



HAL
open science

Bayesian Clustering of Housing Preferences and Lifestyles: a Comparative Analysis of Home Ownership in South-Eastern France

Giovanni Fusco, Floriane Scarella

► **To cite this version:**

Giovanni Fusco, Floriane Scarella. Bayesian Clustering of Housing Preferences and Lifestyles: a Comparative Analysis of Home Ownership in South-Eastern France. 2014. halshs-01505436

HAL Id: halshs-01505436

<https://shs.hal.science/halshs-01505436>

Preprint submitted on 11 Apr 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Bayesian Clustering of Housing Preferences and Lifestyles: a Comparative Analysis of Home Ownership in South-Eastern France

Giovanni FUSCO, Floriane SCARELLA

UMR CNRS 7300 ESPACE

Université de Nice-Sophia Antipolis

98 Bd Herriot - BP 3209, 06204 NICE cedex 3, France

giovanni.fusco@unice.fr, floriane.scarella@unice.fr

ABSTRACT. One of the most important housing research issues is the interplay of socio-cultural factors (lifestyles) and of objective constraints (increasingly important in tense metropolitan housing markets) in shaping households' housing preferences. This paper presents a modelling application for interview/survey data analysis on home ownership within metropolitan areas. The goal of the application is to develop an exploratory methodological protocol to analyse the interplay of lifestyles and objective constraints in home ownership decisions, as well as the spatial inscription of residential trajectories and of mobility behaviours. Its final result is the identification of coherent socio-geographic profiles capable of segmenting new home owners within the study areas of Provence and of the French Riviera. Socio-geographic profiles are identified and characterized by a new five step clustering protocol using Bayesian Networks (BN). BN modelling can integrate both qualitative and quantitative data, and characterise profiles in a probabilistic framework. Socially appreciated suburban and as well as urban lifestyles are present in Provence and on the French Riviera, shaping particularly similar socio-geographic profiles. Peculiarities of the geographical configuration of the two metropolitan areas and of their housing stocks can explain the residual differences among the corresponding profiles.

RÉSUMÉ. Les interactions entre les facteurs socio-culturels (styles de vie) et les contraintes matérielles dans la formation des préférences des ménages sont au cœur des recherches sur le logement. Cet article présente une modélisation pour l'analyse de données d'entretiens issus d'une enquête sur l'accession à la propriété en contexte métropolitain, réalisée au sein des métropoles azuréenne et provençale. Un protocole méthodologique exploratoire permettant l'analyse des interactions entre styles de vie, contraintes, trajectoires résidentielles et pratiques de mobilité est ainsi développé. Une nouvelle méthode de classification par réseaux bayésiens est proposée, intégrant conjointement des données quantitatives et qualitatives, elle identifie et décrit sous une forme probabiliste des profils socio-géographiques cohérents. Ainsi, des profils similaires, liés à des styles de vie socialement valorisés plutôt urbain ou périurbain, apparaissent au sein des deux aires d'étude, quelques différences résiduelles liées à leurs spécificités (configuration spatiale et parc-logements) sont toutefois observées.

KEY WORDS: Housing Preferences, Lifestyles, Bayesian Networks, Socio-Geographic Profiles, Provence, French Riviera.

MOTS-CLÉS: Choix Résidentiels, Modes de Vie, Réseaux Bayésiens, Profils Socio-Géographiques, Provence, Côte d'Azur.

1. Introduction

Local authorities and urban planners are presently confronted with complex issues in the control of urban development. Construction of new sprawling residential subdivisions, urban densification, city centre decay or revival, are apparently different outcomes of housing policies within a city. Social segregation within expanding metropolitan areas is also a main issue in urban policy. Attempts to control physical urban form as well as social form through plans and regulation are, however, often thwarted by agents' autonomous behaviours which are difficult to master. Home ownership decisions by households seem to play an important role in such behaviours and even more within the context of integrated metropolitan areas, where city-centres, suburbs and rural villages constitute overlapping housing markets. It becomes thus important to study the complex interweaving of preferences, lifestyles, motivations and objective constraints of home ownership decisions of households (Kauko 2006, Aero 2006, Rougé 2007). Socio-economic analysis should thus integrate more explicitly elements of the spatial context within which these decisions are taken (Da Cunha *et al.* 2004, Debusschere 2005, Skifter Andersen 2009, Carpentier 2010) as well as elements describing daily and weekly mobility behaviours of households within the metropolitan area (Kaufmann 2000, Rougé 2007, Carpentier 2007, 2010). Integrating space and mobility would open new perspectives for applications stemming from spatial data infrastructures. Nevertheless, traditional databases on housing markets (including census data on housing) describe housing stock characteristics and constraints accurately, but are unable to apprehend the preferences, behaviours and motivations of home buyers. Data are thus to be produced within new research designs in order to address emerging research issues more specifically. New data can then open the way to new modelling applications, taking into account quantitative as well as qualitative information on households behaviours.

The modelling application presented in this paper was developed within a wider interdisciplinary research project (Donzel *et al.*, 2011) federating sociologists (UMR LAMES / University of Aix-Marseille, France, and VECT Mare Nostrum / University of Perpignan, France) and geographers (UMR ESPACE / University of Nice, France). The objective was to study the determinants and motivations of new home owners in metropolitan contexts within southern France. The research was exploratory: new questions were formulated and tested over small population samples. Three fields were selected: the metropolitan area around Marseilles, the metropolitan area of the French Riviera and the smaller urban region around

Perpignan. A semi-structured interview protocol was elaborated and used in order to question 160 households having bought their dwelling within the last 10 years. The case studies covered by this paper are the two metropolitan areas of Marseilles and of the French Riviera, where control of urban development appears most problematic confronted to the emergence of integrated metropolitan areas and fragmented governance (Fusco and Scarella 2013, Decoupigny and Fusco 2009).

Within this context, we wanted to test new methodologies for the numeric treatment of interview/survey data, combining sociological aspects revealed by interviews (or open question surveys) and spatial information extracted from the residential trajectory of the households. The objective was to identify coherent socio-geographic profiles capable of segmenting new home owners within the study area. The coherence of the profiles would result from grouping recurrent logics within revealed preferences, behaviours, socio-economic conditions and spatial inscriptions of residential and daily mobility. Home ownership decisions can thus be considered in the broader context of lifestyles (Bourdieu 1979), going beyond contingent considerations of available housing units and/or transportation facilities. Looking for socio-geographic profiles is also a first attempt to generalize the logics within individual household interviews, although in an exploratory way, given the extremely limited size of the panel interviewed. This attempt was made through the application of new clustering protocols based on Bayesian artificial intelligence.

The next section (2) will present the main issues of the present debate on home ownership determinants. The study areas and the interview design will be presented in section 3. A further section (4) will introduce the new methodology employed for data segmentation, based on Bayesian networks. Section 5 will show the results obtained for the two study areas and will propose some elements of comparative analysis. A final section (6) will resume these results and evaluate the advantages and shortcomings of the modelling application. Perspectives of future research will also be highlighted.

The present paper is more encompassing than a previous text presented at the Agile 2012 Conference (Fusco and Scarella, 2012), whose aim was only to present the interest of the new methodology of data segmentation. The comparative analysis of two metropolitan areas and the positioning within the broader debate on housing preferences gives more insight on the potentials of the new methodology to address current research issues.

2. Home Ownership Between Lifestyles and Constraints: Emerging Research Issues in Housing Studies

Economic analysis of housing preferences has a long tradition but has been criticized (Maclennan 1977, Maclennan and Tu 1996) for its exclusive focus on consumers' rational choice based on functional aspects of housing (surface, comfort, building type, etc.) and economic constraints of households, eventually integrating

functional aspects of housing location (accessibility, presence of service, etc.). New research perspectives have thus been opened by sociology on the role of lifestyle and socio-cultural influence on housing choice. Bourdieu (1979) suggests that housing consumption, like all other kinds of consumption, is directed by the subtle interplay of factors of cultural and symbolic capital. Households assume and publicly show lifestyles as a manifestation of their share of cultural as well as economic capital. Lifestyles are moulded by the *habitus* that individuals receive during their upbringing and continuously develop during their life in socially situated interaction with others, influencing the way they perceive, think and act. Lifestyles thus become a socially classifiable system of practices. A communicative aspect of consumption prevails over the classical functional aspect of consumption assumed by economic analysis. As long as households have degrees of freedom in housing choice (which is the case of most middle- and upper-middle-class households in not too constrained housing markets), lifestyles would play a fundamental role in housing behaviours and residential mobility patterns.

Bourdieu remarks that the upper classes have a fundamental role in shaping lifestyles that become the social norm for the rest of society. More particularly, the upper classes disposing of the highest cultural capital define what should be socially considered as “good taste” within a “legitimate” culture. This has major impact on consumers’ choice, as for example in the case of housing. The upper classes have been the first to develop suburban lifestyles in XIX century England and America (Mumford 1961). They are at present reformulating new lifestyles which are more concerned with locations in the vibrant centres of the global cities and with their cultural life. These new lifestyles related to cultural capital could consequently upset traditional assumptions on housing choice: location would take a more important role than functional aspects of the dwelling and “soft” aspects of location (perception, symbols, socially shared tastes) would take over “hard” aspects (accessibility, service). Ley (1986) highlights that even in North-America the preferences of households disposing of the highest levels of cultural capital have shifted to urban topophilia, reverting the cultural bias towards suburban environments that had been previously hegemonic within anglo-saxon cultural contexts: a flat or a town-house in a highly prized urban context would thus be preferred to a detached house within what has become mass-consumption suburbia.

Adding to Bourdieu’s analysis of the early formation of *habitus* in personal history, Clapham (2005) points to a certain path dependency in housing. Individuals’ housing aspirations and preferences tend to be influenced by their housing career. This opens the way to a more thorough consideration of the residential trajectory of individuals and households, exploring their past mobility between geographical locations, housing types, tenure types, etc.

