Patrolling TripAdvisor: Re-purposing large data-sets in Tourism research.

Par Saskia Cousin, Gaël Chareyron, Jérôme Da-Rugna et Sébastien Jacquot. Le 29 August 2014

From the beaches of Costa Rica to the Beauval Zoo or the Royal Palaces of Abomey, making and breaking the *e-reputation* of restaurants, hotels and tourist attractions, TripAdvisor has become an essential tool for millions of tourists. With 125 million reviews/comments and 57 million members, TripAdvisor publicises itself as « the world's largest travel site » and finds itself increasingly subject to the analytical gaze of a research community which, more often than not, tends simply to mention TripAdvisor's ranking of a given tourist site, or draws on TripAdvisor when developing marketing models or proposals for the improvement of *e-reputations*. Currently, the most interesting research considers the emergence of the non-professional (amateur) review(er) and the concurrent transformation of expertise. Yet even in this context, research with a focus on the metadata associated with such reviews, or which considers the site of such commentaries as an object in itself, is very rare[1]. Such work requires distinguishing between two very different types of data : firstly that concerning the company's communication practices, observable in the manner in which partner's adverts and individual('s) opinions are presented and also, less visibly, in the algorithms favoured by the site ; secondly the web users comments themselves, which require contextualising in terms of social, national and stylistic origins, as well as their enrolment in logics of imitation or distinction (to what extent are comments and reviews produced in reaction to other comments/reviews, for example). These are the concerns of a research programme currently being undertaken on digital footprints[2]. Here, we offer something more modest and playful : based on data available on TripAdvisor, we propose to *trip-patrol* certain data and render it in cartographic form, revealing an alternative account to that offered by the communications company TripAdvisor itself.

TripAdvisor, or the new specialists.

Map 1 : The world according to TripAdvisor. A map of sites referenced by TripAdvisor in April 2014 ; only those sites containing GPS coordinates are drawn. Source : Gaël Chareyron, Saskia Cousin, Jérôme Da Rugna and Sébastien Jacquot, 2014.

The site works on a very basic principle : after their trip, tourists — or at least a small number of them — post comments/reviews and award from 1 to 5 points to each place or activity. Such ratings allow the ranking of activities by category and produce consumer advice aimed at prospective travellers. To achieve this, TripAdvisor uses an algorithm — the formula of which is a well guarded secret — which integrates as key variables the quantity, quality and time elapsed since registration of the ratings.

Combining advice and tourist's reflections on their consumption, TripAdvisor is thus participating in a contemporary revolution in the tourism industry : the demise of a model based on placing trust in professional experts and/or businesses providing touristic goods and services, in favour of the emergence of intermediaries, who, like TripAdvisor, have invented an economic model based on the valorization of non-commercial or interpersonal exchanges (advice between peers, apartment swaps, couch surfing, etc.) and *e-reputation* consultancy services offered to traditional service providers. As Edmund Burke wrote bitterly in 1791, « the world is governed by go-betweens ». Yet this model presents itself as the antithesis of the traditional go-betweens, who were themselves the producers of knowledge concerning sites, activities and destinations. Thus, in 2009, in reference to his partnership with UNESCO, the managing director of TripAdvisor suggested that knowledge is produced by « the collective wisdom and support of TripAdvisor's millions of travellers, and their trusted insights » (UNESCO statement, 20th July 2009).

The majority of commentators on websites believe they are participating in a strictly egalitarian system. Each person believes they are in the right and that it is worth giving their opinion on a particular restaurant, hotel, destination or World Heritage site. They receive, by way of a counter gift, differing opinions, other advice, all contributing towards a « collective intelligence ». This love of exchange, but also the success of marketing the gift exchange, raises fascinating questions for the anthropologist.

Sociologists can uncover, in both the discourses of the websites and the Internet users themselves, questions concerning the practices and legitimacy of the non-professional, the amateur (Flichy 2010). Similarly relevant are Passeron's three stages of cultural action (1991) : *legiticism* which

aims to convince people to confirm expert choices ; *populism* which confers on popular culture, here the *vox populi* of TripAdvisor, a legitimacy to say what is good or not ; and finally, *revolutionism* which, according to Passeron in regard to cultural politics, sets out to erase the boundaries between the viewpoints of people and elites. The sociologist might also remark that such touristic democracy raises at least as many problems as the democratisation of holidays which preceded it : besides the recurring question of fake recommendations[3], biases are innumerable, while equality of opportunity to participate and the legitimacy of any of the viewpoints remain a matter of opinion.

