Selecting qualitative cases using sequence analysis: A mixed-method for in-depth understanding of life course trajectories

Le Roux, Guillaume; Studer, Matthias; Bringé, Arnaud; Bonvalet, Catherine

How to cite

LE ROUX, Guillaume et al. Selecting qualitative cases using sequence analysis: A mixed-method for in-depth understanding of life course trajectories. In: Advances in life course research, 2023, vol. 56, p. 100530. doi: 10.1016/j.alcr.2023.100530

This publication URL: https://archive-ouverte.unige.ch/unige:167746
Publication DOI: 10.1016/j.alcr.2023.100530

Selecting Qualitative Cases Using Sequence Analysis: A Mixed-Method for In-Depth Understanding of Life Course Trajectories

Guillaume Le Roux[a),*], Matthias Studer [b)], Arnaud Bringé [a)], Catherine Bonvalet [a)]

a) Institut national d'études démographiques — French Institute for Demographic Research (INED), 9 cours des Humanités CS 50004 93322, Aubervilliers Cedex, France.

b) University of Geneva, LIVES NCCR & Institute of Demography and Socioeconomics, 105 bd Carl-Vogt, CH-1205 Geneva, Switzerland.

* Corresponding author: e-mail address guillaume.le-roux@ined.fr

DOI: https://doi.org/10.1016/j.alcr.2023.100530

Preprint of the article published in *Advances in Life Course Research*, Volume 56, June 2023, 100530.

Abstract:

In this paper, we propose a sequence analysis-based method for selecting qualitative cases depending on quantitative results. Inspired by tools developed for cross-sectional analyses, we propose indicators suitable for longitudinal study of the life course in a holistic perspective and a set of corresponding analysis guidelines. Two complementary indicators are introduced, *marginality* and *gain*, that allows labeling observations according to both their typicality within their group and their illustrativeness of a given quantitative relationship. These indicators allow selecting a diversity of cases depending on their contributions to a quantitative relationship between trajectories and a covariate or a typology. The computation of the indicators is made available in the TraMineRextras R package.

The method and its advantages are illustrated through an original study of the relationships between residential trajectories in the Paris region and residential socialization during childhood. Using the *Biographies et Entourage* [Event history and entourage] survey and qualitative interviews conducted with a subsample of respondents, the analysis shows the contributions of the method not only to improve the understanding of statistical associations, but also to identify their limitations. Extension and generalization of the method are finally proposed to cover a wider scope of situations.

Keywords: mixed methods, sequential explanatory design, sequence analysis, case selection, residential trajectories

1 Introduction

In the social sciences, life histories are increasingly used to analyze behaviors and social processes. While sociologists, geographers and demographers may not analyze them in the same way, the dissemination of the life course perspective through various disciplines over at least three decades expresses the renewed interest in the individual actors and in the importance of time across the social sciences (Hägerstrand, 1970; Elder, 1975; Abbott, 1990; Courgeau & Lelièvre, 1991; Bernardi, Huinink, & Settersten, 2019). The life course paradigm further stresses the importance of considering past experiences; the role of the socio-economic, historical and cultural contexts in which individuals live/have lived; the interdependence of individuals; and connections between life domains (for example residential, family and occupational trajectories). The need to take so many relationships into account is highly demanding in terms of methods, and has stimulated many methodological developments.

Broadly speaking, two main methodological approaches can be distinguished. Some studies rely on qualitative methods and retrospective interviews, or longitudinal (i.e. repeated) interviews to understand changes over the life course. Other studies use quantitative methods and data, such as event-history surveys and panel data, reconstituting the trajectories of individuals in different life domains. While they are mostly used separately, these two approaches are highly complementary in life course research (Giele & Elder, 1998; Bidart & Cacciuttolo, 2013). For instance, quantitative analysis might reveal statistical regularities in the timing, succession and occurrence of some life events. However, it is often limited to understand how the meaning given to these transitions varies among different sub-populations (Heinz, 2003).

In this paper, we propose a sequence analysis-based method for selecting qualitative cases based on quantitative results, a strategy often called “case selection” (Seawright & Gerring, 2008), but also “purposive sampling” (Patton, 2002). Sequence analysis allows describing and explaining the diversity of life courses. Considering life courses as ordered sequences of states, it renders trajectories as a whole and differs from methods that focus on the occurrence of specific life events. Specifically, it allows reconstructing the overall logic of trajectories and to better account for fuzzy or complex transitions (GRAB, 2006).

Building on the work of Duvoisin (2017), our method relies on holistic and inferential approaches within the sequence analysis framework. More precisely, it allows cases to be selected according to their *representativeness* or *illustrativeness* of a given quantitative relationship between, on the one hand, trajectories and, on the other hand, a covariate or typology. This procedure has several advantages. First, it allows situating qualitative data within the whole sample, thus improving its representativeness. Second, a deeper understanding of quantitative relationships can be achieved using qualitative material. Finally, by looking at divergent cases and counterexample, the limitations of the quantitative analysis can be identified. More broadly, the proposed method extends case-selection techniques to complex data such as trajectories by providing tools to identify cases illustrative of, and counterexamples to, a given quantitative relationship. It aims to explicitly link the quantitative and qualitative analyses. For this reason, it is also useful if the qualitative material was selected using another (more traditional) case selection method.