The integration of a mobility component in housing choice goes beyond mere residential mobility. Mobility sociologist Kaufmann (2000, 2007) demonstrates the existence of a “mobility continuum”: long-range migration, residential migration within a given urban area, tourism, week-end leisure mobility and daily mobility are different interrelated aspects of this mobility continuum. New lifestyles emerge

shaped by the ability to master time and space graduating the different kinds of mobility as needed. Lifestyles in housing choice could hardly be interpreted without taking into the picture the other kinds of mobility of individuals within a household. This had already been pointed out in previous studies of suburban lifestyles and the peculiar patterns of automobile dependency within them (Newman and Kenworthy 1998, Dupuy 1999, Duany *et al.*, 2000). The appropriation of space through both housing choice and mobility behaviour is thus at the base of the “dwelling modes” defined by Stock (2004), which could be seen as a projection of lifestyles in geographical space.

High “motility” according to Kaufmann characterizes widening social groups, but is not the common lot of every household. This is even more evident when residential mobility becomes the focus of the analysis. Of course, not all households are able to freely choose their housing location, due to constraints of housing supply and disposable income. Housing market tensions are typical in the most affluent metropolitan areas of Western Europe. According to Wiel (1999), the integration of local housing markets is enhanced by the increased daily mobility of households within a metropolitan context. This produces higher real estate prices, as a wider demand has access to limited coveted housing stock, especially at the higher end of the housing market and in the most central locations where competition for space also comes from other metropolitan functions (company headquarters, commerce). Planning policies can finally play an important role. According to Bootsma (1998) the recent attraction of city centres in the Netherlands is less the result of a change in household preferences than the consequence of tensions on the housing market resulting from planning policies limiting the peripheral growth of Dutch cities.

These different points of view have impregnated empirical research on housing preferences, behaviours and lifestyles in recent years. In the brief overview that will follow, we will focus both on main results and methodological options from a selection of recent empirical works, mainly in Northern-European contexts.

Directly influenced by Bourdieu’s new sociological approaches, Kauko (2004) explores the relative importance of intrinsic housing qualities *vs* location, as well as of “soft” *vs* “hard” elements of location, within the metropolitan context of Randstadt Holland. The goal of his research is to evaluate to what degree the new socio-cultural aspects can be taken into account to explain household preferences. In-depth interviews are used to question 17 housing market actors in order to assess the relative importance of several factors in households’ housing behaviours. The analysis proceeds from theoretical inference, generalizing from interpretations and not directly from the data. Even if spatial contexts are taken into account (Amsterdam *vs* Rotterdam-The Hague *vs* peripheral developments around these metropolitan centres), the final result is not an explicit segmentation of households according to geographic and behavioural traits, but the identification of main trends within the Randstadt housing market. “Hard” and intrinsic dwelling qualities are thus main factors shaping the housing market in the metropolitan peripheries whereas socio-cultural “soft” factors of location tend to be more important in the city centre of Amsterdam. Most importantly, the study does not explore stated

preferences by households but only overall opinions given by housing actors (developers, real estate agents, planning officers). Moreover, the goal of the study is to assess the relative importance of different attributes for the average household having residential choice (hence upper-middle-class households) and does not aim at segmenting households according to housing preferences and lifestyles.

Aero (2006) more clearly defines a lifestyle approach to housing studies. Based on sociological theory by Bourdieu and Hojrup, he identifies three “modes” in housing choice, encompassing several lifestyles: a pre-modern mode for households who do not chose because they reproduce the conditions experienced in family history; a modern mode where housing is the object of a deliberate choice reflecting household’s lifestyle, social position, values and taste; a late modern mode, where housing choice is constantly redefined in order to respond to different requirements from professional life and/or social engagement.

On this basis, Aero carries out an empirical research on a relatively large sample (813 households) representative of the population of the city of Aarhus (Denmark). The survey protocol addresses questions of housing choice and lifestyle, crossing them with more traditional demographic and socio-economic parameters. As far as space is concerned, particular attention is given to the characterisation of built-up forms of residential districts, whereas the position and the accessibility within the metropolitan area is not taken into account. Multivariate analysis is finally carried out to cross-analyse lifestyle parameters, socio-economic and housing variables.

The interest and the limit of this seminal work is that lifestyles are not inductively searched for. They are postulated and identified from sociological theory and matched with the characteristics of the residential district types. The logit model employed is thus a model of residential district choice, given a certain number of socio-demographic and lifestyle variables. Not surprisingly, it mirrors the statistical models of modal choice in transportation studies. But the matching between lifestyles and district types is not always that obvious. Certain district types can be chosen by households with different lifestyles given the complex interplay of their position in the life-cycle, of their economic constraints and of the tenure type available in the housing stock of residential districts.

Building on Aero’s theory of housing choice, Skifter Andersen (2011) examines the preferences for home surroundings and locations in Denmark. Once again a large sample of households (1500) representative of the Danish population is surveyed. The focus here is more on identifying preferences associated with different phases in the lifecycle of households and on evaluating the relative importance of these preferences in different geographical contexts (urban core, periphery, little town, the countryside, etc.). The survey does not address important realms of lifestyle/dwelling modes (mobility behaviours, housing types) nor financial constraints. At the same time, space is uniquely addressed in terms of location of the present dwelling and residential trajectories are not taken into account. Through factor analysis, the author can nevertheless identify main factors of preference which are later cross-analysed with the position in the life-cycle and geographical location.

A segmentation of the housing market in a metropolitan context is conducted by Debusschere (2009) on a large survey database representative of the housing market of Flanders (Belgium). Both property and private rental markets are analysed within this research. Housing in Flanders is put in the context of metropolisation, which is seen as a unifying force pushing to pan-European housing market integration. Its impact on a local scale is nevertheless overlooked, whereas local housing markets within the whole of Flanders tend to overlap and to be strongly influenced by the main metropolitan centres of Brussels, Antwerp and Gent. Three themes are covered by the Flemish survey: the dwelling, the household and the location (only described through a degree of urbanisation going from metropolitan core to rural area).

As far as the methodology is concerned, the advantages of analysing a large database are offset by the limitations of traditional clustering techniques. Only (quasi-)metric variables are exploited to define clusters, qualitative variables being subsequently used to detail the already identified clusters. In terms of variables, only hard, tangible factors are taken into account. No room is left for “softer” elements linked to the socio-cultural context of housing behaviours (values, perceptions, lifestyles). The segmentation is thus unable to throw new light on the emerging issues of functional *vs* socio-cultural determinants in the housing market. The latter is mainly analysed in terms of financial and lifecycle constraints on households. The study explores the integration of space and time in the analysis. Previous housing situation is thus sometimes taken into account, although not yet in terms of residential trajectory in geographic space. This, together with the relatively poor description of location within the Flemish metropolitan context, brings the author to identify a better integration of the geographic dimension within household segmentation as an important research perspective.

A relatively weak presence of socio-cultural aspects within a robust database also characterises the survey on housing preferences in the urban area of Lausanne, Switzerland (Da Cunha *et al.*, 2004). Once again the geographic space is segmented according to an urban to rural gradient and housing preferences in terms of dwelling and tenure types, location, but also required time for the journey to work are enquired confronting the present dwelling and a desired future one. A few elements of mobility behaviours are introduced (namely travel-time in journey to work). Financial constraints are indirectly taken into account in differentiating three categories of households: active residentially mobile ones (planning a future move), captive residentially immobile (wanting to but not being able to move) and sedentary (not wanting to move). Focussing on housing satisfaction, lifestyles / dwelling modes are nevertheless not directly addressed in this work.

On the contrary, the combination of socio-cultural aspects and economic constraints is well studied by Rougé (2007) in his work on modest households' housing choice in the metropolitan area of Toulouse (France). Discourse analysis carried out on a corpus of 38 interviews allows a deep search of motivations, behaviours and lifestyles of households. Single family homes “must” be the choice for these modest households, given the social norm in Toulouse and their desire to abandon stigmatized social housing. But severe financial constraints reduce location

choice to nothing and the most peripheral locations are thus the only affordable option for home ownership. Daily mobility becomes the adjustment variable in order to ensure a minimal social integration for these households. The result in terms of dwelling mode is strikingly different from the hyper-mobile lifestyles of upper-middle-class suburban households. Longer constrained car trips, reduction of leisure mobility or even isolation at home for spouses and teenagers in one-car households, are the most frequent outcomes. Rougé's research is particularly informative of the role of both cultural and financial constraints in shaping lower-middle class dwelling modes. Its exploratory power is nevertheless offset by the intrinsic limitations of a purely qualitative approach which cannot be used for more consequent databases.

The same willingness to explore both housing choice and mobility behaviours is present in Carpentier's works on Luxembourg (2007, 2010). In his first work, Carpentier carries out an exploratory analysis on housing location and mobility behaviours based on a small sample of 9 semi-structured interviews. The author first applies textual data mining (Lebart *et al.*, 1998) to characterise perceptions of transportation modes and various aspects of household life. Discourse analysis is then developed on the basis of these preliminary results. Carpentier's conclusions confirm Bourdieu's and Kaufmann's theories: household choices (both in terms of transportation and housing) cannot be reduced to simple utilitarian logics but draw heavily on representations, norms and values reflecting households' social identities. As already pointed out by Rougé in the case of modest households in Toulouse, even for more affluent households in Luxembourg, housing choice prevails on mobility behaviours: mobility is the adjustment variable to allow the fulfilment of households' activity agendas whenever the housing location doesn't offer the required services. Perceptions and values are consequently accommodated to justify necessary mobility patterns without questioning the housing choice. Carpentier's ecological approach to mobility is precisely this joint analysis of spatial, social and perceptive factors in apprehending household mobility behaviours.

In the second work (2010), the author analyses a representative sample of Luxembourgish households (600) surveyed on their perception of transportation modes, housing types and desired housing location. Housing location and characteristics are thus better integrated in analysing dwelling modes and lifestyles. Carpentier actually comes to the conclusion that no sharp contrast opposes urban and rural dwelling modes in Luxembourg. A smooth gradient in mobility behaviour and housing preferences can be observed within the little Dukedom, perhaps reflecting the relatively weak capacity of the capital city to catalyse stronger urban lifestyles. The main goal of the work is precisely to evaluate the different aspects of this gradient (households' housing satisfaction, opinions on transportation modes, etc.) according to their localisation and not to identify household profiles. Carpentier's main contribution is more in the empirical demonstration of the interdependence of the two realms of housing and mobility for households. Housing and mobility are not just related in a simple financial trade-off, as often assumed by economic analysis, but are complementary aspects defining more encompassing dwelling modes, interwoven with values, opinions, perceptions and social norms.