In effect, if we put the principal problem to one side — the issue of the *sur*veillance (from above) or of the *sou*veillance (from below) of the Internet along with the potential end of a libertarian utopia (Beaude 2014) — « digital democracy » (Hindman 2009), here understood as the combination of consumerism and collective intelligence, democratisation by "inclusion" (of the totality of the offer) or by participation (of lay consumers) also raises questions (Cardon 2010, Pasquier 2014). In the system of shared recommendations, now effectively standard practice, we co-produce that which we consume (Dujarier 2008) but in a distinctive fashion. In effect, in the same way as with the evaluation of cooking, we observe « a tension peculiar to lay evaluation, as in the end it is always the least lay of the laity, the most professional of the amateurs, who gain prominence » (Beauvisage et al. 2014, p. 201). Thus contributors are differentiated by a position which includes most notably the regularity of their participation in evaluations.

If you need convincing, just look at the way TripAdvisor accords value to the data posted by Internet users, with its famous scoring algorithm, but also the ranking of sites, towns and « favourite » destinations at a world, continental or national scale. Based on the most common queries by French people on the website and an « online survey » of their expectations, Artiné Mackertichian, TripAdvisor's spokesperson, presents us with the favourite destinations of the « French » :

The trends of the TripAdvisor community reflect what is essentially the charm of travelling — adventure, foreign cultures and local colour. This outlines a tendency for this year : French travellers are particularly attracted (*sic*) by the combination of the attractions of big cities and outdoor activities. For those who are planning a journey to these cities, the destination pages offered on our site (TripAdvisor London, TripAdvisor Marrakech…) constitute a mine of information in terms of the Best hotels, restaurants and attractions. Consult them and enjoy an excellent voyage[4].

It's quite clear that such accounts tell us more about the public image of the business than the desires and practices of Internet users. TripAdvisor owns around 20 brands, and has just bought *lafourchette.com*, which represents the antithesis of elitism and is currently dominating media reports concerning tourism, with its almost weekly « revelations » concerning the rankings of hotels, monuments, zoos, beaches, etc.

The challenge here is to develop an analysis which, while drawing on TripAdvisor, offers a different perspective to the company's marketing department and avoids simply reiterating the hierarchy of destinations and activities based on ranking and enquiries, which is the very business of TripAdvisor itself. Taking an approach known as « inclusionist », in other words according equal value to each comment and commentator, two principal "ways in" emerge, which of course need to be combined : an analysis of the content of comments posted on the one hand, and the information available about the posters (*via* the metadata) on the other. This is the challenge of the

experimental programme of studies of digital traces which we have been developing for four years, an amalgam of information technology, geography and social anthropology (Chareyron, Cousin et al. 2014). The work consists in extracting the metadata and other data accessible from a variety of sites, including TripAdvisor, but also Hotel.com, Panoramio, Flickr, etc., and producing analyses *not* directed by marketing objectives, but proceeding according to an inductive process and developing software programs fit for purpose.

