

Searching for Home in the Historic Web: An Ethnosemiotic Study of London-French Habitus as Displayed in Blogs

Saskia Huc-Hepher, University of Westminster

From Big Data to Thick Data

The ethnosemiotic case study presented here lies within the context of a broader doctoral research project on the French community in London. Ethnosemiotics, for the purpose of both studies, and as the architecture of the term implies, is to be understood as a combination of **ethnographic** research and **semiotic** analysis. In the overarching doctoral study, the ethnographic methods comprise immersive, 'on-land' empirical data-collection techniques typical of the field, including interviews, questionnaires, focus groups and field-notes, in conjunction with a more innovative practice-based mechanism, namely, the curation of the London French Special Collection (LFSC) in the UK Web Archive (UKWA), a corpus of online primary sources selected to complement the on-land data and triangulate the ethnographic findings. The semiotic analytical model chosen draws on the *social* semiotic school, in particular multimodality, involving a fine-grained reading of the online resources, the principles of which will be set out below.

In many ways, the objectives of this case study have been defined by its precursor, the AADDA Project case study, as it was through a series of somewhat naïvely over-ambitious geo-indexing and image-tag trials using the JISC UK Domain Dataset (1996-2010), in its nascent, incomplete form, that a crucial lesson was learnt: think *small* when handling *big data*. The focus of this piece of research is, therefore, necessarily on the 'thick description' (Geertz, 1973: 7) of qualitative, 'thick data' (Wang, 2013: 1), as opposed to an arguably more fitting but inherently 'thinner' quantitative approach, the particular object of analysis being London-French *blogs* and the extent to which they are an immaterial manifestation of the material lives and culture of the capital's French community. According to Yoon, blogs 'have the characteristics of personal journals' (Yoon, 2013: 175) and offer a privileged 'window into the past', providing onlookers with evidence of 'social characteristics and changes' (*ibid.*, quoting O'Sullivan, 2005), since the lone blogger cannot be dissociated from the spatio-temporal and sociocultural context in which s/he is writing (*ibid.*, citing Halbwachs, 1992). Furthermore, the design and 'interior furnishings' of blogs mirror the material habitat (Casilli, 2008: 133; Miller, 2012: 156) and, by extension, the internalised habitus of the

blogger, offering insights into their positioning within the migration setting and the extent to which they are 'at home' in London. Further validation for the narrow focus on blogs is provided by the attention Crawford (2013: 1) draws to the dangers of 'data fundamentalism', namely the fallacy 'that massive data sets and predictive analytics always reflect objective truth', a misconception Wang would undoubtedly endorse given her emphasis on the incontestable value of ethnography as a means to 'bridge and/or reveal knowledge gaps' (2013: 1) resulting from the surfeit of information supplied by Big Data. Thus, the ethnosemiotic paradigm posited here, which reintroduces the subjective human to the ostensibly objective numbers, and seeks for thick description and sub-surface meanings in the vastness of the data, attempts to overcome such 'fundamentalism' and inject new, deeper meanings into Big Data through its very concentration on a small set of first-person, French-language primary sources and in its call for ethnographic smallness, with the storytelling of individual lives serving as a guiding light to negotiate the largeness of the web dataset(s).

[A Merging of Models: Bourdieusian Ethnography Meets Kressian Semiotics](#)

In more precise terms, the ethnosemiotic framework supporting this case study is based on the concepts of Pierre Bourdieu (1972; 1979; 1980; et al. 1992; 1994), as regards the ethnography, and Gunther Kress (Kress & van Leeuwen, 1996; 2010; Domingo & Kress, 2013; Domingo, Jewitt & Kress (pending publication)), with respect to the multimodal social semiotics. In particular, Bourdieu's notion of habitus, that is, the dynamic interplay between externalised sociocultural practices / artefacts / representations and an internalised sense of belonging, or otherwise, on the part of the London-French bloggers, examined through a multimodal social semiotic prism. The conceptualisation of Bourdieu's habitus is, for the purpose of this study, threefold,¹ comprising firstly *habitat*, or the physical 'home', that is, the material environment, dispositions and physical spaces of the London French; secondly, *habit*, that is, the everyday practices and behaviours of the bloggers; and, thirdly, *habituation*, viz. tacit expressions of the taken-for-grantedness and embodiment of their situation in the migrant setting. It is therefore these three components of habitus that shall be sought in the data and subsequently examined multimodally in terms of their ethnographic meaningfulness, together with any shifts or developments detected by virtue of the 14-year span (1996-2010) of the JISC UK domain dataset.

1 Maton (in Grenfell, 2008: 55) proposes two components, namely habit and habitat, whilst Jenkins (1992: 179) introduces the notion of habituation, borrowed here as the third element of the habitus triad.

The analytical toolkit devised by multimodal social semioticians studying ‘texts’ in online environments (Domingo & Kress, 2013; Domingo et al. (in press)) lends itself well to the examination of blogs, and proposes six fundamental principles: the affordance of the platform (relating to the semiotic scope of the technology and genre); interest (that is, the underlying purpose of the text from the viewpoint of the sign-maker, but equally shedding light on the target audience and questions of identity); composition (including navigational features, layout, modularity, linearity, image-text weighting, etc.); mode and modal affordance (the meaning potential of distinct modes); design and the multimodal ensemble (including (inter-)modal choices, intra-textual cohesion and framing); design and the multimodal text (relating to coherence, extra- and inter-textuality, and the semiotic potential of the modes). To these multimodal analytical principles, useful for the analysis of a webpage as a whole, can be added Peirce’s (Chandler, 2007: 36–7) basic semiotic concepts, which are of particular relevance to the decoding and interpretation of *images*, and can be broken down into the following, mutually inclusive, sign categories: *iconic* (literally resembling that which it represents, e.g. a portrait photograph), *indexical* (signalling through a visually evident relationship that which is represented, e.g. a no-smoking sign depicted by a crossed-out iconic representation of a cigarette) and *symbolic* (bearing no logical or visible relationship to that which it represents, instead relying on accepted sociocultural convention to make sense of the arbitrary links between signifier and signified; the red cross on the cigarette falls into this category, as it is only through repeated and therefore accepted use, originally within Western cultures, that a diagonal – often red, arbitrarily signifying danger, perhaps aptly in this case – cross carries the meaning of incorrectness; conversely, if the same red cross is rotated to an upright position, the culturally acknowledged significance becomes medical, thus underlining the primarily culturo-cognitive sense of the symbolic sign and its intrinsic arbitrariness). Lastly, Kress (2010: 104), drawing on Michael Halliday’s (1978) communicational functions, delineates three semiotic metafunctions (ibid.: 87) which are pertinent in understanding the relationship between sign-maker, sign and sign-recipient: the *ideational* function (the relationship between the sign and its representational or performative effect in the world); the *interpersonal* function (the relationship between the sign-maker and sign-recipient – or between blogger and blog visitor) and the *textual* function which ‘corresponds to how the other two are presented within the text, i.e. to the shaping of the ideational and interpersonal meanings produced by the textual configuration of the interactive sites/signs’ (Adami, 2013: 8).

Having established the two-fold theoretical paradigm applied to the data collected for the purpose of this case study, it is now necessary to set out the rationale for the identification and extraction of the data from the web archives consulted.

[A la recherche de blogs perdus: Methodological Steps](#)

The primary methodological objective of this study was to identify sufficient web resources within the historic JISC dataset to construct a relatively minute, by Big Data standards, corpus of London-French blog instances archived in different years, which would facilitate the subsequent examination, comparison, revisitation and growth of the material. However, in order to build such a corpus, the challenge remained of how to locate the blogs in the vast expanse of data.