3. Case Studies and Survey Protocol

It is precisely to answer the emerging issues in housing research in partially constrained metropolitan contexts and to overcome the limitations of traditional techniques that our research was conceived. Analysing lifestyles / dwelling modes with quantitative methods is a challenging task. These research issues are indeed usually explored by qualitative approaches. At the same time, discourse analysis cannot be applied to large databases and new ways of generalizing qualitative knowledge must be sought after. Within a more attentive geographic reading of dwelling modes, we also aim at better characterising residential and daily mobility of households within the metropolitan space.

We thus selected two case studies of the emerging metropolitan areas in the Provence-Alpes-Côte d'Azur (PACA) region in South-Eastern France. This region has traditionally been very attractive for both population and tourists nationwide. It is also characterised by a bicephalic urban system. Metropolisation of the PACA region during the last decades has resulted in the emergence of two polycentric metropolitan areas (Fusco and Scarella 2013). In the west, the Provençal metropolitan area is structured around the two poles of Marseille and Aix-en-Provence, but increasingly integrates the further cities of Toulon and Avignon. In the eastern part of the region, the French Riviera metropolitan area is structured around the main coastal cities of Nice, Monaco, Antibes and Cannes. A precise delimitation of the two metropolitan areas is nevertheless not easy, as we are dealing with fuzzy continually evolving geographical entities. We have thus preferred a relatively large definition (Figure 1). The French Riviera metropolitan area covers thus the whole French Department of Alpes-Maritimes as well as the Principality of Monaco and the eastern part of the Department of Var. It includes around 1.5 million inhabitants, three quarters of which in Alpes-Maritimes. Stretching over more than 150 km from east to west, the Provençal metropolitan area has around 3 million inhabitants, two third of which in the Department of Bouches-du-Rhône.

Beyond strong similarities (in topographic constraints, the attractive role of the coastal area, the fragmentation of the metropolitan governance, the common belonging to the same Mediterranean culture) the two case studies offer an interesting opportunity for comparative analysis. Urban forms and socio-economic functioning characterising the port and the industrial areas north of Marseilles are for example hardly to be found on the French Riviera, where the touristic and residential economy is much more present. The housing stock also presents important differences, as can be seen from data from the 2009 French Census. Secondary homes make up for 23% of the housing stock in Alpes-Maritimes, but only 3.5% in Bouches-du-Rhône. Within main homes, single family houses are more common in Bouches-du-Rhône (39.2%) than in Alpes-Maritimes (26.7%, the second lowest value in France after the Department of Paris). Several factors could explain the stronger weight of flats on the French Riviera: the influence of the secondary home market (pushing developers to build small flats more than houses), the stronger topographic constraints and more expensive real estate and finally the

denser pattern of city centres in the coastal area, reducing opportunities for low rise green-field development. Home owners account for roughly a half of households (but 55% in Alpes-Maritimes vs 51% in Bouches-du-Rhône). The private rental market has a similar weight in the two Departments (around 30%), whereas social housing is more important in Bouches-du-Rhône (15.1% of main homes) than in Alpes-Maritimes (8.5%). Finally, a major difference can be highlighted in the geographical configuration of the metropolitan areas. The metropolitan centres of the French Riviera are all coastal cities and this is true even for secondary urban centres (with the exception of the city of Grasse). On the contrary, the metropolitan and the secondary urban centres in the Provence cover the Mediterranean coast and the hinterland more evenly. The second metropolitan belt (see further) can thus also offer residential opportunities in city centres (like La Ciotat, Manosque or Arles), whereas the second belt for the French Riviera is mainly made of suburban residential districts or villages within a rural/natural alpine countryside.

The interest of a comparative analysis is also to verify to what degree these structural differences in the housing market and in their geographic configurations can influence new owners' housing choices. The role of financial constraints will also have to be evaluated, as these two coastal metropolitan areas are among the most expensive housing markets in France, after Paris. But more specifically, we want to understand how these objective constraints interact with socio-cultural aspects in the two study areas to identify distinctive socio-geographic profiles among new home buyers: elements of lifestyle, opinions, mobility behaviours and residential trajectories in the metropolitan space will be taken into account along with the constraints of the local housing markets. Despite the existence of exhaustive databases on housing stocks (INSEE Census and BD-FILOCOM by the French Ministry of Finance), only a specific survey of a sample of new home-owners in the two study areas can integrate all these aspects of the enquiry.

The polycentric metropolitan structure of the two study areas allows the distinction of three zones in each field, according to a centre-periphery gradient having important implications for the real estate market. Nice, Monaco, Antibes and Cannes, on the French Riviera, and Marseilles, Aix-en-Provence, Toulon and Avignon in Provence, are thus the main centres of the two metropolitan areas, concentrating most of the economic activity. A first peripheral belt around these centres is made up of all the municipalities within 20 minutes from these centres on the road network. The second peripheral belt includes further municipalities, often in the mountainous hinterland.

This geographic subdivision, as well as age class (less than 40, 40-60 and more than 60 years old) have been used to stratify the samples of 54 and 59 households interviewed on the French Riviera and in Provence, respectively, based on knowledge of localization and age of main home purchases thanks to the FILOCOM database (Figure 1). Only households having bought their dwelling within the last 10 years (more precisely between 1997 and 2010) were interviewed. Within this exploratory application, a simple quota sampling method was implemented (in both samples the older households are nevertheless underrepresented).

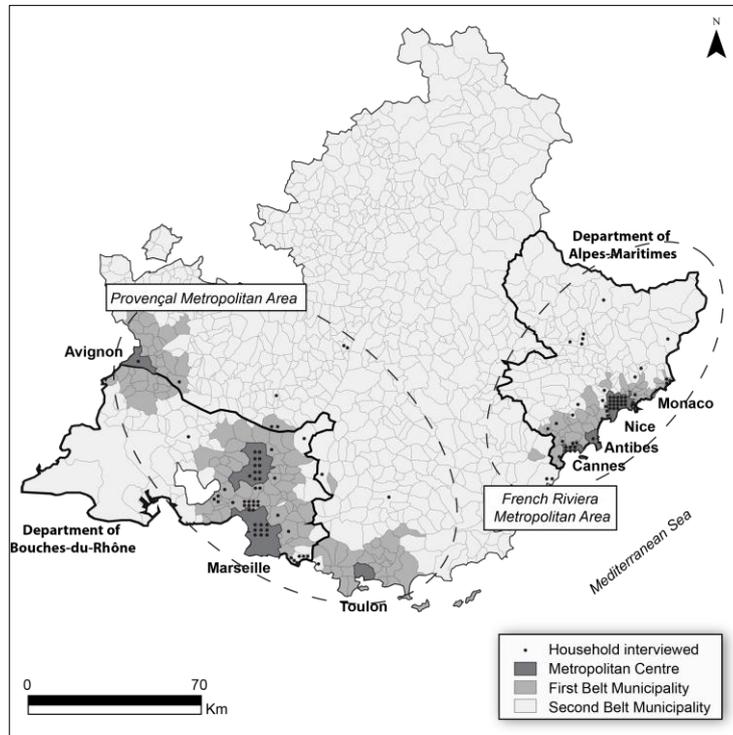


Figure 1. Localization of the Household Sample within the Two Study Areas.

Semi-structured in-depth interviews were realised face-to-face from ten different interviewers over the two study areas between the fall 2009 and the spring 2010. The interviews were recorded and later retranscribed. The interview protocol was conceived from sociologists at the UMR LAMES and was applied with very slight differences to the two study areas. The protocol was formalized by an interview guide covering three chapters: 1. the purchase of the home and the main housing choice criteria, 2. social relations within the neighbourhood and mobility behaviours, 3. future housing projects and desires. More easily measurable parameters were written down by the interviewer at the end of the interview (description of the dwelling, socio-demographic parameters of the household, etc.). Description of the previous dwelling, of its location and tenure type was also addressed during the interview, allowing for the reconstruction of the residential trajectory of the household. The scope of the interview was precisely to address lifestyles / dwelling modes in a relatively broad way and to relate them to housing choice criteria, to the constraints of the home purchase and to the residential trajectories in the metropolitan space. The retranscriptions of the interviews were only the raw material of our modelling application. In order to fully exploit the qualitative data within a quantitative approach, an analysis grid has been applied *a*

posteriori to the interviewees' texts in order to transform the information freely given to the interviewers in 46 mainly qualitative variables (Table 1):

- 6 factual variables on the household (revenues, composition and number of active people, age, socio-professional class, region of birth of the reference person).

- 4 variables for the location of the present and previous dwelling both at the metropolitan scale (metropolitan municipalities / first belt / second belt) and more locally in terms of morphological characteristics of the neighbourhood (city centre, urban periphery, suburb, rural village). One of the characteristics of the study areas is the presence of suburban contexts even within the metropolitan municipalities.

- 11 variables for objective data of the present dwelling (number of rooms, construction period, housing type and the presence of several comfort elements in the dwelling or in the condominium/subdivision). Number of rooms and housing type are also recorded for the previous dwelling.

- 4 variables for stated motivations of the home purchase, for tenure type and for the eventual ownership of a secondary home.

- 8 variables on the financial constraints and on the home purchase (year of purchase, search time, presence of intermediaries, cost, use of mortgage, etc.) and on eventual works which had to be realised in the dwelling.

- 6 variables for the mobility behaviours of the household (visited space for daily needs and leisure, journey to work, transportation modes used, possession of cars).

- 7 variables on the stated perceptions of households of their neighbourhood, on their engagement in the local community, on their relations with neighbours, their degree of satisfaction and their future housing project.