The method and its advantages are illustrated through an original study of the relationships between residential trajectories in the Paris region and residential socialization during

childhood. Using the *Biographies et Entourage* [Event history and entourage] survey conducted in 2001 and qualitative interviews conducted the following years with a subsample of respondents, the analysis highlights the importance of localized social and family networks in the study of residential choices.

The method is suitable when qualitative data are available on a subsample of quantitative data. However, the proposed method is much more general and not restricted to this specific data configuration. It can be used to select a qualitative subsample based on a quantitative analysis (Palinkas, Horwitz, Green, et al. 2015). Second, it can also be used when the qualitative material is not a sub-sample of the quantitative data: it is often possible to go through qualitative material and systematically code the value of the quantitative variables through a process called "quantitization" (Sandelowski, Voils & Knafl, 2009). The qualitative sample is then "added" to the quantitative part before running the analysis. Third, the method is also useful without qualitative material. It allows passing from large samples to small samples, which can then be analyzed in a qualitative way by directly compiling information from quantitative data (Lieberman, 2005; Sharland, Holland, Henderson, et al., 2017) or by returning to raw survey materials (Lelièvre & Trabut, 2019). Finally, the method can be used when a large number of interviews (e.g. more than 50) are collected to facilitate and guide their analysis, which might be difficult to achieve in a purely qualitative manner.

For simplicity, we focus the presentation on the study of the link between a single explanatory covariate and trajectories. However, in the final section, we discuss the inclusion of several explanatory variables and how it can be used in conjunction with a sequence-analysis typology. The proposed indicators can be computed with the `dissindic` function of the TraMineRextras R package. The associated help page also provides a walkthrough example.

The article is organized as follows. We start by reviewing case selection methods. We then briefly introduce our illustrative application before presenting the proposed case selection procedure, which works in four major steps. First, the trajectories need to be coded as state sequences and described within the sequence analysis framework. Second, two indicators are proposed and computed on the quantitative sample. Third, we discuss the selection or labeling of cases according to the values of the proposed indicators. Fourth, we discuss how to integrate the quantitative labeling of cases in the analysis of the qualitative material. We finally illustrate the benefit of the procedure for the interpretation of the qualitative interviews by analyzing the selected cases, before discussing extensions of the methodology.

2 Mixed Methods and Case Selection

Case selection can be used as a "mixed method" to combine quantitative and qualitative approaches to add "breadth and depth of understanding and corroboration" (Johnson, Onwuegbuzie & Turner, 2007, p. 123). Within mixed methods (Tashakkori & Teddlie, 1998; Creswell & Plano Clark, 2011; Hollstein, 2014), it is mainly used in "sequential explanatory designs" (Ivankova, Creswell & Stick, 2006), where the qualitative follows the quantitative phase. The qualitative analysis is then meant to deepen the results obtained by quantitative ones.

2.1 Case Selection

In the context of sequential mixed methods designs, case selection techniques aim to sequentially link quantitative and qualitative analyses by situating qualitative cases according to quantitative results (Lieberman 2005; Seawright & Gerring 2008; Palinkas, Horwitz, Green, et al. 2015). It is often used to guide the qualitative subsample selection based on quantitative analysis (Kluge, 2001; Heinz, 2003; Palinkas, Horwitz, Green, et al. 2015). However, its uses are much more general, as it allows explicitly linking the qualitative and quantitative analyses. This link can be used to inform qualitative data interpretation of a specific case or, on the contrary, to deepen or highlight the limitations of a quantitative analysis.

Seawright & Gerring (2008) distinguish two main approaches and uses of case selection, which pursue different aims. These two approaches are not contradictory, but rather complementary, as it will be shown with our illustrative application.

First, "confirmatory" techniques select observations identified as representative, typical or influential from a statistical point of view. The aim is then to confirm, deepen or reject the quantitative conclusions by looking at qualitative data. This can further be used to assess a "causal" interpretation of the quantitative relationship. For instance, interviewing someone who stayed in the same neighborhood might help understanding why and how people do this. It might confirm the role of socialization, or indicate a potentially spurious quantitative relationship.

Second, "exploratory" case selection aims to select atypical, deviant or extreme cases to document the limitations of quantitative models and conclusions. It might typically reveal key variables omitted from the models, or clusters of cases following different logics not captured by the statistical trends. For instance, it might help to understand why some people move out of the neighborhood where they grew up, even if all factors had led us to expect them to stay.

Case selection methods can be used to purposefully select the qualitative subsample from the quantitative one according to the two above-mentioned approaches. However, it is also useful when the qualitative data has already been collected to inform and make a link with the quantitative analysis. The labeling of qualitative cases as "representative" or "deviant" may guide the interpretation of the qualitative material. For instance, knowing that a case is a statistical "counterexample," we might focus the qualitative analysis on why this individual followed an atypical path. These methods allow qualitative and quantitative interpretation to be linked when quantitative information is available for the qualitative data. This link then enables back-and-forth movements between qualitative and quantitative analyses, which may help to improve the quantitative analysis iteratively (Lieberman, 2005) or explore the diversity of a studied phenomenon.