The Shine (Version 1.0) prototype interface presents three search routes: a basic search in the sample mode, a basic search in the full mode and an advanced search. Within these routes, **searches** can be tackled from two diametrically opposed conceptual standpoints, referred to here as 'bottom-up' and 'top-down' searches. In the bottom-up case, the search functions – advanced or basic – identify and propose web resources, whose presence in the archive is previously unknown to the researcher, on the basis of a search term or terms. The researcher approaches the data cold, akin to a treasure hunter, equipped with a metal detector, entering unexplored terrain, in the hope of unearthing a valuable hoard, but realistic in the knowledge that one or two gems, if any, are more likely finds – after much painstaking and time-consuming searching. The top-down search, however, requires a cognizant understanding of the terrain. Here, the dataset is approached with predetermined points of reference, for example, a specific URL or host, the researcher hoping to discover if it has been captured and is easily identifiable in the mass of data. In this model, it is a matter of finding a particular needle in a field full of haystacks, rather than entering a potentially empty terrain whose possible – but unlikely – treasures lie hidden deep beneath the surface.

To increase the chances of success in finding previously unknown blogs, due to their extinction on the live web, but possibly lying dormant in the dataset, and/or recognised blogs identified on the active internet and systematically recorded, a trialling of both search pathways was required: bottom-up and top-down. Having obtained almost no useful results in the AADDA search trials

applying a bottom-up approach, a concrete, top-down one was considered the more promising option initially. In the knowledge that the LFSC contained a series of London-French blogs, it was therefore consulted as a starting point in the top-down search, on the basis of which a list was compiled of the relevant URLs, the language (French, English or both) of the blogs and the inaugural post dates, as identified in the *internal archive* of the blog itself.

While this systematic process boded well in principle, in practice, the resultant spreadsheet revealed that the majority of blogs were created post-2009 and, as such, would not be found in the historic (1996–2010) JISC UK domain dataset. Furthermore, it became immediately apparent that most of the blogs were either from the <.com> or <.fr> domain, which again would automatically preclude them from the dataset under exploration, despite their being *London*-based in all but domain.

In order to overcome the ensuing and inevitable lacunae, which were indeed confirmed in subsequent top-down trials, two strategies were devised: 1) to compare the London-French blogs (in all but domain) from the LFSC, itself a selective, permissions-dependent, open-access 'micro archive' (Brügger, 2005: 10) inside the live, overarching UKWA, with earlier versions of the same blogs found in the US *Internet Archive* (IA), again taking a top-down approach using the IA's search tools; or 2) to take a bottom-up approach with the pre-2010 JISC dataset, as the absence of pre-2009 blogs from my collection was by no means proof of the non-existence of London French blogs prior to that time, rather, of the Google search engine failing to locate them on the live web, or of their having since become obsolete, a fact which would increase the historical value of the case study and in turn the research value of the dataset.

Both search options, at different times, using different web archives and different datasets, were therefore put to the test:

- Top-down in July–August 2014 for recent instances of the blogs via the public UKWA search pathway or, if unsuccessful, the LFSC sub-archive route; in addition to the Internet Archive for earlier (or the earliest) instances of the same blogs, both in conjunction with the internal archive of the blog itself. This was a fruitful exercise and provided sufficient data for analysis, but was frustrating in its neglect of the JISC UK Domain Dataset.

- Bottom-up in September–December 2014 via the prototype Shine (Version 1.0) interface of the JISC UK Domain Dataset (1996–2010), with its newly configured search and save functions, and increasingly complete state. This series of trials was far more productive than previous bottom-up AADDA attempts and culminated in the compilation of an initial corpus of over 30 resources. The Corpus-building potential of the dataset proved to be a time-efficient, analytically friendly and historically enlightening resource, even at the most superficial of levels, as the first resource added to the corpus illustrates.²

At this point, it should be added that both search approaches were significantly improved by the necessarily restrictive use of the French language in the search terms. The language alone served as a formidable filtering mechanism for UK-domain websites, hence the results often being far more manageable than those of their English-language equivalents.

Methodological Challenges and Considerations:

While both search approaches outlined above were fruitful, enabling the customisation of two (evolving) corpora within the JISC dataset and a useful inter-archival comparison, several pragmatic challenges remain when dealing with archived web material. The first of these is the visually/multimodally ‘deficient’ (Brügger, 2014: 20) archived version available to the researcher. Images are sporadically present and absent – sometimes in the same blog post –, layouts, even fonts can be misrepresented in relation to the live version, and videos are sometimes playable, sometimes not, seemingly arbitrarily. Granted, incomplete and/or damaged objects of analysis are a typical and accepted reality of the historian’s challenges in the physical world, but when conducting a multimodal analysis, these apparently *minor* deficiencies and inconsistencies present *major* challenges, ultimately jeopardising the very validity of the multimodal semiotic approach. For notions of design, interest, composition, modal affordance, intermodal relationships, metafunctions and the cohesion of the multimodal ensemble are all undermined, or at least difficult to assess with any accuracy, if the original semiotic affordances of the web resource are compromised.

² ‘London, A Year in the Shit – August 2005’.

An additional challenge is the difficulty in ascertaining the geographical positioning of the blogger, which although often deducible, cannot be inferred consistently, or with any authority, precision or certainty. For instance, a comment posted by a blog visitor, 'te voici de nouveau en Angleterre' ('so you're back in England') ('Ma Chronique Radio Paris ment, Radio Paris ment ...', archived in French Community Blogs Corpus in the Jisc Dataset (1996–2010)

<http://web.archive.org/web/20100714051717/http://nemesse1.blog.co.uk:80/2010/07/09/radio-paris-ment-radio-paris-ment-8943311/>, captured on 14 July 2010), confirms that the blogger is indeed based in the UK, but does not specify *London* as the ultimate destination or place of abode, which naturally casts doubt on the reliability of the blog as a primary source for a study of the French community in London. Furthermore, while the Shine (Version 1.0) interface provides researchers with a useful postcode identification tool, it is of little help in clarifying this form of geographical nebulosity given that the postcode provided is one associated with the *host* as opposed to the blogger or even his/her technological 'home'. That is, since blogs, unlike websites proper, are presented to the World Wide Web on external platforms, the postcode linked to the platform bears no relation to the physical location of the physical blogger. Furthermore, and reinforcing the above-mentioned challenge posed by version deficiency, this difficulty in reliably gleaned a sense of place is accentuated considerably when visual parameters are absent.

Another genre-specific and fundamental consideration when blogs are the object of analysis is whether there is a genuine need for them to be stored in a web archive at all, since blogs are intrinsically archival, equipped with their own integrated archive. This raises the question of what a macro archive can provide that a blog's integral archive cannot. With respect to live blogs, the advantages of the institutionally archived snapshots may *a priori* be deemed minimal for the researcher. However, in the case of obsolete blogs, a historic archived version is ordinarily the only resource available. Furthermore, and as the following inter-archival analysis will reveal, the integrated blog archive is, in fact, not as integral as it may first appear.