This grouping of variables doesn't correspond to distinct dimensions of the analysis but to practical aspects of data type (objective/stated, geographic/socio-demographic/financial, housing/mobility behaviours).

By coding texts after a careful interpretation of the interviewees' discourse, we also limited the biases linked to households' statements (identified, for example, by Carpentier 2010). Within interviewees' discourse, contradictions may be present. It was for example common for people to overstate the "good" or "very cordial" relationships with neighbours, whereas the further development of the discourse revealed a substantial absence of relations, eventually punctuated by minor conflicts. By interpreting the answers in the context of the whole interview, we could thus avoid a naïve reading that would have emerged from the semantic textual statistics.

To give a quick glimpse at the main characteristics of the two samples, 76% of interviewed households on the Riviera bought a flat (85% already lived in a flat previously), mirroring the strong presence of flats in the local housing market. In Provence, 54% of interviewed households bought a single-family house, even if 80% of them previously lived in a flat. A general move from collective to individual housing thus characterises new home buyers in Provence. Overall satisfaction with the new housing is nevertheless similar in the study areas, as different degrees of dissatisfaction account for 60% of the sample in Provence and 56% on the Riviera.

4. Bayesian Analysis of Interview/Survey Data

Coding interview texts produces two databases with the following features:

- Data are mainly qualitative (yes/no, very satisfied/quite satisfied/dissatisfied).
- Quantitative data are discrete or have been discretised.
- Variables should make it possible to relate housing preferences to lifestyles, financial constraints and residential trajectories in geographic space
- The databases are very small (59 and 54 records, respectively).
- A few data are missing (almost randomly among the variables, but essentially in the Provençal sample).

How can socio-geographic profiles be identified among new home owners in Provence and on the French Riviera, in order to generalize the narratives of every single interview? We propose to take up this challenge by a new clustering application using Bayesian Networks (BN, Pearl 2000, Jensen 2001, Korb and Nicholson, 2004, Pourret *et al.* 2008). The application follows a pioneering protocol which has been lately applied in marketing studies (Bayesia-Repères 2008). BN modelling can integrate both qualitative and quantitative data, hard and soft modelling hypotheses and deal with uncertainty issues. BN allow two main kinds of applications. The first is causal knowledge discovery: BN can infer the most probable causal links among the variables from a database of observed records, even in the presence of missing and/or uncertain data. The second kind of application is knowledge inference through Bayesian simulation. Once the causal model is established, BN can perform probabilistic simulation with extreme freedom in the choice of input and output variables. Combining these two kinds of applications, BN can also become powerful clustering tools, as we will show through our application. More specifically, the clustering protocol is structured in several phases, coupling supervised and unsupervised learning from the database. The protocol was implemented using the BayesiaLab software (Bayesia 2009).

Step 1 – Unsupervised Learning of Associations

The first step is the search of probabilistically strong links among the 46 survey variables through unsupervised learning of a BN, a classical knowledge discovery application. At this stage, no previous knowledge is entered by the modeller over the BN structure. Despite the small size of the databases, it was possible to produce a network containing fairly robust links from a probabilistic point of view.

After having eliminated the weakest edges (identified through an edge strength analysis), the same 22 variables and 22 edges for the Riviera and 26 variables and 22 links for Provence are obtained through the application of different BN learning algorithms (Figures 2 and 3). The main difference between the two BN models is the presence of a single network for the Riviera and of a main network with three disconnected network fragments for the Provençal metropolitan area.

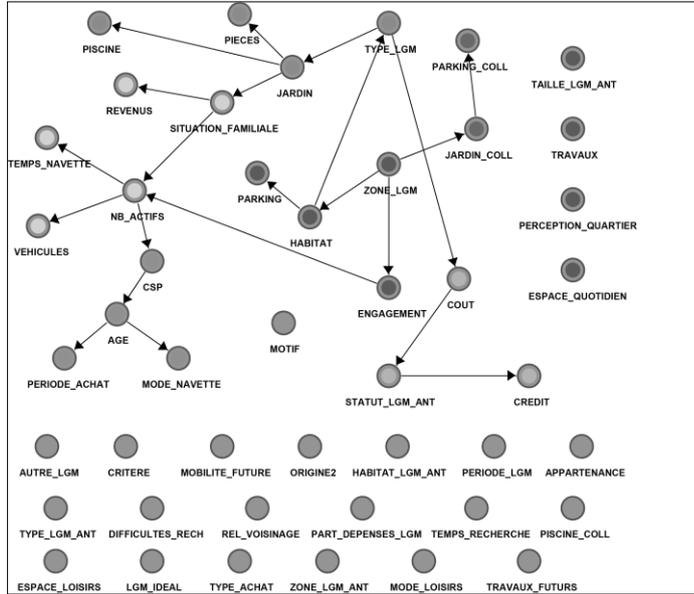


Figure 2. BN Learned from Data and Variable Segmentation for the French Riviera.

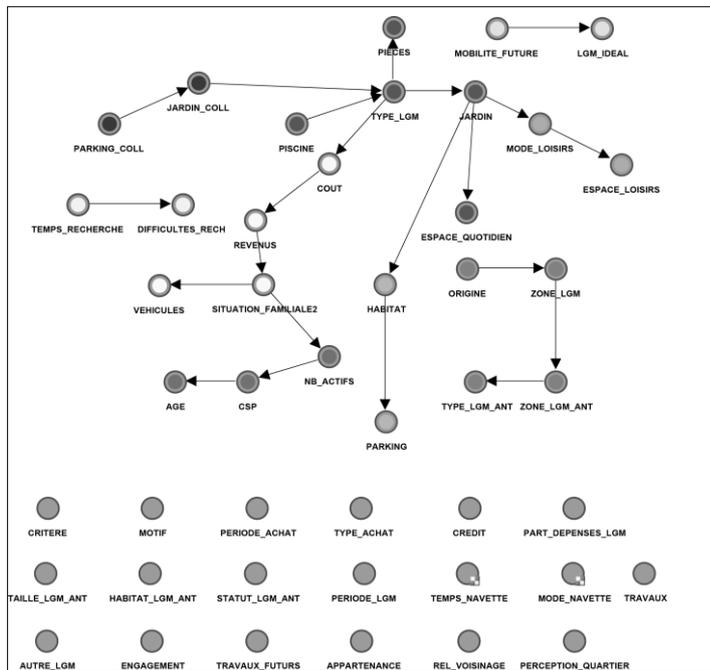


Figure 3. BN Learned from Data and Variable Segmentation for Provence.

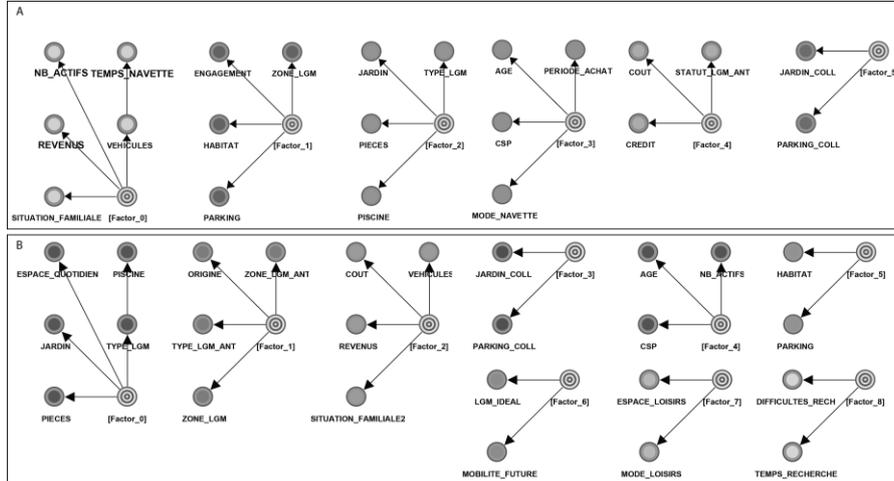


Figure 4. BNs Determining Factors for the French Riviera (A) and Provence (B).

Step 2 – Variable Segmentation

The produced BN are analyzed by a hierarchical clustering algorithm in order to detect groups of closely linked variables, which can be ascribed to a more general concept (Figure 4). Six variable groups could be identified for the French Riviera, reflecting the following concepts:

- Factors 0 (variables of household composition, income, number of active people, commuting time, number of motor vehicles owned) and 3 (variables of socio-professional status and age of the reference person, mode used for commuting to work, period of home purchase) describe the socio-economic characteristics, the position in the life-cycle, but also the logistic organization and the mobility constraints of the household.
- Factor 2 describes the intrinsic characteristics of the dwelling (variables of dwelling type, number of rooms, presence of a garden and of a pool).
- Factor 1 focuses on the context of the dwelling (metropolitan zone, neighbourhood morphological characteristics, availability of private parking, owners’ engagement in local life). Private parking is thus more a descriptor of the interaction with the geographic context of the dwelling, whereas garden and swimming pool participate in its intrinsic level of comfort.
- Factor 5 reveals the status of the subdivision or of the condominium of the dwelling (availability of collective parking and garden within the subdivision or condominium).

- Factor 4 deals with the real estate transaction (total cost, part financed through loan, status of occupation of the previous dwelling).

The more fragmented BN model for the Provence metropolitan area can be segmented in nine groups of variables, making sometimes different associations among variables and finer distinctions among concepts:

- Factors 2 and 4 correspond to factors 0 and 3 on the Riviera (socio-economic characteristics, position in the life-cycle, logistic organization of the household) but don't include variables on journey to work, whereas Factor 2 integrates the cost of the purchase of the new dwelling (which contributes to Factor 4 on the Riviera).

- Factors 5 and 1 detail the context of the dwelling in terms of neighbourhood (5) and position in the metropolitan space as well as residential trajectory (1). Taken together, they correspond to Factor 1 on the Riviera.

- Factor 0 describes the intrinsic characteristics of the dwelling. It corresponds to Factor 2 on the Riviera and integrates the space used by households on a daily basis.

- Factor 3 describes the status of the subdivision or of the condominium of the dwelling, just like factor 5 on the Riviera.