When it comes to TripAdvisor, the variety of information available relates to comments, ratings attributed to sites (attractions, hotels, restaurants), cross referenced with the metadata associated with both the comments (date of online publication, the particular tourist site, language) and the commentators (stated country and town of origin, stated gender, stated age, other sites visited, number of comments). So we are working with massive amounts of data : hundreds of thousands of photographs, millions of elements of information. The subsequent step involves developing queries that draw on sociological categories, territorial data or temporal sequences. From this, we see the emergence of concentrations and dispersions, of fluxes and of circulations, the recurrence of particular words or images. In short, like archaeologists, we work to reconstitute routes, paths and itineraries, in time and in space, from a few (million) traces and also retrieve the history of those who left them. In the following section, drawing on a number of maps, we outline what the use of such data can offer, regardless of the content of the comments or the skill of the reviewers.
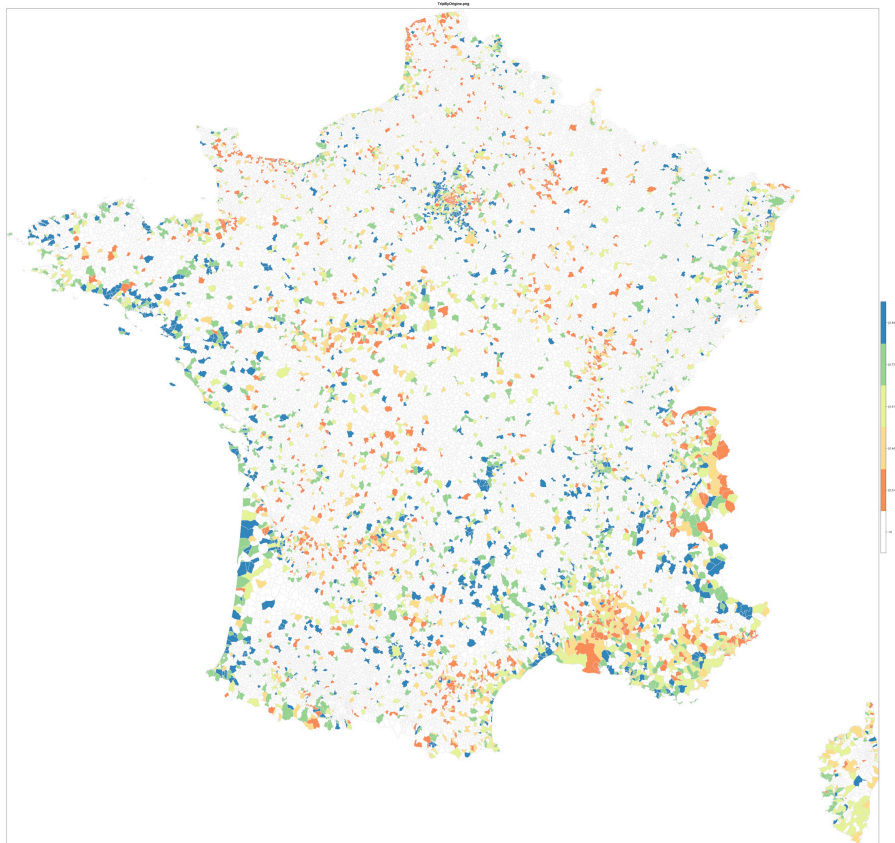
The Reviewer's France.

Map 2 : French anamorphoses. Map of French communes coloured according to the percentage of comments posted by French nationals. The categories rely on quartile calculations. The anamorphosis has been

achieved using the program ScapeToad based on the number of comments per commune. Communes with less than 10 comments where the originating source is included are not represented. N.B. In the instance of chloropleths, Paris, Lyon, and Marseille are divided by arrondissement. Source : Gaël Chareyron, Saskia Cousin, Jérôme Da Rugna and Sébastien Jacquot, 2014.

This second map is an anamorphic image based on the number of comments posted per French commune. It allows a reading of the hierarchy of French touristic space which is in sharp contrast to that proposed by TripAdvisor in two different ways. Above all, this map is distinguished from the site's orientation towards « activity » by a grouping together of tourist attractions, hotels, restaurants and shopping spaces and a focus on territorial analysis. Most importantly, on this map, weighting is not due to a score as calculated by the algorithm, but to the number of comments, their mass : it is clear that the France of mass opinions is not the same as that of positive opinions.

By focusing on the quantity of opinions offered, we can see the crushing weight of the Parisian metropolis emerge, like a great splash on the map of France. A few destinations, situated at the edges of the country do manage to compare with the Parisian metropolis : the beaches of Côte d'Azur, the Northern Alps, Alsace, Lourdes. In contrast, a France characterised by a sparse tourism appears in North East Lorrain and Ardennais, at the edges of the Paris Basin and across a large part of the Massif Central. The map also indicates the ratio of French citizens/foreigners in the comments posted for each of the French communes. TripAdvisor itself proposes a ranking of comments by the language used, but we have chosen to abstract, in preference, the nationality as declared by the user, hence comments where the nationality is not available have not been retained. The colours are functions of the quantiles : in blue, communes with the highest proportion of comments posted by the French, in dark orange, communes with the highest proportion of comments posted by non-French people (thresholds on the right-hand side of the map). This automatically generated map makes visible a France of French holidaymakers on the Atlantic coast — the Western camp sites —while the areas relating to comments made by non-French Internet users are more numerous in Paris, the Loire Valley, the Dordogne, Normandy, sites commemorating the D-Day landings, and most of all, the south-west of the Alps and the Côte d'Azur.