Also warranting reflection is the problematics of defining and cataloguing the blog as a genre. The basic generic boundaries between distinct online resources are far less clear-cut than in physical environments (Domingo *et al.*, 2014: 12) where texts are confined and defined by their very materiality. This haziness not only complicates the application of specific, well-defined metadata to ill-suited, generically ambiguous archived resources, but also serves as justification for including

captures of discussion board threads or reader comments in a corpus originally assigned to blogs. Clearly, these two examples cannot be referred to as blogs per se, because they do not conform to the structural, functional or technological characteristics of the blog genre. However, they do 'provide heterogeneous and original first-person testimonies' (Gomes & Costa, 2014: 110) of London-French experience, perhaps with increased 'sincerity' and validity, as they are often blunt expressions of genuine concerns or queries about life in London, devoid of the 'design' motivations (Kress, 2010: 133) and 'l'intérêt' ('the interest') (Bourdieu, 1994: 149) of the blogger who is producing the web object for an 'intended audience' (Yoon, 2013: 181); and they are too valuable a form of evidence to be relegated from the analytical corpus simply because they do not squarely map on to the generic label. While not constituting an insurmountable obstacle to the successful creation of corpora for this case study, given its manageable size, when scaling up this kind of categorisation exercise at an institutional level and attempting to restrict the unwieldy, organic assortment of web material housed in a macro archive to distinct classes and types of information, the difficulty and inappropriateness of the task (if attempted in the first place) would be compounded.

This leads on to the challenge of defining the analytical object itself. A blog roll, in form, is akin to a never-ending scroll (hence the phonetic proximity of the terms), which poses no challenge when navigating the material online, in its born-digital environment on the live web, or even in its reborn-digital context in the archive. It does, however, present difficulties when attempting to *analyse* and *re-present* the blog data within the body of a critical text. It raises questions over the boundaries that should be set to frame it for analytical scrutiny and the terminology to be used: for example, whether it is apposite to refer to the analytical object as a web 'page' when, in its nascent form, it is not self-contained on a single page or screen. It would be more accurate, though less succinct, instead to refer to 'screenshots of sections of web pages', or more precisely, of 'digital blog scrolls'. Reducing the object of analysis to a spatially and temporally suspended screenshot is analytically possible and easier to manage from a presentational – especially if a physical copy is produced – and audit-trail-validity perspective, but is unfaithful both to the born-digital live web object and reborn-digital archived web object, which aims, as Brügger points out, to maintain the "'Internet" dynamics ... and to some extent the recipient-specific dynamics' (2005: 33) of the original resource. In terms, therefore, of semiotic 'authenticity' (Lyman, 2002: 3) in relation to the initial blog design, as well as to its intrinsic visual and navigational affordances, the doubly-reborn-

digital object, that is, the once reborn archival version re-reborn within the constraints of a single screenshot for the sake of analytical convenience, offers the viewer a substantially altered, if not inferior, experience.

A supplementary methodological limitation, or at least risk, worth acknowledging is that while this paper attempts to illustrate evolution over time in the bloggers' habitus, through the changes noted in the various archived versions of their blogs, it is entirely plausible that the said bloggers might not be responsible for modifications, having perhaps recruited the services of an external web designer to refurbish their blogs. If this were the case, it would inevitably cast doubt over the validity of the findings as reliable indicators of habitus transformation. However, in view of the fact that the bloggers would have needed to approve such design changes, the legitimacy of the conclusions drawn is deemed sufficiently compelling.

A final consideration worthy of mention is the notion of curator subjectivity and the unforeseen objective power generated in bottom-up search results. While the apparent arbitrariness of the finds may sit uncomfortably with the rigorous researcher, ever striving for scientific credibility and an irrefutably robust methodology, it could be regarded a refreshing objectivising tool by the curator of a web corpus, seeking a means of making the final collection less a reflection of him/herself, as a result of the necessarily subjective selection choices made, and more about the material itself. In the bottom-up search model, however serendipitous the results may seem, the resources have the advantage of 'speaking for themselves', countering Kitchin's scepticism over the attractiveness of Big Data's 'empiricist epistemology' (2014: 4), and arguably giving the curator *and* researcher greater scientific objectivity.

Analysing the Blog Data

1) A Hybrid Habitus: *Tea Time in Wonderland & Lost and Found in London*

Taking a comparative inter-archival, multimodal analytical approach revealed several unexpected changes in the resources over time, as well as shedding light on more predictable developments in their technical affordances. Indeed, by examining the blogs multimodally, not only were multiple ethnographic conclusions able to be drawn regarding the increasingly hybrid habitus of the French Londoners, but it was also possible to ascertain which archives provided the most 'complete'

representation of the original web resource. *Tea Time in Wonderland* was one of the few .co.uk domain blogs identified and as such lent itself to a top-down search in the Jisc UK Domain Dataset. The contrast between the earliest version of the blog³ now permanently archived in the customised French Community Blogs Corpus (FCBC), and the most recent archived version,⁴ from the LFSC in the UKWA, is flagrant; and in both instances there appears to be a high degree of authenticity in relation to the born-digital resource.

While the soft colour palette remains analogous in both, there has been a significant change to the blog design. The banner has undergone the most substantial renewal, evolving from an essentially iconic 'London Tourist' representation to a 'French Londoner' one, which undoubtedly reflects the transformed habitus and increased sense of belonging of the blogger. The images in the updated banner improve intratextual and extratextual coherence between the blog name – 'Tea Time in Wonderland' – and the visual experience of its landing page. That is, there is a dynamic modal interplay between the banner words and images: the White Rabbit from Lewis Carroll's classic *Alice's Adventures in Wonderland* (1865) is at once an indexical sign, forming an intertextual relationship with the well-known (on both sides of the Channel) literary work, and a symbolic sign, evoking London as a *wonderland* in its own right, which also coheres with the sub-title of the banner 'La *magie* de Londres' ('the magic of London' – my italics). This bears witness to London as a place of mysterious allure, enchanting many a French young person through its candid eccentricity and celebration of otherness, as repeatedly confirmed in the interviews conducted for the overarching study (for example, by Séverine, in Interview 20, 'à Londres, l'excentricité est encore admise et respectée' ('in London, eccentricity is still allowed and respected')) and other blogs, such as *Lost and Found in London*, where the city is referred to as 'cette ville excentrique, dynamique et si attachante' ('this eccentric, dynamic town that is so lovable'). Indeed, the recent design implies a development in the mindset, or more aptly the habitus, of the blogger, Coralie, as if she has embodied an air of Londoners' eccentricity herself, becoming unwittingly habituated to it and explicitly embracing it through the motivated signs of the banner. Although the reference in the blog name to 'teatime' reinforces the intertextual link to Carroll's work – particularly, and befittingly, to the Mad Hatter's tea party – and in so doing adds to the coherence of the blog landing page as a multimodal text, 'teatime' speaks more loudly of its clichéd Franco-French

³ <http://web.archive.org/web/20090628092128/http://teatimeinwonderland.co.uk/>

⁴ <http://www.webarchive.org.uk/wayback/archive/20140108230609/http://teatimeinwonderland.co.uk/>

association with 'Les British'. This reliance on stereotypical evocations of London(ers) can perhaps be explained by the choice of name predating the redesign of the banner, probably corresponding to Coralie's recent migration to the city, when still very much embedded in the habitus of origin.