- Three further factors are identified among the variables of the Provençal metropolitan area: housing satisfaction and future projects (Factor 6), leisure habits (Factor 7) and the search of the new dwelling (Factor 8).

Step 3 – Determining Synthetic Factors

Latent non-observable variables (factors) are created in order to summarize the information contained in every group of variables. The number of factor values is automatically determined by segmentation algorithms according to a likelihood maximization approach within a clustering naïve BN architecture (Figure 4). Factors are very robust and the individual records of the database can be assigned values for each factor with very limited uncertainty. Links among variables contributing to each factor and high mutual information between variables and factors also make factors robust to addition/suppression of variables.

Step 4 – Identifying Profiles that Summerise Factors

A Bayesian classification of individual records can now be performed using the factors (more precisely, for the French Riviera, we added the unconnected but highly informative variable MOTIF to the six factors, describing the reasons of home purchase). For each study area, a new non-observable variable is added, the socio-geographic profile of the home-buyer, considered to be the latent cause of the seven or nine factors, respectively. Once again, the number of profiles is determined automatically through likelihood maximization. Six profiles could thus be identified in the French Riviera sample and five in the Provence sample. The uncertainty in assigning individual records to profiles is extremely low for the French Riviera (clustering purity between 96.52% and 99.37% according to profile) and relatively

low in Provence (clustering purity between 89.06% and 99.89%). Probabilistic assignment of individuals to profiles allows detection of both archetypical and atypical households within each profile, supplying a heuristic for a more detailed analysis of interview recordings. Profile interpretation from factor values is nevertheless unpractical. It will be done in the next step of the analysis using the original 46 variables.

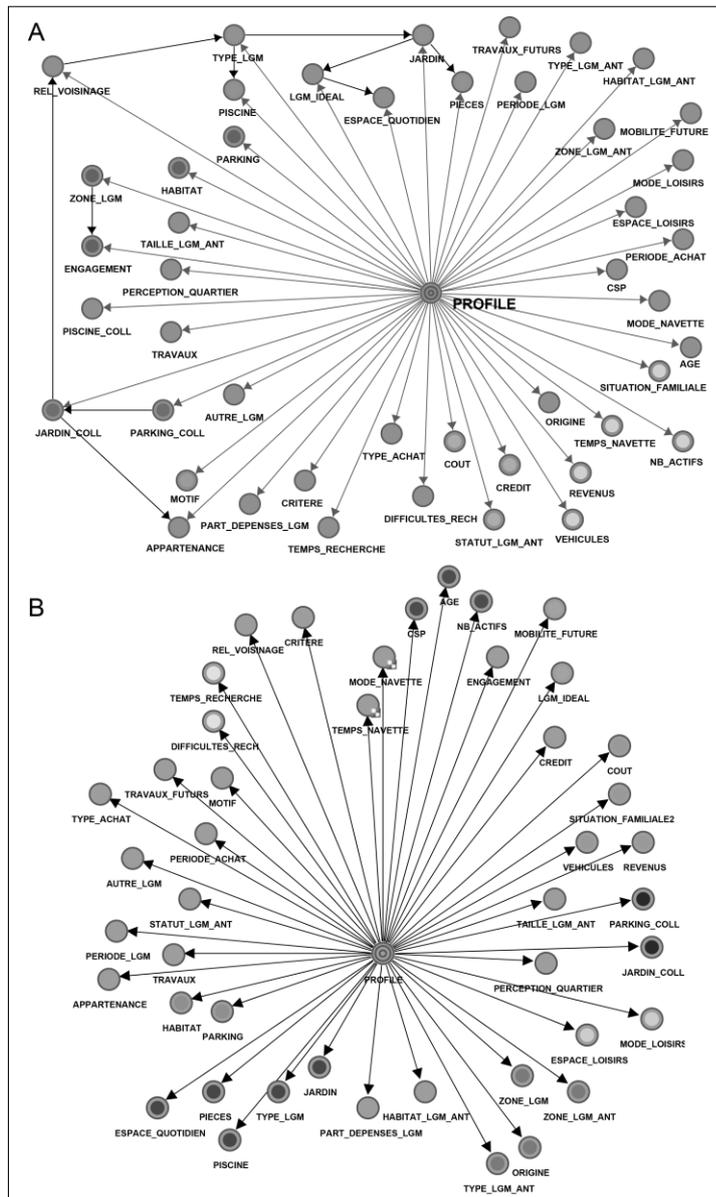


Figure 5. *The BNs Used for Profile Characterization, for the French Riviera (A) and Provence (B).*

Step 5 – Characterizing Profiles

Supervised learning of two new BNs is now performed in order to characterize socio-geographic profiles of new home buyers from the original data plus the newly determined profiles. It is finally possible to easily interpret the socio-geographic profiles previously determined in terms of probabilities of the combination of values of the original variables. The new BNs (having augmented naïve architecture, Figure 5) can precisely be used to infer the characteristics of each profile in the two study areas probabilistically. Variables don't contribute equally to every profile. They can then be ranked with respect to mutual information with profiles. The comparison of the variable ranking for the two study areas is very informative, anticipating a few main differences that will be observed in the socio-geographic profiles. For the French Riviera, the variable MOTIF has the highest mutual information (22%) with the variable PROFILE, justifying its inclusion in profile identification retrospectively. The single most important modality for MOTIF is the quest for a better housing environment. In Provence the desire to become home-owner is the first motive (it is the case for 60% of the sample, even if first time owners make up the same share in the two samples). The widespread desire to become owners makes the MOTIF variable less informative for the socio-geographic profiles in Provence.

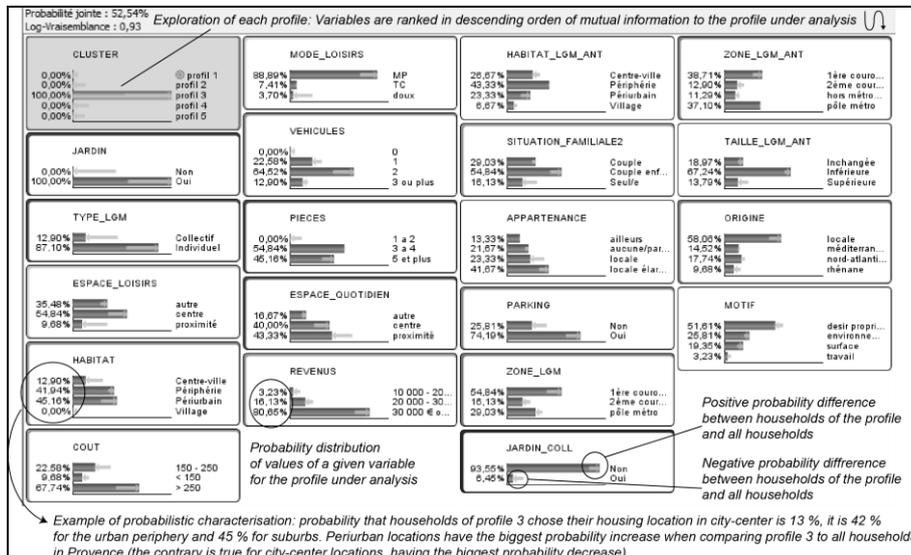
The morphological characteristics of the neighbourhood is one of the most informative variables in both study areas (14.5% of mutual information for the Riviera, 22% for Provence), being closely linked to lifestyles and housing preferences, as shown in literature (Aero 2006, Carpentier 2010). The period of construction of the dwelling is very informative in both case studies, with the two extremes of old (before 1945) and new dwelling being more frequent in Provence.

In Provence, the two most informative variables are nevertheless those describing the intrinsic characteristics of the dwelling (dwelling type and presence of a garden, with 50% and 24.5% mutual information, respectively), which is not the case for the Riviera. Lifestyles associated with individual housing in suburbs and rural settings should be more clearly opposed to lifestyles associated with flats in the Provençal metropolitan area. Finally, the variables associated with leisure activities are more informative for socio-geographic profiles in Provence, as could be already anticipated by the fact that a specific factor for leisure emerged from the original BN. Affecting individuals to one of the profiles is performed with extremely low uncertainty (minimum purity of 99.97% for the Riviera and of 98.68% for Provence), knowing the variables derived from interviews.

5. Socio-Geographic Profiles of New Home Owners: A Comparative Analysis

Through the five steps, six socio-geographic profiles could be identified and characterized within new home buyers on the French Riviera and five in Provence. Some profiles are more recurrent in the samples (namely profile 3 in Provence, as we will see) but the small sample size doesn't allow any extrapolation of the relative weight of these profiles within all new home buyers in the two study areas. The description of each profile will be based only on the variables having the highest mutual information with the belonging to the profile in question and will be done in a probabilistic framework (see Box 1, where the probability distribution of a given profile is also compared to the one of the global sample). Spatial inscription of socio-geographic profiles becomes even more interesting when the whole residential trajectory of households is taken into account. A precise map of profile distribution in space wouldn't be very informative, given the almost uniqueness of spatial trajectories of households on such a map. We thus need to visually aggregate these trajectories using the logical categories of our analysis of neighbourhood and metropolitan space. Figure 6 schematizes and visualizes these trajectories, which will be further commented in the characterization of the profiles.

One of the major results of our clustering application is the striking similarity of the overall results obtained for the two study areas. When merging profiles 2 and 4 on the Riviera, the resulting five socio-geographic profiles of new home buyers are in a one-to-one correspondence with the five profiles in Provence. Examining the details of the profiles, several important differences can nevertheless be highlighted between the two study areas, making the comparative analysis the more interesting.



Box 1: An Example of Probabilistic Characterisation for Profile 3 in Provence.

5.1. *The Wealthy Suburbanites*

Three socio-geographic profiles of wealthy suburban families emerge in the two study areas: profile 3 in Provence (accounting for 52% of the sample), and profiles 2 and 4 on the Riviera (13% and 20% of the sample, respectively). The common features of these three profiles are that they concern wealthy families having bought relatively expensive suburban single-family houses with gardens. The car is their favorite transportation mode both for the journey to work and for leisure, confirming a typical suburban lifestyle. Their dwelling mode is weakly grounded in the neighborhood, both for daily needs and for leisure activities.