Map 3 : Communes through the prism of the reviewer's origins. A map of French communes coloured according to the percentage of comments posted by French citizens (TripAdvisor). The categories rely on quartile calculations. Communes with less than 10 comments where the originating source is included are not represented. Source : Gaël Chareyron, Saskia Cousin, Jérôme Da Rugna and Sébastien Jacquot, 2014.

Map 3 — which also focuses on nationality and the ratio French/non-French by commune, with the same colours but without anamorphosis — demonstrates that the proportion of foreigners is not just a function of the mass of comments. Sites in the Champagne region, flattened in the anamorphic map, emerge strongly internationalised on this map, which is however created only from those communes with at least 10 comments where the metadata includes nationality. However, the impression is still left of a great tourism vacuum, revealing the strong territorial selectivity operated by Tripadvisor's commentators, rendering a large part of the country invisible.

The age and gender of the *tripadviser.*
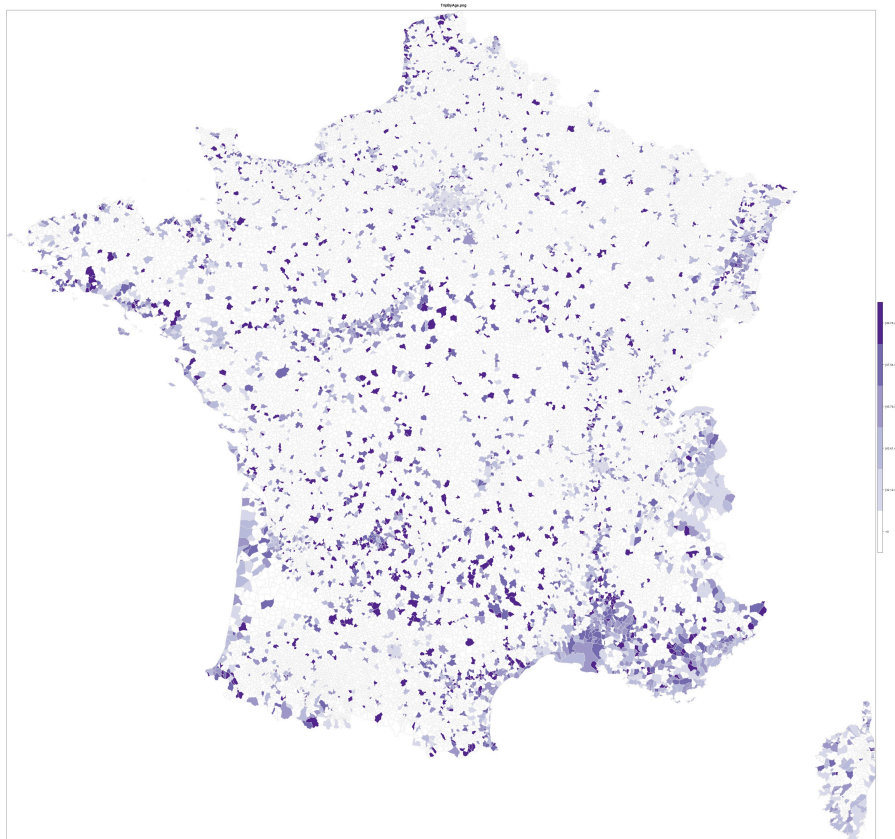
Map 4 : The age of reviewers. A map of French communes coloured according to the average age of tripadvisers. The categories rely on quartile calculations. Communes with less than 10 comments where the age is included are not represented. Source : Gaël Chareyron, Saskia Cousin, Jérôme Da Rugna and Sébastien Jacquot, 2014.

The proportion and characteristics of TripAdvisor users vary depending on the site in question : as a function of their geographical origin (see Map 2), but also their age and/or gender. Consequently, we can hypothesise that such variations might tell us something about the ways of engaging with a site, an activity or a destination and this at a variety of scales. Map 4 presents, at the national scale, the average age of TripAdvisor commentators, by commune, according to a division by quantile. The higher the average age, the darker the colour. This allows for a geography other than that of the opposition French/non-French, bringing into focus the use of French cities by a much younger population with, on the other hand, a tourism in rural spaces notable by higher average ages. We also see an opposition between mountain ranges, between the Northern Alps, with younger *tripadvisers*, and the Massif Central.