Nevertheless, a comparison between the original banner contained in the Jisc dataset and the updated one housed in the LFSC brings to light a clear shift from the stereotypical to the typical. London's physical habitat is suggested indexically – in contrast to the iconic image dominating the earlier version of the banner – through the 'double-decker' bus, but as a whole the current banner is a tacit articulation of the nonconformity that could be said to epitomise London in the eyes of many French Londoners, finding it a liberating and appealing quality comparatively rarely encountered in France, as indicated in the interviews conducted for the wider study: 'Ici on a plus de liberté. Chacun a le droit d'avoir son style; en France on est tout de suite jugé' (Interview 17, Chantal, ('You have more freedom here. Everyone has the right to have their own style; in France you are judged straight away')). The blogger's self-appropriation of the idiosyncrasy that is permitted in London, and her cultural sensitivity more generally, are made manifest again in the subtle intertextual referencing of the hearts in the background of the banner, the indexical allusion being to the Queen of Hearts from the children's story; while the playful inversion of the hearts on the right-hand-side of the banner, in apposite topsy-turviness, is equally suggestive of playing cards, and perhaps a further multimodally cohesive sign of London's 'magie'. Likewise, on a symbolic level, the hearts function as a subtle depiction of the blogger's fondness for the city, echoing the words of one interviewee, when asked to send a photograph of 'her' London: 'J'aurais envie de vous envoyer une photo d'un coeur: c'est là qu'il y a ma famille, c'est là qu'il y a mon amour, c'est là qu'il y a mes intérêts et mes passions' (Interview 2, Jacqueline, ('I would like to send you a photo of a heart: London's where my family is, where my love is, where my interests and passions are')).

Despite the clear transformative process that has taken place between the pre-2010 blog instances collected in the Jisc UK Domain Dataset and the 2014–15 ones housed in the LFSC, the blog, as a multimodal semiotic ensemble, is not entirely 'anglicised', being instead a London-French hybrid, which doubtless reflects the current identity of the sign-making blogger. This is exemplified through the attempt made to depict the quintessentially British habit of physically having tea through the pot and the chair – but, significantly, the former, in its shape, is more reminiscent of a

French coffee pot than a British *teapot*, just as the latter is not a representation of a typically English wheelback or Chesterfield, but a traditional Louis XV-style armchair. It could be argued, therefore, that this blog subtly expresses something of halfway habitus, straddling the cultural habitat and habits of both France and London, yet demonstrating internalised habituation vis-à-vis the quirkiness of the city and its inhabitants.

In the 2014 instance of the *Lost and Found in London* blog⁵ the banner has also undergone significant alteration since the 2010 capture,⁶ this time identified through a top-down Internet Archive search, given that it is hosted by the .com domain. It has, like the *Tea Time in Wonderland* banner, gone from being a fundamentally iconic photographic representation of famous London landmarks to an intricate spatial dialogue which merges indexical signs for physical London places in a bird's-eye-view map of the city with simultaneously iconic, symbolic and, to a certain extent, indexical, through their social implications, superimposed childlike images. The crudely sketched green forms in the foreground signify, in primarily iconic terms, trees. On an indexical level, and in the context of this multimodal ensemble, they relate more broadly to London's green spaces, back gardens and acres of parkland; whereas, symbolically, they suggest the French appreciation of such verdure, particularly in comparison to the relative lack thereof in Paris, as frequently recounted in interviews: 'Londres, c'est les squares avec des petits jardins, des parcs' (Interview 7, Marie ('London, it's squares with little gardens and parks')) or 'Je n'habite pas dans une banlieue en béton [comme à Paris]; j'habite dans une banlieue où il y a plein de petites maisons à deux étages, avec un parc, avec des arbres' (Interview 3, Sarah, ('I don't live in a concrete suburb [like in Paris]; I live in a suburb where there are lots of little two-storey houses, with a park, with trees')); Mulholland and Ryan (2013) also allude to this veneration: 'London offers a generally higher quality of life than Paris, with more green space'.

While semiotically complex compared to the earlier banner, this multimodal ensemble works as a coherent and cohesive whole. As regards its compositional principles, the directionality of the images, from left to right, are geographically, and therefore indexically, coherent in relation to their physical spatial positioning from east to west, and each icon portrayed in the sequence is also cohesively meaningful, reinforcing the social-semiotic affordance of the banner as a multimodal

⁵ (<http://web.archive.org/web/20141224060430/http://lostandfoundinlondon.wordpress.com/>)

⁶ <http://web.archive.org/web/20100121160120/http://lostandfoundinlondon.wordpress.com/>

unit. At first glance, the suburban terraced houses to the left of this 'web element' (Brügger, 2014: 5) appear to correspond to the entrenched conception of London streets among the French, where 'les maisons se ressemblent toutes' (field-notes, 5 July 2014; 'all the houses look alike') and possibly to where the blogger lives, West London being one of the residential community hotspots. Under closer scrutiny, however, the naively sketched houses show differences in their architectural features which could, if interpreted in relation to their left/west positioning, indexically allude to the Tudor edifices of the Hampton Court district, notably the home of Hampton Court House School (an independent Anglo-French bilingual school accredited by several French bodies and popular among the more affluent ranks of the French community, hence the symbolic potential of the image); likewise, moving right/east and more central (geographically and compositionally), the iconic quality of the representation of the house differs from the others, the simplicity of its symmetrical, rectangular lines perhaps acting as an indexical signifier of a Georgian terraced house, quite possibly one found in Kensington, given the geographical coherence of the banner's composition, and therefore carrying symbolic weight given that Kensington is considered by many to be the established 'home' of a certain – affluent – French community in London. These architectural details demonstrate an in-depth knowledge of the West London habitat that appears to form part of the material habitus of this blogger (a former lawyer, whose blog makes repeated visual and written allusions to her visits to Richmond Park, Ealing, etc.) and is indeed inhabited by many members of the socio-economically comfortable sub-community of the London-French population as a whole; as one interviewee put it, 'Les Français "posh", ils sont à South Kensington; les Français "hip" sont plus vers Brick Lane' (Interview 19, Miranda, ('The posh French are in South Kensington; the hip French are more around Brick Lane')).

While the photograph in the 2011 version of the banner is equally evocative of London, its negation of the presence of the West London habitat of the blogger and of London's financial centre, the City – depicted indexically in the 2014 instance through St Paul's Cathedral, the Gherkin and Tower Bridge – is significant, for not only does it deny the blog visitor/sign-recipient the insight into the habitus of the blogger that was provided in the later instance through its habituated referencing to her habitat and, by extension, her likely habits, its exclusive focus on central London's West End (as was the case with *Tea Time in Wonderland*) places the blogger in the position of tourist as opposed to inhabitant, suggesting greater attachment to and positioning in the habitus of origin than that of adoption. Whereas the inclusion of iconic edifices whose

collective presence forms an indexical link to the City in the updated version of the banner implicitly sheds light on the working habitat of many highly-skilled French migrants (Mullholland & Ryan, 2013; Favell, 2006, 2008), and as such indicates a growing embeddedness in the habitat and working practices of the migration setting. Indeed, evidence from an article titled 'The French Phenomenon' archived in the French Community Blogs Corpus, within the Jisc UK Domain Dataset (1996–2010), on 11 Nov. 2014 (http://web.archive.org/web/20070610042613/http://news.efinancialcareers.co.uk:80/NEWS_ITEM/newsItemId-10155, site captured 10 June 2007), triangulates this semiotic reading: 'BNP Paribas alone employs more than 400 French people in its Harewood Avenue office ... French talent is underpinning London's leading position in the OTC derivatives markets ... Another recruiter alleges that 80% of quants at Goldman are French' (accessed 10 Feb. 2015). The representation of the City, therefore, in the updated web element, is an accurate reflection of the professional habitat and habits of a segment of the highly-skilled London-French population.

Tate Modern is another introduction to the recent version of the banner. Placed in a coherent position relative to the map and the semi-iconic representation of waves depicted through a 'hand-drafted' set of blue, unfurling conical forms, indexically referencing the River Thames running along the bottom right-hand-side of the banner, the minimalist depiction of the Tate gallery serves to echo the habits of the London French and their appreciation for free entry to London's cultural institutions, a point articulated repeatedly in other blogs and verbal accounts, such as: 'Les musées, c'est gratuit, donc j'adore' (Interview 9, Brice, ('Museums are free, so it's great')) or 'There's a culture of free things which is strange in a way' (Interview 6, Antoine); the interviewee in question was a Tate member and cited the British Museum and the Greenwich Observatory as other prized cultural attractions.