2.2 Case Selection in Life Course Research

Various life course studies have collected quantitative and qualitative data within a single project, but most have combined the information from the two approaches in the interpretation phase using "triangulation" (Giele & Elder, 1998). We have found few contributions combining them in the analysis phase within a sequential explanatory design. These contributions can be classified according to their use of one of the two quantitative approaches to the study of life course trajectories, sequence analysis or event history analysis (Aisenbrey & Fasang, 2010; Billari, 2005; Courgeau & Lelièvre, 1991). Logically, the chosen case selection method depends on the quantitative approach.

First, sequence analysis, which is rooted in the *algorithmic* culture, takes an exploratory approach and a holistic perspective on trajectories. It is typically used to build a typology of recurrent paths, where the types are thought to describe the main mechanisms driving the trajectories. A number studies have used this quantitative approach in conjunction with qualitative interviews. Latcheva and Herzog-Punzenberger (2011) and Verd and Andreu (2011) used a typology established using sequence analysis to select respondents for the qualitative phase. Similarly, Abbott (1995), Heinz (2003), and Remillon and Lelièvre (2018) selected qualitative interviews to illustrate the life course patterns identified by the typology and further explore the mechanisms driving these trajectory types. However, in such approaches the selection is usually made by focusing on the most central observations from each type (i.e. the medoid). This approach has two main limitations. First, by using it, we tend to ignore trajectories lying at the margins of the typology. This tends to understate the diversity and complexity of the trajectories associated to each ideal-typical trajectory, which is crucial to evaluating the quality of the typology (Piccaretta & Studer, 2019). Second, this approach does not allow cases to be selected according to the relationship between trajectories and key covariates such as gender or cohort. For instance, in our application, we are interested in selecting cases that illustrate the relationships between socialization and future residential trajectories.

Event history analysis, in contrast, is rooted in *statistical modeling* culture, and takes an explanatory perspective, aiming to estimate the relationships between covariates and the occurrence of events or transitions within the life course. In this perspective, case selection is generally used to gain a better understanding of quantitative relationships and to offer a qualitative description of the underlying processes. The most common strategy is to stratify cases according to individual characteristics or life events[1] known to be associated with specific outcomes in the life course (such as gender, level of education, or occupation). Next, random sampling is conducted within each stratum to ensure maximum variation of cases (Heinz, Kelle, Witzel & Zinn, 1998; Weymann, 1999; Elliott, Gale, Kuh & Parsons, 2011; Østergaard & Thomson, 2020). [2] Selected cases are then explored to study a specific quantitative association between an event, an outcome or a transition and individual characteristics, through the in-depth study of "on-the-line" cases (Lieberman, 2005), but also outliers (Portes & Fernández-Kelly, 2008). Here, case selection strategies are based on single events or transitions and not on the entire trajectories. However, the life course paradigm stresses the importance of situating events and outcomes within the whole trajectory. Furthermore, qualitative cases are generally not labeled using the quantitative results, which might provide further information when analyzing the qualitative cases, as we will illustrate below.

In this paper, we propose a new case selection method based on sequence analysis that can be used in either an explanatory or exploratory perspective. It allows cases to be selected according to their statistical illustrativeness or representativeness for a given quantitative relationship between trajectories and a covariate or a typology.

[1] Some authors also proposed to use events (Østergaard & Thomson, 2020) or a typology of trajectories (Legewie & Tucci, 2021) to select case according to their association with a specific outcome in later life.

[2] Readers interested in the criticisms of the use of case selection methods in qualitative research are referred to Small (2009). However, most of these criticisms do not apply to our context, where the mixed method is driven by quantitative analysis.

3 Illustrative Application

We demonstrate the usefulness of the proposed method with an original study on residential trajectories within the Paris region. We focus on the link between these trajectories and individuals' place of socialization during childhood.

During the second half of the 20th century, Paris underwent a profound urban transformation that increased urban segregation. Part of the working and middle classes moved to the suburbs, a new migrant working class arrived, and the city center started its gentrification. The study of residential trajectories offers a key way to understand these changes and to highlight inequalities in access to housing and urban resources (Bonvalet & Brun, 2002; Dureau, Dupont, Lelièvre, Lévy & Lulle, 2000; Clark, 2009; Ohnmacht, Maksim & Bergman, 2009).

How and why people move is not a new topic (Rossi, 1955), but the phenomena involved are complex and might be linked to the degradation of the residential environment, homeownership desire, the quest for comfort, the affirmation of a new social status, and the choice of a new lifestyle (Clark & Dieleman, 1996; Bonvalet & Dureau, 2000; Bonvalet & Brun, 2002; Authier, Bonvalet & Lévy, 2010). The life course perspective deepened these results by integrating past experiences and the interactions between different life domains (residential, family, professional, social) (Courgeau & Lelièvre, 1991; Bonvalet & Lelièvre, 2016). Qualitative research also highlighted the agency of individuals and households, and the diversity of strategies they may employ to become a homeowner, to stay in the neighborhood where they live, or to affirm a social position (Bonvalet & Fribourg, 1990).