Several interesting points differentiate the three profiles. Households in the relatively homogeneous profile 3 in Provence bought their house either in suburban areas or in urban peripheries and their residential trajectories are bound with high probability within the first metropolitan belt. Together with the desire of becoming owners, a quest for a better environment or for a bigger dwelling are also important motivations for the purchase. These households keep on visiting or traditional city-centres, as well as emerging suburban centres. Suburban profiles on the French Riviera are different in their residential trajectories. Profile 2 is closer to its Provençal counterpart, but has higher probability of including households having left city centres and of households buying a flat, even if in a suburban neighborhood. Most interestingly, it concerns professional newcomers on the Riviera and/or younger families looking for a suburban environment not too far away from the city, motivated by the quest for a better environment or a bigger dwelling. Profile 4 is also characterized by departures from the city-centres, but households in this profile head to further suburban spaces in the first or even in the second metropolitan belt. They are wealthier older couples, whose attitude to housing choice reflects Aero's second category particularly well: a deliberate choice of residential environment at the higher end of the real estate market, the fulfillment of the dream of the ideal home showing the social status of the household and a certain desire of natural amenities. The city is kept at distance but is still regularly frequented for leisure by households, who are mainly of local or Mediterranean origin. The presence of children can eventually reinforce the use of neighborhood services. Even if these three profiles participate in a new emerging car-dependent lifestyle trend in a Mediterranean context, none of them shows a complete disaffection from traditional city centres, as it can be observed among North-American suburbanites.

5.2. *The Wealthy Urbanites*

Two socio-geographic profiles of wealthy urban households also emerge in Provence (profile 2) and on the French Riviera (profile 5), with a similar share in the two samples (around 13%). They mainly concern singles or young couples without children having chosen to buy a flat within the metropolitan city centres. If an urban

lifestyle is common to the two profiles, a few differences exist in their residential trajectories and mobility behaviors.

Urbanites on the Riviera are essentially locals, whereas urbanites in Provence have a high probability of being newcomers from North-Atlantic France (mainly Paris). This explains a higher degree of local belonging for the Riviera city dwellers, who were previously living in peripheral neighborhoods around city centres or in villages and see the arrival in the urban core as a way of improving their housing environment (closer to urban amenities). Some of these households are families with children or empty nesters, optimizing their localization in the urban space in the course of their life-cycle. Urbanites in Provence who were already living in inner cities, are much more likely to be singles and their main motivation for buying (beyond becoming owners as in most Provençal profiles) is a professional move. The localization in the city is thus a desire for an urban life-style as well as a career opportunity/constraint. Urbanites in Provence take the most advantage of their central location by using neighborhood service and leisure opportunities, whereas urbanites on the Riviera also visit other destinations in the metropolitan area. Better knowledge of the region through household biography and overlapping influence of metropolitan centres on the Riviera could explain these behavioral differences. Consequently, urbanites on the Riviera are also more heterogeneous in terms of mobility behaviors (namely car use vs. walking and transit).

5.3. The Middle-Class of the Urban Periphery

Middle-classes households in Provence and on the French Riviera are not attracted only by suburban or city-core contexts. Many of them chose an urban periphery neighborhood, as is the case for profile 1 in both samples. Households in these profiles are mainly characterized by centrifugal residential trajectories from the city centres to the periphery or from internal movements within the urban periphery (with a few movements from city centre to city centre in Provence). In both cases, these movements do not concern the second metropolitan belt. The social profile of these households is relatively homogeneous: it is the middle-class of employees (eventually also cadres in Provence). The housing type is not very informative for these two profiles, implying small differences from the probabilities of the whole samples, hence a preference for collective housing on the Riviera and for individual housing in Provence. This also mirrors stronger real estate constraints in the Riviera cities than in Provençal cities (both metropolitan municipalities and secondary urban centres). New housing projects are nevertheless preferred (above all in Provence) and alternatively recent, modern dwellings. This relatively standardized housing offer, often sold by promoters or real-estate agents, seems less capable of mirroring their owners' personal taste and culture (as in the case of the urbanites), but shows more of their established social position. Matching social expectations in the housing purchase puts an important financial burden on the household income. Typically 30 to 50% of household income is thus absorbed by

housing costs, and sometimes more than 50% on the Riviera. Lifestyles / dwelling modes come to terms with these financial constraints. Use of neighborhood services and leisure is thus preferred, even if city centres (which are not too far away for these households) are also visited or leisure activities.

The main difference between the two study areas is the relative importance of these profiles within the sample: 24% on the Riviera (the modal profile) vs. 5% in Provence (the smallest one). Other differences can be highlighted in the search of the present dwelling. The latter was longer and more difficult for households in Provence, constrained by insufficient budget and a preference for logistical criteria (importance of location in terms of access to place of work and/or services). The extreme exiguity of the profile in the Provence sample should nevertheless warn against too much generalization, as over fitting of the profile on the few corresponding households cannot be ruled out.

5.4. *The Middle-Class Utilitarians*

Households of profiles 5 in Provence (12% of the sample) and 3 on the Riviera (17% of the sample) could be characterized as being middle-class utilitarians. In both profiles, households are mainly (exclusively in Provence) motivated by the desire to become owners within an economic rationale: the housing market is becoming increasingly expensive in the two study areas (average prices have roughly doubled between 1998 and 2008) and it is economically sound to invest disposable savings and mortgage credit opportunities in a dwelling. We are thus dealing with first-time owner households having bought their home in the very last years (2008-2010) and considering this purchase as an economic opportunity in the perspective of future resale. The purchased dwellings are most probably small or medium-sized flats in newly constructed buildings of higher status (with common parking spaces and gardens), considered a surer investment. The search for a good price/quality ratio in the purchase makes these households particularly attentive to the intrinsic characteristics of the dwelling, and not necessarily to location. It is thus difficult to determine the most probable residential trajectories within the metropolitan space.

Only a few nuances distinguish the two profiles. Households in Provence tend to be wealthier, have several motor-vehicles and to be more car-dependent. Their residential trajectories are mainly contained within the two peripheral metropolitan belts, with a certain preference for urban periphery as final destination, whereas on the Riviera they also concern movements to and from the metropolitan municipalities. Longer residential movements within the Riviera (for example from metropolitan centre to the second belt) can also imply more important readjustment in lifestyles, which are not to be detected in Provence.

5.5. *The Modest Households of the City Centres*

Profiles of modest households with strong financial constraints are relatively similar in Provence (profiles 4) and on the French Riviera (Profile 6). It is to be remarked that these profiles concern the most modest households within new homebuyers, and not the poorest households in the metropolitan areas, who are tenants in the private and in the social public sectors.

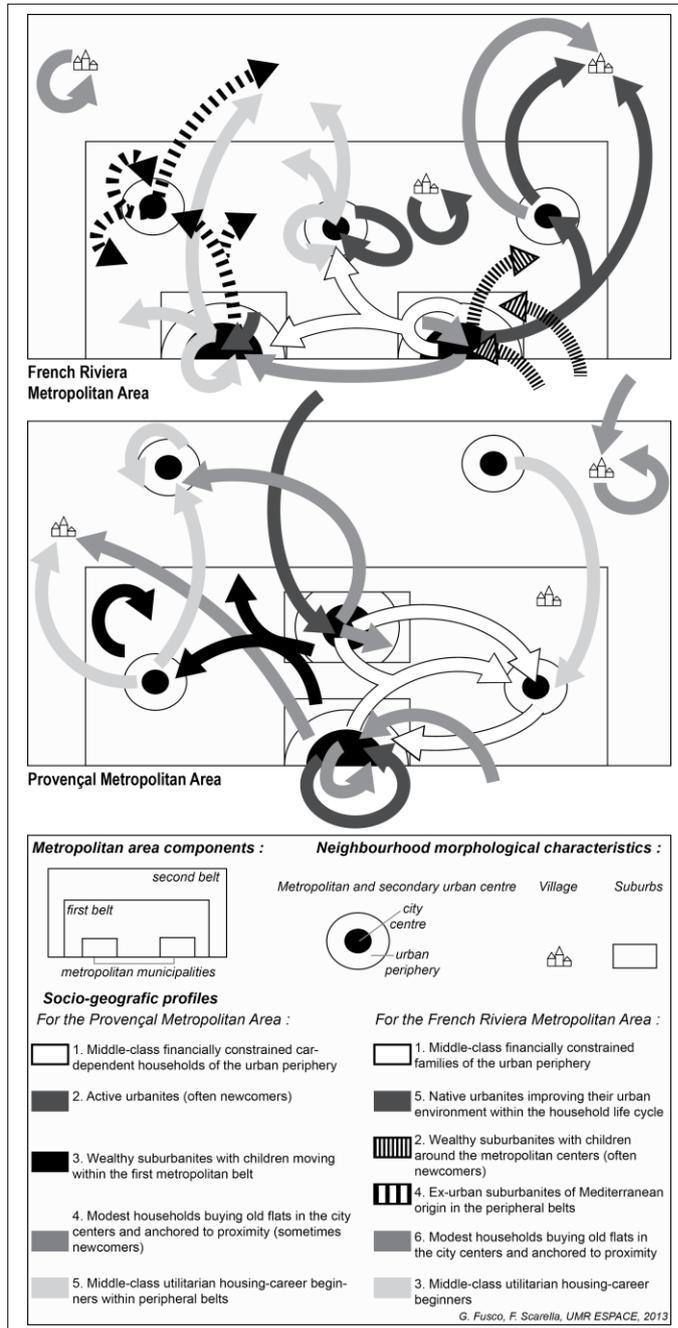


Figure 6. Spatial Inscription of Socio-Geographic Profiles.