The various examples above, by means of comparative, multi-archival, multimodal analysis, have enabled the identification of an evolving online habitus in the blogs themselves and, by extension, in the 'on-land' habitus of the bloggers. The archived versions reveal the move from the literal and the iconic outsider's representation of the city, to a more personal and symbolic insider's representation of *their* city, which, in turn, substantiates further the theory of habitus transformation among the London French, and an increasing sense of London being home.

2) Interpersonal Meaning in Fine-grained Reading: *Londres Calling* & *Good Morning London*

Among all the blogs housed in the LFSC, only four were from the UK domain, one of which was *Tea Time in Wonderland*, analysed above, the others being from the .com domain, or, to a considerably lesser extent, the .fr domain (telling associations in themselves, in terms of a lack of allegiance to the country of origin). Furthermore, the remaining three were all created post-2010, hence their absence from the Jisc UK Domain Dataset (1996–2010). For this reason, the following blog analyses are consistently based on inter-archival comparison between captures identified in the UK Web Archive, its sub-collection, the LFSC, the US Internet Archive (IA) and/or the integrated blog archives.

As has been evidenced above with respect to the hybridisation of London-French habitus, far from keeping an authentic record of past posts, the in-built blog archive is open to manipulation according to the changing identity, positioning and interpersonal intentions of the blogger. This has an effect on the blog design, its semiotic affordances and its textual function. The shift in cultural reference points that accompanies the refurbished blogs, and in turn narrows the accessibility of the multimodal text among blog visitors unfamiliar with the London-French hybrid habitus, is exemplified compellingly through the 2010⁷ and 2014⁸ versions of the *Londres Calling* blog, supported by the canalblog.com platform, and is indicative of a new audience being targeted.

Once again, the most noticeable change relates to the banner. London is no longer depicted indexically through the iconic representation of a red telephone box alone, but by far more subtle, ‘hand-drawn images’ (as is often the case in London-French blogs, recalling France’s long tradition of graphic novels), which assume greater knowledge of and integration into London/British culture. In the 2014 version, there is an iconic depiction of a Dalek, the indexical significance of which would no doubt be lost on a Franco-French audience, generally unacquainted with the half-century-old science-fiction BBC television series that has become an integral feature of the British audiovisual landscape, and would suggest a change in the television-watching habits of the blogger. The inclusion of this make-believe ‘character’ from a famously quirky series, whose protagonist plays on the eccentricity of the ‘mad scientist’ figure, resonates with the White Rabbit depicted in the *Tea Time in Wonderland* blog as regards its function within the multimodal text,

⁷ <https://web.archive.org/web/20100325194043/http://londrescalling.canalblog.com/>

⁸ <https://web.archive.org/web/20140615152847/http://londrescalling.canalblog.com/>

but differs in terms of its interpersonal function, as its culturo-symbolic insignificance in France would make it meaningless to a Franco-French audience, unlike the widely-known Carroll character. The narrowing of the audience in the 2014 version is emphasised by the other artefacts extracted from the blogger's on-land habitat, which Rowsell describes as 'fractal habitus' (2011: 333), for inclusion in the online banner, such as several quintessentially English ice-creams (99 Flake, Twister, etc.), which, ideationally, are meaningful for a British audience but not a Franco-French one, giving a sense of deep, even childlike, and therefore well-rooted, belonging and integration. An indexical sign pointing to an art gallery is another addition, echoing the inclusion of Tate Modern in the updated *Lost and Found in London* blog and, in so doing, suggesting a degree of online community cohesion through the shared appreciation of London's cultural institutions experienced on-land. Other culturally meaningful, as well as culturally exclusive, features of the banner indicative of the blogger's evolving tastes are the Victoria sponge cake and the jar of Marmite, which, owing to the seeping of advertising into the mores of British society, acts as a symbolic signifier of a love-hate relationship, applicable perhaps to London culture as a whole given the blogger's decision to include it in the *Londres Calling* banner, and presumably expressing a love of both condiment and culture. Reflecting another dimension of the culture of the migration setting is the England footballer, which may well indicate a change in the habits of the blogger, with several interviewees in the wider study commenting on their support of a particular London team and the sense of belonging it gave them to the local community. Taking central position, however, is a sign that necessitates no in-depth knowledge of London culture to be meaningful, yet is doubly-symbolic: a Union-Jack-clad heart, the heart being a conventional symbol for love, just as the red, white and blue flag is for the UK. The standardised coding of the message conveyed through the combination of both is culturally exportable in this case and ensures that the fundamental purpose of the blog is immediately accessible to London-French, Franco-French and Francophile audiences.

All these images sequentially combine together to form the multimodal ensemble that is the updated *Londres Calling* banner, a web element that shows a considerable shift in the habitus of the blogger and the interpersonal function of the multimodal text, evolving from a somewhat simplistic,⁹ stereotypical and predominantly semiotically indexical representation of London

⁹ <https://web.archive.org/web/20100325194043/http://londrescalling.canalblog.com/>

through an iconic depiction of a telephone box (which was successful in acting as a cohesive device in relation to the blog name, but signifying little else), to a semiotically manifold and culturally sophisticated set of symbolic and indexical signs, tacitly and exclusively directed at a (French) audience based in London, or at least sensitive to the subtle materialities constitutive of British culture. However, the Britishness of the objects acts in semiotic duality with the stylistic Frenchness through which they are depicted. The pastel colour palette, their hand-drafted quality and the equal hierarchical standing awarded to the culinary as to objects of ‘high’ culture are all typically French features. This semiotic hybridisation could be interpreted, therefore, as indicative of a distinct London-French habitus and, by extension, a specifically London-French audience and interpersonal function.

Similarly, the *Good Morning London* blog, hosted by the .fr domain, bears witness to metafunctional transformation. There are repeated indications in the blog posts and related comments that the blogger, Aurélie, reaches a Franco-French audience in addition to the London-French community closer to ‘home’. However, under closer examination of the earliest (via the IA) and most recent captures of the blog (via the LFSC in the UKWA), subtle changes to the content of the web objects suggest a shift away from the habitus and audience of origin. When placing a screenshot of the first instance¹⁰ of the inaugural blogpost captured in 2011 and identified top-down (by inputting the live URL to the IA ‘Browse History’ field) alongside the first instance of the ‘same’ blogpost found in the (micro) internal blog archive (via the (meso) LFSC of the (macro) UKWA), captured three years later in 2014,¹¹ almost unnoticeable, yet ethnosemiotically meaningful, modifications can be detected. Irrespective of the font change in the title of the post, in the 2014 instance, a blending of French and English typographical norms is evident: the space before the colon is typical of French standards, and found in both instances, whereas the reliance on the exclamation mark – used far more liberally in French texts than their English equivalents – has been reduced, as its absence from the updated article title confirms. Likewise, the incongruous placing, by British standards, of the sterling symbol (between the pounds and the pence) in the 2011 instance, has been ‘corrected’ in the 2014 version, thus corresponding to English cultural norms (and to the expectations of readers who have been exposed to them for sufficient time).