Such data can also highlight practices within a single destination and allow us to note, for example in Paris, activities and places which have the greatest proportion of women amongst the reviewers, as is illustrated in the following table :

| Nom du site	Ratio femmes / N	N = Nombre total de commentaires pour lesquels le sexe du posteur est précisé
O Kari Hammam	94.4 %	36
Le Cafe des Chats	85.7 %	49
Caulaincourt Square Hostel	78.2 %	78
Cook n With Class	76.4 %	327
Paris Muse Private Tours	76 %	75
La Cuisine Paris - Cooking Classes	75.8 %	203
Culinary Tours of Paris	75 %	60
THATLou, Treasure Hunt at the Louvre	73.1 %	26
Ricki Stevenson s Black Paris Tours	72.7 %	22
Le Massyl Restaurant	72.4 %	29

Table 1 : The ten sites and activities in Paris with the highest proportion of women amongst the reviewers. Source : Gaël Chareyron, Saskia Cousin, Jérôme Da Rugna and Sébastien Jacquot, 2014.

These sites are not, far from it, the most reviewed amongst the Parisian destinations and not all of them are specifically tourist sites. O Kari Hammam, for example, is evaluated very largely by people from greater Paris. In addition, in certain cases, the information generated in this table is at times self-evident : the Hammam makes it clear that it is « for women only ». Furthermore, in order to know whether the ratio men/women is representative of the Internet users who have posted comments, it would be necessary to verify whether men and women have the same tendency to indicate their gender and this for each type of attraction, hotels, etc. However, once these reservations are taken into account, a feminine touristic Paris does emerge, focused around activities like cooking courses and cultural discovery. On the contrary, the sites with the highest proportion of male commentators are essentially restaurants and brasseries. Do these differences, which necessitate drawing on a great number of examples, reveal different touristic practices or differences in the manner of hierarchising and recounting touristic experience ? In other words, do we have a woman's Paris, taking cooking courses, snapshots of cats, engaging in playful cultural activities, and a man's Paris, eating, eating again, and, perhaps, paying the bill ? Or are we simply observing a gendered account of that which one *ought* to have seen, tried, and, most of all, that which one *should* tell people about ? Asking such questions at the scale of the individual destination or for all sites could allow us to access the gendered dimension of practices and accounts. Of value both in themselves and to the extent that they open up new routes of enquiry, it's becoming clear that the questions we ask and the results are increasingly ethnographic.

So, thanks to the information provided by Internet users — the metadata — it is possible to sketch out a variety of maps of practices and touristic experiences based on differences in provenance, age, gender and all of this at a variety of scales, by country, region or individual commune. All of this assumes a clear ethical practice so as to preserve the anonymity of informants.

Relational Territories.
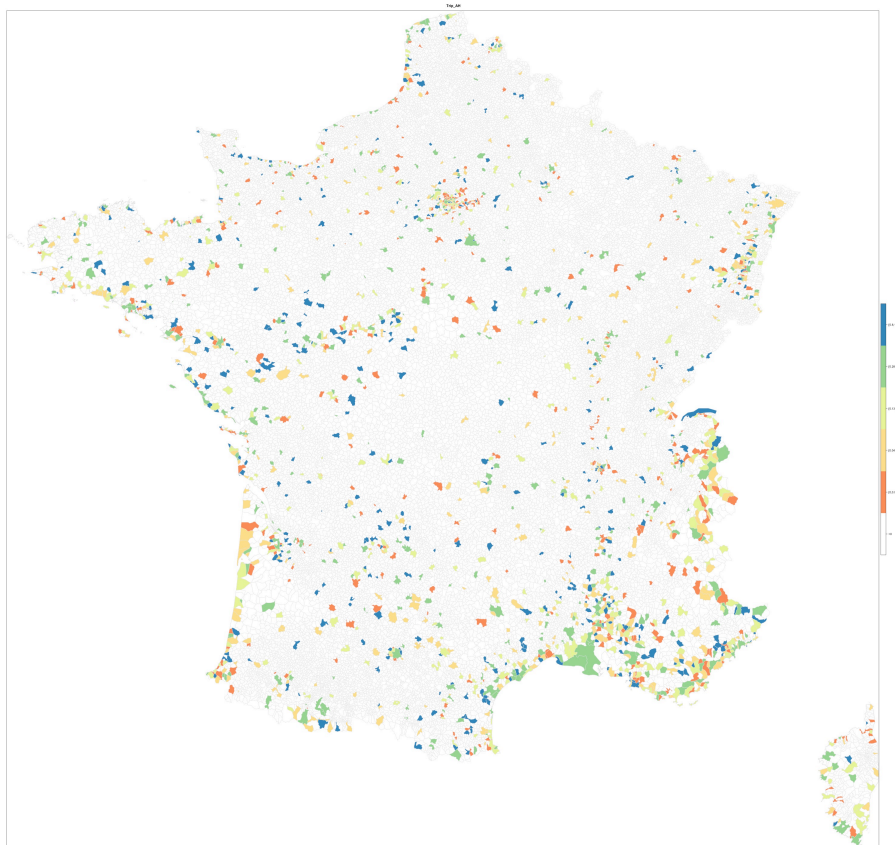
Map 5 : Relations between touristic services. A map of French communes coloured according to the ratio between the number of reviews of attractions/the number of reviews of hotels. Communes with less than 10 reviews on hotels and 10 reviews on attractions are not represented. Source : Gaël Chareyron, Saskia Cousin, Jérôme Da Rugna and Sébastien Jacquot, 2014.