In this article, we focus on the effect of individuals' place of socialization during childhood on their adult trajectory, which is little studied in the quantitative literature. Following the life course perspective, we explore the hypothesis that past place experiences structure subsequent trajectories, as suggested by qualitative studies (Bonvalet & Gotman, 1993). Places where people grew up might become reference places tied to residential choices during adulthood. We aim to show that, by combining qualitative and quantitative approaches, the relationship between residential socialization during childhood and subsequent residential choices can be better assessed and understood, and that the procedure highlights a wider diversity of social mechanisms.

To explore this hypothesis, we rely on the quantitative and qualitative data from the *Biographies et Entourage* survey conducted by INED (2001 for the quantitative component, 2002-2004 for the qualitative component) (Bonvalet & Lelièvre, 2016). Quantitative data was collected from 2,830 respondents from the 1930–1950 birth cohorts living in the Paris region (Île-de-France). Information on geographical, residential, occupational and family trajectories beginning at birth was collected retrospectively. 141 in-depth interviews were conducted with respondents to the quantitative survey in 2002 and 2004. This subsample was not constructed using a specific case selection method.

4 Case Selection Using Sequence Analysis

We aim to extend case selection techniques to the study of the link between trajectories and a categorical covariate. The latter can be a sequence analysis typology, in an exploratory

perspective, or an explanatory covariate, such as the place of socialization, as in our illustrative application.

The procedure works in four steps, which are developed in this section using our illustrative application as an example. First, we rely on sequence analysis to quantitatively study the relationships between a covariate and the trajectories coded as state sequences. Second, we propose two indicators that allows situating qualitative cases according to the quantitative analysis. These indicators allow identifying typical or atypical cases, but also cases that are illustrative of or discordant with a quantitative relationship. Third, these two indicators can be used to select a subsample for qualitative analysis or to label qualitative cases if they were previously collected as in our illustrative application. Finally, the previous information is used for the qualitative analysis of the selected cases.

4.1 Sequence Analysis

The first step is to quantitatively study the relationships between trajectories and our covariate using sequence analysis. Regularly identified as one of the key approaches for life-course analysis (Shanahan, 2000; Mayer, 2009; Piccarreta & Studer, 2019), sequence analysis aims to analyze trajectories described as a sequence of categorical states in a holistic perspective. It regroups methods ranging from visualization to explanatory methods and provides a comprehensive overview of the observed trajectories taken as a whole. The first step of any sequence analysis is to code trajectories as state sequences. This implies specifying the state occupied by an individual at each time point.

4.1.1. Describing Trajectories

Based on the retrospective survey, we focus on residential trajectories between ages 20 and 49. This age range corresponds to the rise of homeownership's aspirations and family formation, which are key aspects of residential choices (Bonvalet & Brun, 2002). Residential trajectories are multidimensional, including housing location, tenure (ownership or rental), type of dwelling (house, apartment) and type of neighborhood. In this article, we focus on the two main competitive dimensions: location and housing tenure. These two aspects are the result of tradeoffs constrained by household resources and housing supply configurations. We therefore code the trajectories using two parallel state sequences for location and housing tenure, which is made possible with multichannel sequence analysis (Gauthier, Widmer, Bucher & Notredame, 2010).