¹⁰ <https://web.archive.org/web/20111007110655/http://www.goodmorninglondon.fr/>

¹¹ <http://www.webarchive.org.uk/wayback/archive/20140720220816/http://www.goodmorninglondon.fr/bon-plan-cine-40-de-reduction-dans-les-salles-odeon-jusquau-13-octobre/>

This suggests that the blogger herself has become more accustomed to British typographical practices over time and could be making a greater effort to 'assimilate' or self-appropriate local customs. In this example, therefore, by comparing the institutionally archived version of the blog to its in-built 'archive', punctuation has been seen to act as a *mode*, that is, a form of expressing meaning, both wittingly and implicitly, and provides an indication of the evolving habits and audience of the blogger.

The Welcome module on the right of the screen also attests to a more targeted London-French audience through typographical amendment and lexical economy. The punctuation is again illustrative: the 2011 exclamation mark after 'Welcome !' has this time been replaced by a 'smiley face', ' :) ', reiterating the blogger's diminished reliance on the typically French exclamation mark, choosing instead to use an intrinsically more 'light-hearted' form of engaging the reader through the mode of punctuation, which could be said to reflect the English sense of humour to which many interview respondents alluded positively: 'J'aime bien l'humour anglais, quand je le saisis' (Interview 11, Brigitte, ('I really like the English sense of humour, when I get it')), or 'Je pense que les Anglais ont un sens de l'autodérision qui n'est pas du tout quelque chose qu'ont les Français' (Interview 5, Charles, ('I think the English have a sense of self-derision which is not at all something the French have')). Moreover, the 2014 welcoming paragraph, or web element, is a considerably abbreviated iteration of its former self, again more in keeping with English rhetorical norms than French. Gone is the redundant use of synonym and adjectival description whose sole purpose is to add stylistic weight to the sentence, 'Ce blog est dédié au partage de bons plans divers et variés pour profiter au mieux de la superbe capitale british' ('This blog is dedicated to the sharing of diverse and varied top tips to make the most of the superb British capital') is condensed to 'Ce blog a pour objectif premier le partage de bons plans sur Londres' ('The main aim of this blog is the sharing of top tips about London'), followed by several concrete examples. This discursive shift from the ornate and superlative to the streamlined and concrete is in keeping with the types of changes made when translating and/or adapting information from French into English, and as such reflects both habitus transformation – in particular regarding writing *habits* – on the part of the blogger, as well as metafunctional alteration. That is to say, although the text has not been translated into the official language of the migration setting, the stylistic adaptation towards English writing conventions is indicative of an interpersonal and ideational function now geared towards a London-French audience to a greater extent than in the previous version of the blog

archived in the IA.

Gone also is the exaggerated Anglo-French linguistic hybridisation of the text, brought about by the liberal code switching between French and English. While this might on the surface indicate *less* blending into the ‘host’ culture over time, a more convincing argument is, on the contrary, that the scattering of English lexical items found in the earlier, 2011 version of the Welcome text is a self-conscious act to portray a fashionable and ‘connected’ blogger, who is *au fait* with the latest English terms being used *in France*, which reinforces the hypothesis of a Franco-French audience being prioritised initially. Therein lies the paradox, for the – almost involuntary – code-switching that occurs among long-standing French Londoners for the sake of efficiency or non-equivalence, when ‘le mot anglais vient d’abord, en fait, et finalement l’équivalent ne vient pas’ (Interview 8, Sadia, (‘the English word comes first, actually, and the equivalent doesn’t end up coming at all’)), is not the same as the code-switching practised by French “stayers” (Favell, 2008: ix), either in terms of the lexical items substituted or the contexts in which the transposition takes place, with the latter making – entirely voluntary – use of English terms for effect. This can be seen in the earlier capture of the web element, where ‘british’ and ‘enjoy’ appear within the French sentences – both of which have easily accessible alternatives in French – and imply a self-conscious effort to evoke ‘cool Britannia’ (or the American Dream), whereas in the more recent version, ‘shopping’ is the only English word used in this block of text, relied upon here principally for efficiency on the part of the sign-producer rather than the connoted, symbolic effect it may have on the sign-recipient (regardless of the term’s entry into standardised Franco-French vocabulary), as its French translation would be one of several *three-word* phrases. Thus the affected linguistic ‘style’ found in the 2011 instance has been replaced by sparser, British substance in the integrated – and updated – blog ‘archive’, illustrating an interpersonal refocusing from a Franco-French to a London-French audience, as well as a sense of the blogger being more culturally aware and socio-linguistically ‘at home’.

3) Technological Scope and Limitations: *Good Morning London - About & Home Sweet London*

Developments in the technical and navigational affordances of the *Good Morning London* blog as a multimodal ensemble are also noteworthy. In the 2011 version of the About page,¹² the

¹² <https://web.archive.org/web/20111007112700/http://www.goodmorninglondon.fr/a-propos>

intratextual navigational options are limited to three: 'Accueil' ('Welcome'); A Propos ('About'); and 'Contact'. By 2014,¹³ these options have more than doubled, with eight tabs leading to other spaces deeper within the blog and, accordingly, the habitus of the blogger, such as 'Shopping', 'Food', 'Sorties' ('Going Out'), 'En dehors de Londres' ('Outside London') and a 'Blogroll'. The greater scope of the 2014 version bears witness both to the increased technical possibilities available to Aurélie within the framework of the platform and to transformation in all three of the habitus dimensions posited here. Regarding *habits*, the tabs would suggest that she is now in a position to share her London French shopping and eating practices; in terms of *habitat*, the material recommendations and excursions evidence a greater sense of ownership of, and belonging to, London and its environs as a physical space, within which conclusions can be drawn about the *genuine* on-land geographical positioning and movements of the diasporic population, as opposed to the mythologised spaces, most notably South Kensington, often considered to be the exclusive realm of the French community. The habituation dimension is also represented through the tabular titling alone: the 'naturalised' London French linguistic hybridisation, which assumes a high degree of bi-cultural knowledge, juxtaposing Franco-French anglicisms, such as 'Shopping', with a decidedly London-French use of the word 'Food' (phonetically shorter and semantically broader than its French-language counterpart 'nourriture', and, unlike 'le shopping', not a standardised feature of the French language). Moreover, compositionally, the order (from left to right, as Western convention dictates) is significant here, since the English 'Shopping' and 'Food' are placed in an arguably subordinate position to 'Bons plans en vrac' ('Loads of top tips'), itself a highly colloquial and idiomatic Franco-French utterance which would exclude most non-native speakers of French through its semantic inaccessibility, and is therefore indicative of both a continued attachment to the culture and habitus of origin, and of a sense of the defined London-French community referred to above. Developments in the technological affordances of the blog consequently impact on its textual function, with the 2014 version not only targeting a London-French, rather than the initial more Franco-French, audience, resulting from the increase in navigational scope, but concurrently transforming the text ideationally, bearing in mind that the material *habit* and *habitat* recommendations made are designed to have an effect on blog visitors' practices in the physical world, in a dynamic online/on-land relationship.

¹³ <http://www.webarchive.org.uk/wayback/archive/20140720220350/http://www.goodmorninglondon.fr/a-propos/>

Further alterations to the multimodal affordances of the recent capture¹⁴ are noticeable in the more explicitly modular layout of the web object, in which the written information is neatly framed by means of visible lines, and which is, in turn, one step further away from a traditional, physical book or diary page (see Domingo *et al.* (pending print): 7). The most immediately patent difference is the semiotically iconic image, placed in 'prime' reading position on the left of the text, which now gives the blogger a physical identity denied in the earlier version, either through the former technological limitations of the system or those of the blogger, or perhaps because of her increasingly settled identity within the migrant setting. The extratextual (hyper)links to social media are also noteworthy in their greater number (a manifestation of their growing interpersonal and ideational roles in 'on-land' and online environments), their heading, which has moved from the French 'Suivez-moi' ('Follow me') to a more inclusive, less dogmatic, 'Keep in Touch', and finally their unfaithfulness to the born-digital object, since one of the icons is missing in this recently reborn-digital version. The implications of this 'deficiency' (Spaniol *et al.*, 2009; Brügger, 2014), while not overly detrimental here, are nevertheless of import when conducting a multimodal analysis, where every detail of the multimodal ensemble has the potential to carry meaning.