The use of data from TripAdvisor also allows us to characterise spatially at a variety of scales. In effect, for each commune three types of locale are reviewed : tourist attractions (sites, museums, private tours, etc.), hotels/other accommodation and restaurants. Each territory can thus be defined by a particular relation between these three tourism service categories, a relation which depends both on an objective distribution (number of restaurants, hotels, museums) and the relative importance accorded to each of the activities as an aspect of tourists' experience, suggested by their willingness to report on this. It is equally possible to identify the links made by the Internet users and to model them. Map 5 proposes a France of relations between tourism services, based on the number of reviews by commune and the totality of attractions and hotels. This map thus presents both the great sites of tourism and those touristically dormant communes, in effect the polarisation between the centre and the periphery of touristic space. Implicitly it is possible to see several types of relations between spaces within one single regional or local destination : between highly coveted spaces and spaces ignored by tourism, for example.

To summarise, TripAdvisor retains a great variety of information on its users, their practices, their accounts, but also their distribution across the country. It is these digital traces which interest us. Evidently, it would be inappropriate to move from these users to tourists in general, a whole sociology of the use of social networks by tourists remains to be conducted (Chareyron, Da Rugna

et al. 2014). Anglophone tourists are the most numerous, doubtless more accustomed to the use of social networks in general and TripAdvisor in particular. A large proportion of tourists across the world do not publish comments nor share their holiday photographs beyond Facebook. TripAdvisor is thus characterised by a community of users that we can, on the one hand, attempt to relate to a wider population (tourists), but equally study in itself, in all its variety. For all that, for the moment, our maps correspond to the results of classic tourist visitor surveys. We are aware that because the scale is too small to enter into the details here, we learn very little that is new concerning visitor patterns, but we hope to have demonstrated the legitimacy of our research and the validity of our method of knowledge production. This is already an achievement, but the challenge is elsewhere of course : it's about paying attention to the ways and byways, how they are travelled, articulated and recounted.

In the manner of numerous social networks, but in a more fundamental and explicit way, TripAdvisor can be studied as a community producing a unique description and having thus a particular agency in the world. Effectively, TripAdvisor is much more than simply the reflection of a multitude of individual points of view : what it tells us and represents to us has a performative affect on the travel industry and its organisation, as much as on the practices and the lived experiences of millions of tourists. In fact, what we observe is not a dispersion of touristic flows but rather a hyper concentration, something that doesn't, however, stop us from identifying and working on the outskirts and peripheries of these concentrations (Chareyron, Cousin et al. 2014).

We are focusing on the meta data here, but studying TripAdvisor equally supposes the analysis of the content of the comments, their status and what they tell us about touristic practices, experiences and accounts. There are hundreds of thousands of comments, in disparate styles, in several languages, on an equally wide variety of registers (emotions before the grandiose spectacle, cleanliness of the toilets, the length of the queue, etc.), written in heterogenous styles, concerning touristic attractions and material from across the world, but also restaurants, nightclubs, shops : this is a body of material that appears as still virgin territory to any systematic analysis and we propose, so to speak, to clear the land. With these few examples, it has not simply been a matter of allowing ourselves to enjoy some cartographic *trip-patrolling*, but to suggest that as researchers not oriented towards marketing, whether in ethnography, semiotics or computer science, we are only beginning to be able to see the extent of possible research.