Both profiles are mainly made of singles, either young and active or old and retired, partly explaining their particularly low revenues. The dwellings they buy are similar (small flats, at the lower end of the market prices, with few comfort elements and often in old buildings) and are located in the less expensive neighborhoods within city-centres or in the most peripheral villages, where the offer for older dwellings is most important. Their residential trajectories are thus from centre to centre (even if on the Riviera periphery-to-centre trajectories are also relatively probable). Despite their lower living standards and their central localization, they often have a motor-vehicle (car or two-wheeler) and they can thus arbitrate between walking, transit and private transportation for some of their trips. Although by financial constraint, they finally adopt an urban lifestyle, by often giving priority to neighborhood services and leisure activities. They finally have a similar weight in the two samples (13% on the Riviera, 17% in Provence).

One main difference between these two constrained urban profiles is the professional situation of the households, with higher probability of having active people on the Riviera. Housing choice criteria are thus more often logistical (proximity to work and services) or functional (intrinsic characteristics of the dwelling) on the Riviera. Retired and unemployed people are more probable in the Provençal profile, with a non negligible share of them being newcomers in Provence (as in profile 2, they mainly come from the Parisian region). The housing choice has been strongly restricted by the small budget, but housing costs as percentage of disposable income can be relatively low in the case of retired people having sold a previous dwelling. If we can speak in both cases of a constrained urban/village lifestyle, the modest households on the Riviera appear doubly constrained: from their financial situation and from the obligations of an active life. Modest households thus enjoy more freedom in mobility behaviors in Provence. This, together with a change in professional status (retirement coinciding with the purchase of the new dwelling), probably explains a higher tendency of Provençal modest home-buyers to be pragmatically satisfied with their new housing (probability 0.66 of being in the ideal or almost ideal dwelling, given their financial constraints).

5.6. New Home Buyers in South-Eastern France: An Overall View

Bayesian network clustering produced new insights on housing preferences and lifestyles of new home-buyers in South-Eastern France. Data segmentation revealed a mainstream suburban profile in Provence, corresponding to two more particular suburban profiles on the Riviera, one characterizing professional newcomers and the other the suburban dream of local wealthier households. It also revealed two urban profiles in both study areas (wealthy urbanites by choice and modest urban households by constraint) and two more heterogeneous middle-class profiles preferring peripheral urban locations (one as a lifestyle choice, the other mainly on utilitarian grounds). The strong similarities among the profiles in the two study areas

suggest that the results could be more robust than what the exiguity of the two samples implies.

Despite the many analogies among the different profiles in the two study areas, differences in the geographical configuration of the metropolitan areas, in the available housing stock and in the prevailing economic activities can of course produce sometimes divergences in housing preferences and lifestyles in Provence and on the French Riviera.

The suburban lifestyle and, eventually, individual housing within the urban periphery, have thus the tendency to act as a social norm for middle-class and upper-middle-class households among the local population in Provence (often the urban profiles are made up of newcomers in the region). Nevertheless in Provence, as well as on the Riviera, the attraction of denser and functionally more mixed urban peripheries plays an important role in shaping some profiles. At the same time, urban centres are still attractive places, for the residential choices of the urbanites, as well as for the use of leisure activities by urban periphery and suburban households. The attractiveness of urban centres seems even more pronounced on the French Riviera, where an important share of the local population tends to accomplish a whole housing career within the central cities. Polycentric metropolitan configurations (with four metropolitan centres and several secondary urban centres in each study area) also make urban amenities more easily available to households in peripheral/suburban locations.

Making a difference with the modest households in the Toulouse metropolitan area studied by Rougé (2007), the purchase of suburban individual housing is an important social trend for middle-class families (in Provence) and for upper-middle-class families (both in Provence and on the Riviera). More modest home-buyers make the best of old collective housing in the city centres or in the villages of the second metropolitan belt. By so doing, they seem to enjoy a larger degree of freedom in their mobility behaviors than their counterparts in Toulouse, taking advantage of locally available services and arbitrating between car mobility, walking and transit. For middle-class employees, an apartment in new or recent housing projects in the urban periphery seems a socially appropriate housing choice in the French Riviera cities. Socially appreciated suburban and urban lifestyles are thus both present in the two study areas, with an even stronger presence of the urban model on the Riviera. This substantially differentiates our two Mediterranean metropolitan areas from the case of Luxembourg (Carpentier 2010), where a rural/urban gradient reflects the weakness of the central city in catalyzing urban lifestyles for households having housing choice.

The frantic increase of housing prices in South-Eastern France between 1998 and 2008 has also resulted in the emergence of a particular profile of utilitarian-minded middle-class households trying to enter the housing market before being barred from it by ever-increasing prices (a situation not uncommon in Paris). It would be interesting to explore the housing choices of these households in the new phase of downturn which is presently characterizing the housing market in most French

metropolitan areas, including Provence and the French Riviera. Are they still an important component of the local housing market or have they disappeared among the new home-buyers, waiting for a stronger correction in housing prices before coming back on the market?

6. Conclusions and Perspectives

The results obtained from the proposed analysis protocol show the interest of cross-analyzing variables on conditions, terms and motivations of home purchase with behavioral variables and geographic variables inscribing the residential trajectory within the metropolitan space. The origins and the destinations of these residential movements, taking into account the local characteristics of the residential environment and the localization within the metropolitan area, are indeed essential to understand the underlying logics of home purchase and complete the lifestyle / dwelling mode approach to housing preferences. Lifestyle factors and objective constraints of the housing market, of household finances and needs, all contribute to shape socio-geographic profiles within a given metropolitan area. We stress that the proposed socio-geographic profiles cannot be reduced to lifestyles / dwelling modes both for the integration of constraints and for the fact that they also take into account residential trajectories in space and among lifestyles. Profile 4 on the French Riviera is thus characterized both by an inner city to countryside residential move and from an urban to suburban lifestyle change.

Our analysis also confirms that location and spatial inscription of residential trajectories do participate to socio-cultural aspects of housing preferences. The most utilitarian-minded profiles (for example, profiles 2 on the Riviera and 5 in Provence) are thus the hardest to inscribe in space, because they are the less location sensitive.

A few specificities of the two Mediterranean metropolitan areas in South-Eastern France emerge: firstly, the presence of both socially appreciated suburban and urban lifestyles and, secondly, the fact that more modest home-buyers can make the best of constrained urban housing adopting proximity-bound lifestyles in inner cities and villages. As already pointed out for other geographic contexts (Kauko 2006, Aero 2006), “soft”, socio-cultural aspects are not the only factors in shaping socio-geographic profiles. In South-Eastern France, tensions on the local housing markets already shape the housing preferences of the most utilitarian-minded households and constrain the housing choices and residential trajectories of many middle-class and lower middle-class households. Prospectively, they put the capacity of these households to enter the property market altogether at risk. Finally, our analysis of socio-geographic profiles and the identification of their corresponding residential trajectories throw a new light on emerging phenomena of residential segregation in South-Eastern France, within the context of regional metropolisation (Fusco and Scarella 2011).

These results couldn't be obtained from more traditional multivariate analysis techniques (Hair *et al.*, 1998, Sanders 1989), as done by Aero (2006), Debousschere (2006) and Skifter Andersen (2011). Our five step Bayesian protocol can be compared more particularly to factor analysis followed by hierarchical clustering. These techniques need quantitative data structures, whereas interviews constitute a rich corpus of data with few numeric variables. Qualitative information is therefore central. Only textual statistics (Lebart *et al.*, 1998) could outmatch BN clustering in taking into account qualitative information. The BN protocol has however the advantage of a controlled reduction of model variables, of easy concept discovery through variable associations and of careful uncertainty appraisal in clustering. Coding interview transcriptions in qualitative variables we also limited the biases linked to households' statements (Carpentier 2010). This is however also a limit of our procedure. Discourse contextualization of answer items is only possible for a limited corpus of interviews. Applying the BN protocol to a more consistent database would require conducting a survey based on an appropriate questionnaire mixing closed and open questions. In-depth interviews with *a posteriori* variable coding are thus to be considered only as an exploratory phase of the research.

Generalization of analysis results is also clearly limited by interview sample size and sample method. As far as sample size is concerned, given the limited number of values of the original variables (2 to 4), samples ten times as big as the ones we used could already give a satisfactory estimation for the parameters of the Bayesian networks employed. An appropriate sampling frame would then be necessary to obtain unbiased stratification. The proposed application is thus more a prototype of Bayesian statistical modelling than a complete analysis of socio-geographic profiles within home buyers in Provence and on the French Riviera. It shows nevertheless the possibility of identifying profiles which are probabilistically quite robust, even from a very limited database. Moreover, profiles are strikingly similar in the two study areas, pointing both to structural similarities in the geographic and socio-cultural contexts of Provence and of the French Riviera, and to robust modeling. The methodology also shows the interest of a human/computer interaction in analysing qualitative data from interviews. This approach could be used as a heuristic in order to determine typical profiles for a more detailed discourse analysis (Renkema 2004) of interviews, opening the way to interdisciplinary research with sociologists.

Other potential developments can be foreseen for scenario building through probabilistic simulation in the BN. We could thus infer probable residential choices of growing households or of first time home buyers, within a given socio-geographic profile and under the assumption of short-term stability of profiles. Time series analysis could be used in order to prospect the future share of different profiles. The main limit that has to be overcome in order to develop such applications is once again the availability of large databases of interview recordings and/or detailed open answer surveys. A future perspective is thus to apply the proposed methodology to larger databases containing a geographic dimension, like household mobility surveys, as long as they cover all necessary lifestyle and constrain factors identified as emerging research issues in housing studies.