The impact of such visual deficiency is demonstrated unequivocally in the first post on the *Home Sweet London* blog. In the IA instance,¹⁵ captured in May 2011, the modally and compositionally dominant set of five photographs of the Carter Steam Fair has been preserved, providing the multimodal sign-recipient with a tangible sense of the physical reality of the funfair, tinged with an air of nostalgia through the blogger's decision to present the images in black and white. Indeed, this choice appears to be a deliberate attempt to accentuate the antiquity of the funfair (John Carter was, according to the blogpost, a collector of early fairground trappings). Here, therefore, the monochrome hues function as a mode in tandem with the written mode, in which the semantic field is predominantly 'old', with the following lexical items included in the brief, seven-line text: 'autrefois', 'traditionnelle', 'd'époque', 'd'antan', 'datant du', 'début du XXème siècle' and 'restauré' ('yesteryear'; 'traditional'; 'vintage'; 'of old'; 'dating from'; 'early 20th century'; 'restored'). The intermodal relationship between the historic lexis of the text and the black-and-white images serves to communicate more powerfully both the quaintness of the bygone fairground and, perhaps more importantly, the nostalgia of the blogger (in London for some eight

¹⁴ Ibid.

¹⁵ <https://web.archive.org/web/20110526200110/http://homesweetlondon.wordpress.com/>

years at the time of writing the Carter Steam Fair post) for now remote – geographically and temporally – childhood experiences, often felt with greater intensity among migrants, in their dislocated state from the homeland and home-family ('family' being the prevailing response to an interview question on the most-missed aspect of France, as Brice, Interview 9, reflects: 'Qu'est-ce qui me manque le plus ? Pas grand-chose ... Amis, famille... mais dans l'ensemble, ça va' ('What do I miss most? Not a lot ... Friends, family ... but overall, it's fine')). The absence, therefore, of the photographic ensemble in the 'same' post preserved in the UK Web Archive in 2014, is significant.¹⁶ All that remains of this telling, intermodally meaningful, set of images in the UKWA version is a tantalising acknowledgement of its original born-digital presence through an automated missing image icon. A geographical sense of place is specified through the reference to the fairground being set up 'dans le parc à côté de chez moi à Ravenscourt Park' ('in the park near where I live in Ravenscourt Park'). However, this useful clarification as regards the (non-South-Kensington) external habitat of the blogger lacks the underlying meanings relating to her internalised habitus conveyed through the images, hence the importance of a macro archiving system that can preserve the micro modal detail of born-digital web resources.

Taking Stock

The bottom-up search route proved wholly worthwhile in allowing the identification and thematic preservation of previously unknown – and now obsolete – material. Although relatively undeveloped in this case study, it brought to light evidence not found elsewhere in the London French blogs, for instance, the unusually critical content of the aptly named blog, *A Year in the Shit* (in an intertextual nod to Stephen Clarke's *A Year in the Merde*) (<http://web.archive.org/web/20060112224755/http://london-adventure.blog.co.uk:80/main/index.php/london-adventure/2005/08>, French Community Blogs corpus in the Jisc UK Domain Dataset (1996–2010), captured 12 Jan. 2006), where the detritus observed in the London habitat is photographed and denigrated, alongside the 'inappropriateness' of several of its native inhabitants' habits, such as wearing flip-flops to the workplace or routinely consuming excessive quantities of alcohol (practices repeatedly disparaged in interview and field-note evidence). This historic blog data is of added value in that it contrasts fundamentally the rosy picture painted of London in every other London-French blog encountered, instead echoing the

¹⁶ <http://www.webarchive.org.uk/wayback/archive/20140109004953/https://homesweetlondon.wordpress.com/page/18/>

sentiments of those whom one migrant referred to as the French 'cliques' (field-notes, 2009), who appear to take pleasure in collectively bemoaning the shortcomings of the city (and whom she, like many others, attempted to avoid, as François explained in Interview 10, 'aller à l'étranger pour se replier sur son groupe ethnique; c'est pas mon truc' ('going abroad to withdraw into one's own ethnic group; that's not my thing')). The bottom-up searches also enabled access to, and preservation of, material unharvestable in the LFSC due to permission not being granted by the host owner, as opposed to the blogger herself, namely posts written by historian and social commentator (and French Londoner in her own right) Agnès Poirier, for *The Guardian's* 'Comment is free' blog. An article by Poirier, titled 'Pauvres Britanniques!'

(http://web.archive.org/web/20080222015108/http://commentisfree.guardian.co.uk/agnes_poirier/2007/08/pauvres_britanniques_1.html, French Community Blogs corpus in the Jisc UK Domain Dataset (1996–2010), captured 22 Jan. 2008), was discovered serendipitously, during a bottom-up search using the Shine (Version 1.0) interface and is now archived in the London French blog corpus within the overarching, authorisation-free, historic Jisc dataset, offering scope for analysis at a later date. Finally, the bottom-up search provided alternative forms of empirical evidence, such as online fora and comment threads, useful for the purpose of triangulation in the broader London French project, for example, comments on the [sofeminine.co.uk](http://www.sofeminine.co.uk) 'Having a Baby and Parenting' discussion board relating to women's unexpectedly positive experience of the National Health Service: 'il faut faire confiance au systeme [*sic*] anglais. Les choses ne se font pas comme en France mais finalement je me sens moins stressé [*sic*], tout ce [*sic*] passe pas trop mal' ('you have to trust the English system. Things don't happen like in France but I actually feel less stressed here, and everything goes pretty well')

(http://web.archive.org/web/20060329010819/http://forum.sofeminine.co.uk:80/forum/maternite1/f8_maternite1-Any-french-future-mother-around.html, French Community Blogs corpus in the Jisc UK Domain Dataset (1996–2010), captured 29 March 2006). This counters the deep-seated preconceptions present in the habitus of origin that the NHS is an outdated, inefficient and inferior system to its French counterpart, and again demonstrates the evolving opinions and dispositions of the London French, in addition to an emerging sense of community through the empathetic comments exchanged among the small group of French women simultaneously experiencing pregnancy for the first time in the diasporic habitat.

When practicable, the top-down search pathway demonstrated the historic inauthenticity of in-

built blog archives, that is, the non-equivalence or incompleteness of captures of the 'same' material, together with the impact of technological advances on blog design, digital habitus and, in turn, self-representation. However, owing to the domains and dates of all but one of the London French blogs identified for this case study, the top-down approach was almost impossible within the Jisc dataset. As a result, a multi-archival, multimodal, cross-temporal analysis of habitus was conducted, which succeeded in confirming the transformation from Franco-French outsider to London-French insider, in the space of a relatively few number of years, and the progressive targeting of a like-minded audience, capable of tapping into the same unsaid cultural references as themselves, and therefore connoting the reality of a particular London-French community identity.