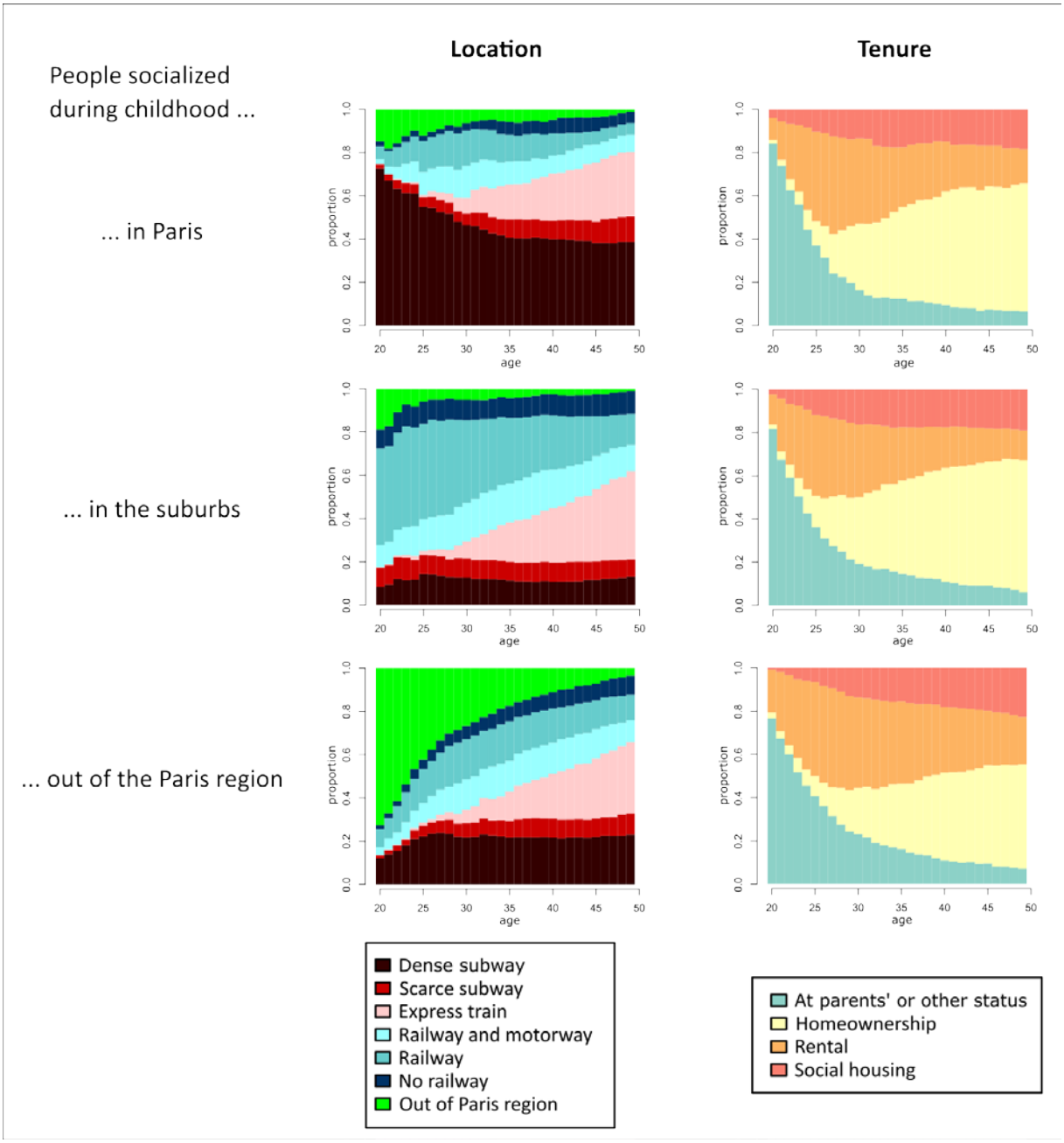
The first sequence codes the location trajectory by specifying for each year between ages 20 and 49 the location of an individual at that time. The locations are coded according to the accessibility level of the municipalities of residence, defined by the level of transport facilities at the time respondents lived in the municipality using seven categories (Le Roux, Imbert, Bringé, & Bonvalet, 2020). This is not only a key factor in location choices, but also reflects the relationship of individuals to centrality in a city undergoing constant transformation. Indeed, it determines a relative location as the city expands and transport networks develop: peripheral municipalities may become relatively central as urban sprawl proceeds. The most accessible

locations feature a dense distribution of subway stations, generally the Parisian *arrondissements* and some well-connected municipalities of the inner suburbs. The other categories distinguish different suburbs types depending on estimated travel times to the city center. Second, the tenure trajectories are coded using four categories in the same age range: ownership, rental (private market), social housing, and being hosted by family or other types of tenure.

Sequences analysis provides several visualization methods for states sequences. Figure 1 presents the chronograms of the two parallel sequences. A chronogram presents the distribution of the states at each time points, and allows identifying the most frequent situations at each age. We used a separate chronogram for each place of socialization to highlight the main difference between socialization places. The latter is measured by looking at the longest time spent before age 20 in the following three locations, which reflect the major distinctions in urban context, lifestyles and representations (Pinçon & Pinçon-Charlot, 2008): within Paris, in the suburbs, and outside the Paris region.

According to Figure 1, individuals socialized in Paris have the highest proportion in the most central locations, regardless of their age, even if many of them leave the city center. Those socialized in the suburbs tend to stay out of the most central locations, but many benefit from the expansion of the express train network. They tend to access homeownership more frequently than individuals socialized outside the Paris region, and even more so after 30 years. Finally, those socialized outside the Paris region occupy an intermediate position in terms of residential location. These first results show that there is an association between place of socialization and the following residential locations.

Figure 1: Chronograms of location and tenure trajectories between the ages of 20 and 49 according to the place of socialization during childhood



Source: Biographies et Entourage survey (INED, 2001)

4.1.2. Discrepancy Analysis

The chronograms presented in Figure 1 allow describing the association between trajectories and a categorical covariate. The discrepancy analysis framework provides methods to study the strength and statistical significance of the relationship (Studer, Ritschard, Gabadinho, & Müller, 2011). The method proposed in this article relies on this framework. We therefore start with a short reminder, and we refer interested readers to Studer, Ritschard, Gabadinho and Müller (2011) for the full presentation.

Conceptually, this framework is based on the study of the discrepancy between (or variation among) the trajectories. This discrepancy is measured using a distance measure. This measure quantifies the dissimilarity between trajectories, according to the timing, the time spent in each

state, and the sequencing of the states. Several distance measures are available, and the choice among them should be grounded on the research questions. Studer and Ritschard (2016) provide a detailed discussion and guidelines for this choice. In our illustrative application, we used multichannel distance based on optimal matching (OM) with constant costs. Multichannel distances allow jointly studying location and tenure trajectories (Pollock, 2007; Gauthier, Widmer, Bucher, & Notredame, 2010).