Bibliographie

Beaude, Boris. 2014. *Les fins d'Internet*. Limoges : Fyp.

Beaude, Boris.2015. « Making explicit digital footprints, revealing urbanity », in *A Cartographic Turn*, Elsa Chavinier et Jacques Lévy (dir.), Lausanne : EPFL Press (à paraître)

Beauvisage, Thomas et al. 2014. « Une démocratisation du marché ? Notes et avis de consommateurs sur le Web dans le secteur de la restauration » *Réseaux*, n° 183 : p. 163-204.

Burke, Edmund. 1791. « An Appeal from the new to the old whigs, in consequence of some late discussions in Parliament, relative to the reflections on the French Revolution » in *The Works of the Right Honourable Edmund Burke*, Vol. IV (Projet Gutenberg).

Casilli, Antonio A. 2014. « Anthropologie et numérique : renouvellement méthodologique ou

reconfiguration disciplinaire ? » *Anthrovision*.

Chareyron, Gaël, Jérôme Da-Rugna et Bérengère Branchet. 2013. « Mining Tourist Routes Using Flickr Traces » *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, Ontario, Canada, septembre 2013.

Chareyron, Gaël, Saskia Cousin, Daniel Gabay, Se?bastien Jacquot et Je?ro?me Da-Rugna. 2014. « La me?tropole du data mining : ce que l'exploration du web nous apprend des pratiques et imaginaires me?tropolitains » *Les cahiers de la métropole*, Hors-série : p. 54-57.

Chareyron, Gaël, Je?ro?me Da-Rugna, Saskia Cousin, Maxime Michaud, Sairi Piñeros et Bérengère Branchet. 2014. « Observer les pratiques touristiques en croisant traces numériques et observation ethnographique. Le projet de recherche Imagitour » *Espaces*, n° 316 : p. 99-107.

Dujarier, Anne-Marie. 2008. *Le travail du consommateur. De Mcdo à Ebay : comment nous coproduisons ce que nous achetons*. Paris : La Découverte.

Flichy, Patrice. 2010. *Le Sacre de l'amateur. Sociologie des passions ordinaires à l'ère numérique*. Paris : Seuil.

Hindman, Matthiew. 2009. *The myth of digital democracy*. Princeton : Princeton University Press.

Pasquier, Dominique. 2014. « Les jugements profanes en ligne sous le regard des sciences sociales » *Réseaux*, n° 183 : p. 9-25.

Passeron, Jean-Claude. 1991. *Le Raisonnement sociologique*, chap. XIII, « Figures et contestations de la culture. Légitimité et relativisme culturel ». Paris : Nathan.

Note

1 L'affirmation est appuyée par une étude de mesure d'audience.

2 Pour un état des lieux des travaux sur ces questions, voir Beaude 2014.

3 Cette recherche est notamment financée par la Région Centre Val de Loire (France) dans le cadre d'un Programme Pilote Régional (PPR) intitulé *Imagitour*.

4 Alors que des hôteliers et associations de consommateurs s'inquiètent dans différents pays d'avis postés en masse, pour augmenter artificiellement la notoriété d'un élément ou dénigrer un concurrent, TripAdvisor a refusé de signer la norme Afnor mise au point pour réguler les avis des consommateurs, en arguant que ses logiciels étaient plus puissants et mieux adaptés.

5 Communiqué de presse du 24 juin 2014, « TripAdvisor dévoile les tendances de voyage pour cet été. Où et quand les Français comptent partir ? ». On y indique « Enquête menée auprès de 1363 voyageurs français du 16 au 19 juin 2014 ».

Article mis en ligne le Friday 29 August 2014 à 10:48 –

Pour faire référence à cet article :

Saskia Cousin, Gaël Chareyron, Jérôme Da-Rugna et Sébastien Jacquot,"Patrolling TripAdvisor: Re-purposing large data-sets in Tourism research.", *EspacesTemps.net*, Publications, 29.08.2014
https://www.espacestemps.net/en/articles/etudier-tripadvisor/