.
<<https://doi.org/10.1002/isaf.1393>>
- YOO, K.-H.; SIGALA, M. i GRETZEL, U. (2016). «Exploring TripAdvisor». A: R. GGER, I. GULA i D. WALCHER (eds.). *Open tourism: Open innovation, crowdsourcing and co-creation challenging the tourism industry*. Berlín: Springer, 239-255.
<https://doi.org/10.1007/978-3-642-54089-9_17>
- YUAN, Y.-L. i HO, C.-I. (2015). «Rethinking the destination marketing organization management in the big data era». A: *ASE BigData and SocialInformatics 2015*. Nova York, NY: ACM Digital Library, article 60.