Conceptually, the underlying idea of discrepancy analysis is that high average distances are associated with large discrepancies (variations) among the sequences, while low average distances are linked to lesser discrepancy between trajectories. Extending the ANOVA (analysis of variance) framework, the method then computes the share of the discrepancy between the sequences that is explained by a covariate. The statistical significance of the relationship is estimated using permutation tests. In our application, the place of socialization during childhood explains 2.9% of the variation of the parallel residential trajectories, and the relationship is statistically significant.[3]

We now turn to a brief presentation of the mathematical details of the framework, which are used to develop our indicators. The ANOVA is based on the analysis of the sum of squares. The total sum of squares, SS_T can be decomposed into within (SS_W) and between-group (SS_B) sums of squares, leading to the following relationship.

$$SS_T = SS_B + SS_W$$

These sums of squares can then be used to compute a R^2 measuring the share of the variation among the sequences explained by a covariate. Importantly, the sum of squares can be computed using the squared Euclidean distance between observations. By replacing the squared Euclidean distance by another distance measure relevant for sequences, such as optimal matching, we can generalize the ANOVA framework to the sequence analysis case.

These pseudo-sums of squares can then be computed as follows:

$$\begin{aligned} SS_T &= tr(\mathbf{G}) \\ SS_B &= tr(\mathbf{HG}) \quad (1) \\ SS_W &= tr[(\mathbf{I}-\mathbf{H})\mathbf{G}] \end{aligned}$$

Where $\mathbf{G}$ is the "Gower matrix," defined as $\mathbf{G} = -\frac{1}{2}(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}')\mathbf{D}(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}')$, where$\mathbf{1}$ is a vector of ones of length n. Conceptually, it is a transformation of the $n \cdot n$ matrix of distances $\mathbf{D}$ required for the computations. $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is the idempotent "hat matrix" used in linear regression based on $\mathbf{X}$ the $n \times m$ matrix with the values of m covariates using contrasts for coding factors and including a first column of ones for the intercept. Conceptually, it summarizes the information of the covariates. $\mathbf{I}$ is the identity matrix. Readers interested in the mathematical details are referred to McArdle & Anderson (2001). We further discuss the interpretation of these matrices in the next section.

[3] This percentage might seem low, but this is generally the case. Indeed, trajectories are complex objects measured using many variables, and the share of diversity explained by one variable is thus usually low. See Liao & Fasang (2020) for a thorough discussion of the issue.

4.2 Case Selection Indicators for Sequence Analysis

The selection of cases is often based on an indicator highlighting the relevant cases (see Seawright and Gerring 2008 for a review). Several indicators have been proposed, which work as follows. First, a value is computed for each observation. Second, a selection is made according to the value of this indicator. Here, we propose two new indicators relevant to study trajectories within the sequence analysis framework. First, the *marginality* index aims to distinguish typical and deviant cases. Second, the *gain* indicator aims to identify illustrative cases and counterexamples of a quantitative relationship.

4.2.1 Marginality

When a case study follows a regression or a bivariate analysis, *typical* or *deviant* cases can be selected using regression residuals (Seawright and Gerring, 2008). Observations with a low residual are then considered as typical of the studied cross-case relationship and representative of the population if the model is correctly specified. These observations can be used to deepen our understanding of whether and how the statistical cross-case relationship operates within each observation. It is therefore mainly used in a *confirmatory* perspective.

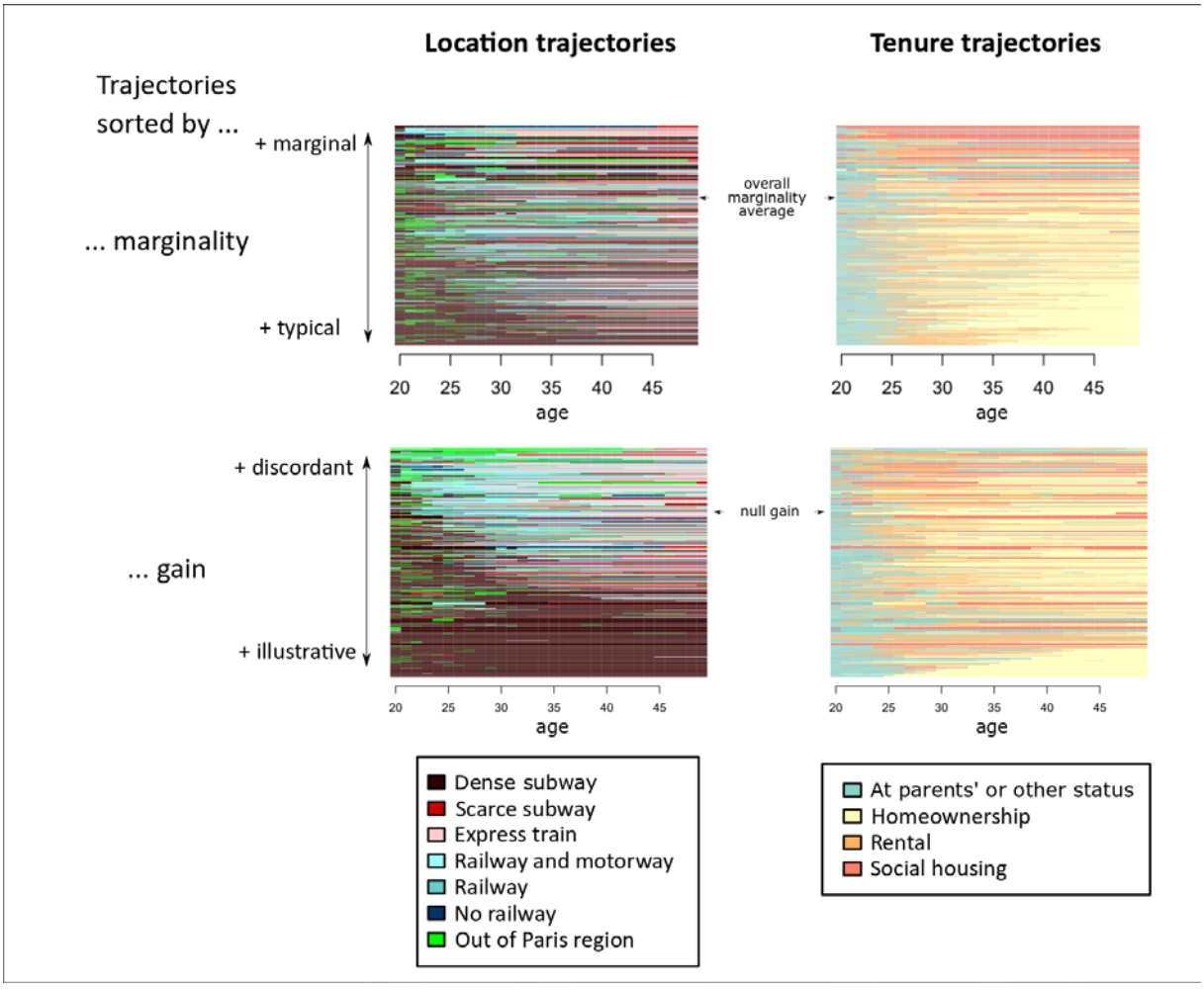
Similarly, one might select deviant cases using high regression residuals. These observations can illustrate the diversity that is not captured by the statistical model. As such, they are mainly useful for *exploratory* purposes. As pointed out by Seawright and Gerring (2008), these observations are, by definition, not representative of the association captured by the model. They should therefore not be interpreted as such when analyzing the qualitative material. Deviant cases can shed light on forgotten covariates, model misspecification, or limitations of the quantitative association. They allow the variation of the selected cases to be maximized. They may also bring out new research questions to be tested quantitatively.