More time is needed to study the commonalities, as proposed at the outset of this project, between the blogs analysed in this paper and the more recent additions to the ever-growing historic corpus constructed in the Jisc UK Domain Dataset, which would serve to test further the existence of an online/on-land 'community' per se. At present, the most striking inter-blog patterns emerging are that, without exception, they appear to belong to women, which may or may not relate to the demographic distribution of the community on-land; evidence also suggests that the bloggers correspond to a more affluent sub-category of the London-French community at large, which could be related to their having more time at their disposal, for instance, than French Londoners working as employees in the hospitality sector, or could be explained by the relatively large proportion of bloggers freelancing in the media industry who use blogs as a self-promotional tool; there is a predilection for 'hand-drafted' imagery in a number of London-French blogs, reflecting the status and embeddedness of the 'bande dessinée' (comic book) in French culture and a maintained attachment to the habitus of origin; finally, food recurs as a leitmotif running through the blogs, which again speaks typically of the centrality of the culinary in French culture, at home, but is equally illustrative of the cultural dynamics taking place within the migration setting, for the majority of the blogs are a celebration of *English* rather than French comestibles, confirming therein the habituated adoption of London habits and the materialisation of a hybrid habitus.

To conclude, 'thinking small', by means of a fine-grained analysis of a modest corpus of resources, has indeed bridged some of the 'knowledge gaps' (Wang, 2013: 1) between Big Data and sociocultural meaning. The inter-archival examination of the blogs was effective in revealing,

through a text-based alternative to numeric 'Data fundamentalism' (Crawford, 2013: 1), the interplay between internalised and externalised lived experience, with the London French observed to become increasingly at home in London, yet disinclined to relinquish entirely the habitus of origin. In this distinctly hybrid habitus common among the bloggers, they were not torn between two cultures, but over time embodied both, through a mutual merging of French and British habits, habitats and attitudes, giving rise to a sense of community and personal growth. As Catherine (Interview 13) epitomises, 'Je me sens pas tiraillée entre les deux pays, maintenant j'ai intégré les deux, j'ai trouvé un bon équilibre' ('I don't feel torn between the two countries, I have integrated both now, I've found a good balance').

References

- Adami, E. (2013) 'A social semiotic multimodal analysis framework for website interactivity' – MODE Working Paper, London: National Centre for Research Methods [available at http://eprints.ncrm.ac.uk/3074/4/website_interactivity_Adami.pdf accessed 10/02/2015]
- Bourdieu, P. (1972) *Esquisse d'une théorie de la pratique*. Geneva: Librairie Droz, 1972; Paris: Seuil, 2000.
- Bourdieu, P. (1979) *La Distinction*. Paris: Minuit
- Bourdieu, P. (1980) *Le Sens pratique*. Paris: Minuit
- Bourdieu, P. & Wacquant, L. (1992) *Réponses : pour une anthropologie réflexive*. Paris : Seuil
- Bourdieu, P. (1994) *Raisons Pratiques: Sur la théorie de l'action*. Paris: Seuil
- Brügger, N. (2014) 'The Web of Hyperlinks – Challenges for a Historical Approach' (slides presented at BUDDAH Research Meeting held at IHR 17 Sept. 2014)
- Brügger, N. (2005) *Archiving Websites. General Considerations and Strategies*. Aarhus: Centre for Internet Research.
- Carroll, L. (1865) *Alice's Adventures in Wonderland*. [via Internet Archive, as a PDFFreeBook <https://archive.org/stream/AlicesAdventuresInWonderland/alice-in-wonderland#page/n0/mode/2up> , accessed 10 Feb. 2015]
- Casilli, A. A. (2010) *Liaisons numériques – Vers une nouvelle sociabilité*. Paris: Seuil
- Chandler, D. (2002, 2nd Ed. 2007) *Semiotics: The Basics*. London: Routledge
- Clarke, S. (2005) *A Year in the Merde*. London: Black Swan

- Crawford, K. (2013) 'The Hidden Biases in Big Data' in The Harvard Business Review blogspot (01/04/2013). Available at <http://blogs.hbr.org/2013/04/the-hidden-biases-in-big-data/> [accessed 11 Feb. 2015]
- Domingo, M. & Kress, G. (2013) 'Multimodal and Ethnographic Semiotic Analysis of Digital Communication Environments' - MODE Training Document. 27 June 2013. London: IoE/NCRM/ESRC.
- Domingo, M., Jewitt, C. & Kress, G. 'Multimodal social semiotics: Writing in online contexts' in K. Pahl and J. Rowsel (eds.) (pending publication) *The Routledge Handbook of Contemporary Literary Studies*. London: Routledge
- Favell, A. (2006) 'London as Eurocity: French Free Movers in the Economic Capital of Europe' in MP Smith and A. Favell (eds) (2006) *The Human Face of Global Mobility*, New Brunswick: Transaction, pp.247-274
- Favell, A. (2008) *Eurostars and Eurocities: Free Movement and Mobility in an Integrating Europe*. Oxford: Blackwell
- Geertz, C. (1973) 'Thick Description: Toward an Interpretive Theory of Culture' in *The Interpretation of Cultures*. New York: Basic Books
- Gomes, D. & Costa, M. (2014) 'The Importance of Web Archives for Humanities' in *International Journal of Humanities and Arts Computing*. 8.1, 2014:106-123
- Jenkins, R. (2002; Revised Edition 2007) *Pierre Bourdieu*. London: Routledge
- Kitchin, R. (2014) 'Big Data, new epistemologies and paradigm shifts' in *Big Data & Society* April–June 2014: 1–12
- Kress, G. (2010) *Multimodality. A Social Semiotic Approach to Contemporary Communication*. London & New York: Routledge.
- Kress, G. & van Leeuwen, T. (1996) *Reading Images: The Grammar of Visual Design*. London: Routledge
- Lyman, P. '(2002) Archiving the World Wide Web' in Part 4 of *Building a National Strategy for Preservation: Issues in Digital Media* Report. Washington DC: The Council on Library and Information Resources and the Library of Congress.
- Masanés, J. (2006) *Web Archiving*. New York: Springer
- Maton, K. (2008) 'Habitus' in Grenfell, M. (Ed.) *Pierre Bourdieu: Key Concepts*. Durham: Acumen
- Miller, D. & Slater, D. (2000) *The Internet: An Ethnographic Approach*. Oxford: Berg
- Miller, D. (2010) *Stuff*. Cambridge: Polity

Miller, D. (2012) 'Social Networking Sites' in Horst, H. A. & Miller, D. (Eds) *Digital Anthropology*. London: Berg

Mullholland, J. & Ryan, L. (2013) French Capital: A Study of French Highly-Skilled Migrants in London's Financial and Business Sectors – Preliminary Findings Report [via UK Web Archive, site captured 13 Sept. 2013
<http://www.webarchive.org.uk/wayback/archive/20140913220149/http://frenchlondon.co.uk/wp-content/uploads/2012/09/French-Capital-A-Report-on-Preliminary-Findings.pdf> , accessed 10 Feb. 2015]

Rowell, J. (2011) 'Carrying my family with me: artifacts as emic perspective' in *Qualitative Research*, June 2011,11 (3): 331-346

Spaniol, M., Mazeika, A., Denev, D. & Weikum, G. (2009) "'Catch Me If You Can": Visual Analysis of Coherence Defect in Web Archiving' in The 9th International Web Archiving Workshop (IWAW 2009) : Workshop Proceedings, available at
http://iwaw.europarchive.org/09/papers/spaniol_iwaw.pdf [accessed 30 Jan. 2015]

Wang, T. (2013) 'Big Data Needs Thick Data' in *Ethnography Matters* (13/05/2013). Available at
<http://ethnographymatters.net/2013/05/13/big-data-needs-thick-data/#more-4782> [accessed 10 Feb. 2015]

Yoon, A. (2013) 'Defining What Matters When Preserving Web-Based Personal Digital Collections: Listening to Bloggers' in *The International Journal of Digital Curation*, 2013, 8:1, 173–192.