References

- Aero T., “Residential Choice from a Lifestyle Perspective”, *Housing, Theory and Society*, vol. 23, no. 2, 2006, p.109-130.
- Bayesia – Repères, “Bayesian Networks: a new tool for consumer segmentation”, *Skim Conference*, Barcelona, May 28th 2008, 31 p., <http://www.slideshare.net/f.abiven/repres-bayesia-consumer-segmentation-skim-conf08>
- Bayesia, *BayesiaLab Documentation, version 4.6*, Laval, Bayesia, 2009.
- Bootsma H. G., *The myth of reurbanization. Location dynamics of households in the Netherlands*, Amsterdam: NETHUR D Publications, 1998.
- Bourdieu P., *La distinction. Critique sociale du jugement*, Paris, les Éditions de Minuit, 1979.
- Carpentier S., “Modes d’habiter urbains et ruraux : entre continuité et rupture”, *Articulo – Journal of Urban Research*, Special Issues 3, 2010, 20 p.
- Carpentier S., “Une analyse exploratoire des liens entre mobilité quotidienne et ancrage résidentiel. Vers une approche écologique de la mobilité ?”, *Articulo – Journal of Urban Research*, no.3, 2007, 20 p.
- Clapham D., *The meaning of Housing. A Pathways Approach*, Bristol, The Policy Press, 2005.
- Da Cunha A., Bochet B., Both J.-F., Mager C., Vuagniaux L., *Mobilité résidentielle, aspirations des ménages et transformations de l’habitat : l’agglomération lausannoise, rapport de recherche*, Observatoire de la Ville et du Développement Durable, 2004.
- Debusschere E., “Segmentation of the Flemish housing market”, *ENHR Conference*, Prague, 28.6-1.7.2009.
- Decoupigny F., Fusco G., “Mobilités potentielles et émergence de structures réticulaires en région Provence-Alpes-Côte d’Azur”, *Espace géographique*, tome 38, n°3, 2009, p. 267-282.
- Donzel A., Faure L., Fusco G., Girard A., Le Dantec E., Scarella F., *Enquête sur l’accession à la propriété résidentielle dans les régions méditerranéennes françaises, rapport de recherche pour la DREAL PACA, UMR LAMES / UMR ESPACE / UPVD*, Marseille, 2011.
- Duany A., Plater-Zyberk E., Speck J., *Suburban Nation: The Rise of Sprawl and the Decline of the American Dream*, New York, North Point Press, 2000.
- Dupuy G., *La dépendance automobile. Symptômes, analyses, diagnostic, traitements*, Paris, Anthropos, 1999.
- Fusco G., Scarella F., “Looking for Socio-Geographic Profiles from Survey Data: the Case of Home ownership in the French Riviera”, *Proceedings of the AGILE’2012 International Conference on Geographic Information Science*, Avignon, April 24-27 2012, p.135 -140.
- Fusco G., Scarella F., “Recompositions territoriales en Provence-Alpes-Côte d’Azur. Analyse croisée par les mobilités quotidiennes et résidentielles”, *Cybergeo*, 2013 (forthcoming).

- Hair J. F., Black W. C., Babin B.J., Anderson R. E., *Multivariate Data Analysis*, Prentice Hall, 5th Edition, 1998.
- Jensen F.V., *Bayesian Networks and Decision Graphs*, New York, Springer, 2001.
- Kaufmann V., “La motilité : une notion clé pour revisiter l’urbain ?” in Joye D., Kaufmann V., Bassand M., *Enjeux de Sociologie urbaine*, Lausanne, PPUR, 2007, p. 87-102.
- Kaufmann V., *Mobilité quotidienne et dynamiques urbaines. La question du report modal*, Lausanne, PPUR, 2000.
- Kauko T., “Expressions of Housing Consumer Preferences: Propositions for a Research Agenda”, *Housing, Theory and Society*, vol.23, no. 2, 2006, p.92-108.
- Korb K.B., Nicholson A.E., *Bayesian Artificial Intelligence*, Boca Raton, Chapman & Hall, 2004.
- Lebart L., Salem A., Berry L., *Exploring Textual Data*, Boston, Kluwer-Academic Publisher, 1998.
- Ley D., “Alternative explanations for inner-city gentrification: a Canadian assessment”, *Annals of the Association of American Geographers*, vol. 76, 1986, p. 521-535.
- Maclennan D., “Some thoughts on the nature and purpose of house price studies”, *Urban studies*, vol. 14, 1977, p. 59-71.
- Maclennan D., Tu Y., “Economic perspectives on the structure of local housing systems”, *Housing studies*, vol.1, 1996, p. 387-406.
- Mumford L., *The City in History: Its Origins, Its Transformations, and Its Prospects*, San Diego, Harcourt Inc, 1961.
- Newman P.W.G., Kenworthy J.R., *Sustainability and cities. Overcoming automobile dependence*, Washington, Island Press, 1998.
- Pearl J., *Causality. Models, Reasoning and Inference*, Cambridge, Cambridge University Press, 2000.
- Pourret O., Naïm P., Marcot B.G. (Eds.), *Bayesian Networks: A Practical Guide to Applications*, New York, John Wiley & Sons, 2008.
- Renkema J., *Introduction to Discourse Studies*, John Benjamins, 2004.
- Rougé L., « Inégale mobilité et urbanité par défaut des périurbains modestes toulousains », *EspacesTemps.net*, 25.04.2007
- Sanders L., *L’analyse statistique des données en géographie*, Montpellier, Reclus, 1989.
- Skifter Andersen H., 2011, Explaining preferences for home surroundings and locations, *Urbani izziv*, volume 22, no. 1/2011, p. 100-114
- Stock M., “L’habiter comme pratique des lieux géographiques”, *EspacesTemps.net*, 2004, <http://www.espacestemp.net/document1353.html>.
- Wiel M., *La transition urbaine, ou le passage de la ville pédestre à la ville motorisée*, Liège, Mardaga, 1999.

VARIABLE CODE: Variable Definition, Number of Values

AGE: Age of the interviewee (3): <40, 40-60, >60	CRITERE: Priority criterion in the choice of the present dwelling (4): logistics, calm/nature, functional, other
ORIGINE: Geographical origin of the interviewee (4): local, Mediterranean, North-Atlantic, Eastern France/Central Europe	PART_DEPENSES_LGM: Portion of monthly household income devoted to housing (3): <30%, 30-50%, >50%
CSP: Socio-professional status of the interviewee (4): retiree, executive, factory/office worker or clerk, independent or other	PERIODE_ACHAT: Purchase period of the present dwelling (4): 1997-2001, 2002-2005, 2006-2007, 2008-2010
REVENUS: Household income of the interviewee (3): 10-20 K€, 20-30 K€, more than 30 K€	TYPE_ACHAT: Typology of the purchase of the present dwelling (3): real estate agent or developer, privately, other
NB_ACTIFS: Number of active people in the household (3): 0, 1, 2 or more	TEMPS_RECHERCHE: Search time for the present dwelling (3): < 1 month, 1-6 months, >6 months
SITUATION_FAMILIALE: Household typology of the interviewee (3): single, couple, couple with children	DIFFICULTES_RECH: Difficulties encountered within the search for the present dwelling (4): no difficulty, insufficient budget, no supply in the sector, other
ZONE_LGM: Metropolitan zone of the present dwelling (3): metropolitan centres, 1 st belt, 2 nd belt	COUT: Total cost for the purchase of the present dwelling including construction/renovation (3): <150 K€, 150-250 K€, >250 K€
HABITAT: Morphological characteristics of the neighbourhood of the present dwelling (4): city centre, urban periphery, suburb, rural village	TRAVAUX: % of construction/renovation costs for the present dwelling (3): no renovation, <20%, >20%
TYPE_LGM: Typology of the present dwelling (2): house, flat	CREDIT: Portion of loan financing for the purchase of the present dwelling (3): <1/3, 1/3-90%, >= 90%
PIECES: Number of rooms of the present dwelling (3)	ESPACE_QUOTIDIEN: Space frequented daily by the household (3): city-centre, neighbourhood, other
JARDIN: Presence of a private garden in the present dwelling (2): yes, no	ESPACE_LOISIRS: Space frequented regularly by the household for leisure activities (3): city-centre, neighbourhood, other
PARKING: Presence of a private parking in the present dwelling (2): yes, no	MODE_NAVETTE: Transportation mode used for daily commuting (4): car/motorbike, transit, walking/cycling, NA
PISCINE: Presence of a private swimming pool in the present dwelling (2): yes, no	MODE_LOISIRS: Transportation mode used for leisure activities (3): car/motorbike, transit, walking/cycling
PARKING_COLL: Presence of a common parking within the subdivision/condominium of the present dwelling (2): yes, no	VEHICULES: Number of motor vehicles owned by the household (4): 0, 1, 2, 3 or more
JARDIN_COLL: Presence of a common garden within the subdivision/condominium of the present dwelling (2): yes, no	TEMPS_NAVETTE: Travel time for daily commuting for the interviewee (4): <10 min, 10-30 min, >30 min, NA
PISCINE_COLL: Presence of a common swimming pool within the subdivision/condominium of the present dwelling (2): yes, no	PERCEPTION_QUARTIER: Social perception of the neighbourhood of the present dwelling by the interviewee (2): privileged, other
PERIODE_LGM: Period of construction of the present dwelling (4): before 1945, 1945-1975, 1975-present, newly constructed	APPARTENANCE: Feeling of belonging to a geographic space by the interviewee (4): local, metropolitan region, somewhere else, everywhere/nowhere
AUTRE_LGM: Ownership of another dwelling (2): yes, no	REL_VOISINAGE: Relationship of the household with the neighbours (3): no relationship/conflict, weak/cordial, mutual aid
ZONE_LGM_ANT: Metropolitan zone of the previous dwelling (4): metropolitan centres, 1 st belt, 2 nd belt, outside metropolitan area	MOBILITE_FUTURE: Project of future residential mobility (3): yes, no, don't know
TYPE_LGM_ANT: Typology of the previous dwelling (2): house, flat	TRAVAUX_FUTURS: Project of future renovation works in the present dwelling (2): yes, no
HABITAT_LGM_ANT: Morphological characteristics of the neighbourhood of the previous dwelling (4): city centre, urban periphery, suburb, rural village	ENGAGEMENT: Engagement of the household in the social life of the local community (2): yes, no
TAILLE_LGM_ANT: Size difference of previous and present dwelling (3): bigger, smaller, the same	LGM_IDEAL: Correspondence between the present dwelling and the (almost) ideal dwelling of the interviewee (2): yes, no
STATUT_LGM_ANT: Occupancy status of the previous dwelling (3): owner, tenant, free housing	
MOTIF: Main reason for the purchase of the present dwelling (4) : work, environment, surface, desire to own property or other	

Table 1. List of the 46 Variables Coded from Interview Transcriptions.