We propose to use the same approach for sequence analysis. In the discrepancy analysis framework (see equation 1), the diagonal elements of the matrix $(\mathbf{I} - \mathbf{H})\mathbf{G}$ can be interpreted as the contribution of each sequence to the within sum of squares. In a linear regression, these values are the square of the residuals. They are therefore directly linked to the concept of a contribution to the residual discrepancy (Studer, Ritschard, Gabadinho, & Müller, 2011). They can also be interpreted as a measure of distance between a given sequence and the center of gravity of its groups (Batagelj, 1988; Studer, Ritschard, Gabadinho, & Müller, 2011). Conceptually, they are therefore a kind of *residual* of the discrepancy analysis framework.

To distinguish this indicator from the residuals of a regression analysis, we call it *marginality*. It measures the typicality of cases within each group of the explanatory covariate. Theoretically, the values of the *marginality* index range between zero and the maximum possible distance value. However, negative values are possible if the dissimilarity measure does not respect the triangle inequality (see Batagelj, 1988; Studer, Ritschard, Gabadinho, & Müller, 2011). Low *marginality* values indicate cases close to the group center, while high values indicate cases far from it.

Figure 2 presents index plots of the location and tenure trajectories of people socialized in Paris.[4] In these plots, each trajectory is represented using a thin horizontal line. At the top of Figure 2, the sequences are ordered according to their *marginality*, that are computed for both channels simultaneously using the multichannel distances. The sequences at the bottom of these plots have the lowest *marginality* and are thus the most typical of their group. In Paris, these trajectories are characterized by moving from their parents' home, to private rental, before homeownership, while generally staying in a central location. Lower *marginality* is mainly determined by the tenure sequence and its most frequent states.

Figure 2: Trajectories of people socialized in Paris, sorted by marginality and gain



Source: Biographies et Entourage survey (INED, 2001)

4.2.2 Gain

Marginality allows us to identify cases that are the closest to the "usual" or most common trajectory within each group. However, these cases are not necessarily illustrative of the relationship between the sequences and the studied covariate. We therefore propose a second

[4] The paper focuses on the category "people socialized in Paris." For results on the other categories, see ***anonymized***

indicator, named *gain*, which aims to select cases that are either illustrative of, or discordant with, a quantitative association.

The diagonal elements of the Gower matrix $\mathbf{G}$ can be interpreted as the residuals of the null model—i.e. when no covariates are included in the model (see Batagelj, 1988; Studer, Ritschard, Gabadinho, & Müller, 2011). A high value means that a sequence is far from the overall center of gravity of the entire sample, i.e. the most typical situation. A low value indicates a sequence close to the overall center of gravity. Extending our previous discussion on *marginality*, these values are therefore a kind of residual of the null model, i.e. the model without any covariates.

By combining the residual of the null model and the *marginality* (the residual of the "fitted" model), we can identify sequences that are better represented when using the model than without it. These are the diagonal elements of the matrix $\mathbf{G} - (\mathbf{I} - \mathbf{H})\mathbf{G}=\mathbf{HG}$. Logically, according to Eq. (1), these values are also the sequences' contributions to the between sums of squares, a concept directly linked to the explained discrepancy. We call these contributions the *gain* of a sequence, as it measures the statistical gain of information for each case when taking the covariate into account.

The *gain* can be either positive or negative, but the average value should, by construction, be greater than zero for any significant quantitative relationship. A positive *gain* value means that the corresponding sequence is better represented when the explanatory factor is taken into account. The corresponding cases can therefore illustrate the quantitative relationship, as we gained statistical information on these cases with the covariate. In contrast, a negative value means that the sequence is less well represented when using the covariate. These sequences are therefore discordant with the studied quantitative relationships.

The index plots at the bottom of Figure 2 are ordered according to the *gain* of each (multichannel) trajectory. The sequences with high *gain*, which are the most illustrative of the quantitative relationship, are located at the bottom of each plot. We observe a clear trend in the location trajectories, with individuals socialized in Paris tending to stay there. This trend is more clearly highlighted than with the *marginality* indicator.[5] The tenure trajectory patterns are less clear. While individuals still tend to reach homeownership relatively quickly as with the *marginality index*, we also observe other kinds of tenure trajectories, including some in social housing.

4.2.3 Case Selection Using Gain and Marginality

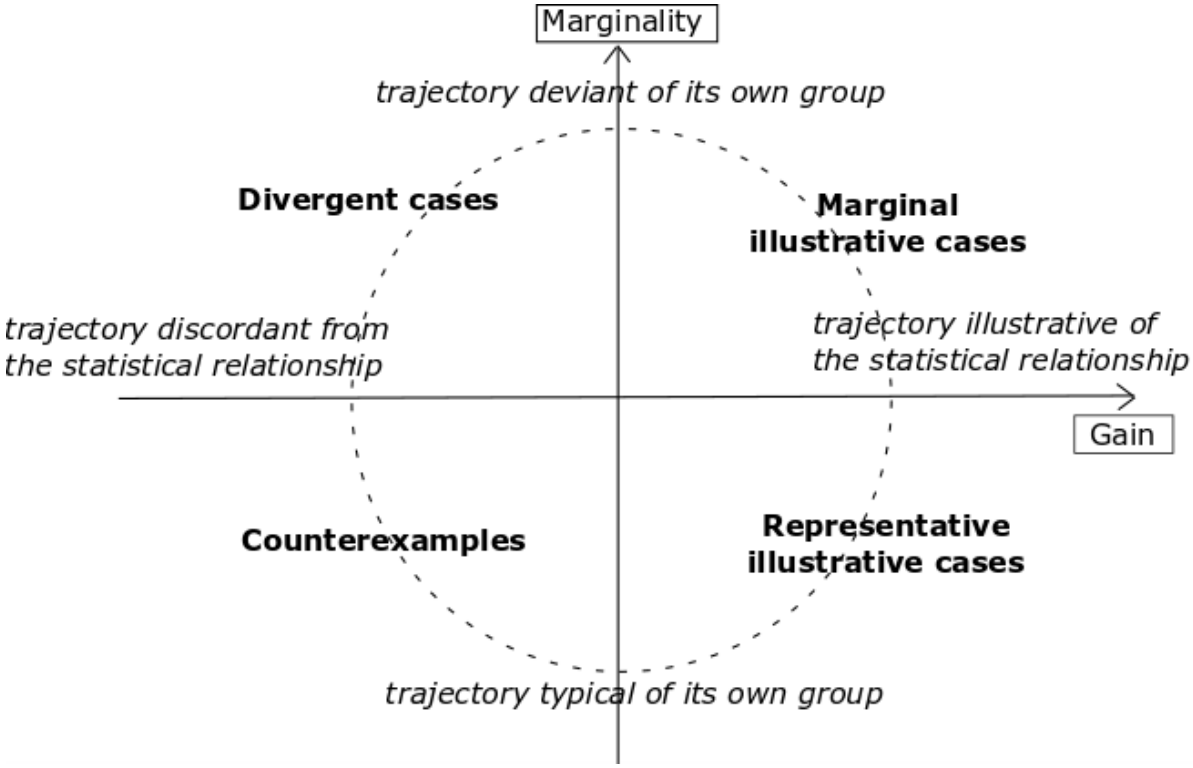
The third step of the proposed framework is to select cases based on the value of the computed indicators. The *gain* and *marginality* provide each distinctive and useful information. We therefore propose to use them both to select relevant cases for qualitative analysis.

We can distinguish four configurations of these two indicators that can be represented schematically using a scatterplot as illustrated in Figure 3. In this figure, each quadrant indicates a different combination of the two indicators. It is constructed as follows. The x-axis represents the *gain*. Here, a clear division can be drawn between positive and negative values,

[5] Similarly, those who spent their childhood in the suburbs tend to stay away from the most accessible or central locations (***anonymized***)

corresponding respectively to observations illustrative of, and discordant with, the quantitative association. The vertical bar is therefore plotted at zero. The y-axis represents the *marginality*. The *marginality* index lacks a clear theoretically driven threshold value. We are mainly interested in the value of a case relative to the others. For this reason, the horizontal line is drawn at the average value of the entire sample. Therefore, this horizontal line should not be considered as clear threshold but rather as a marker within a continuum.

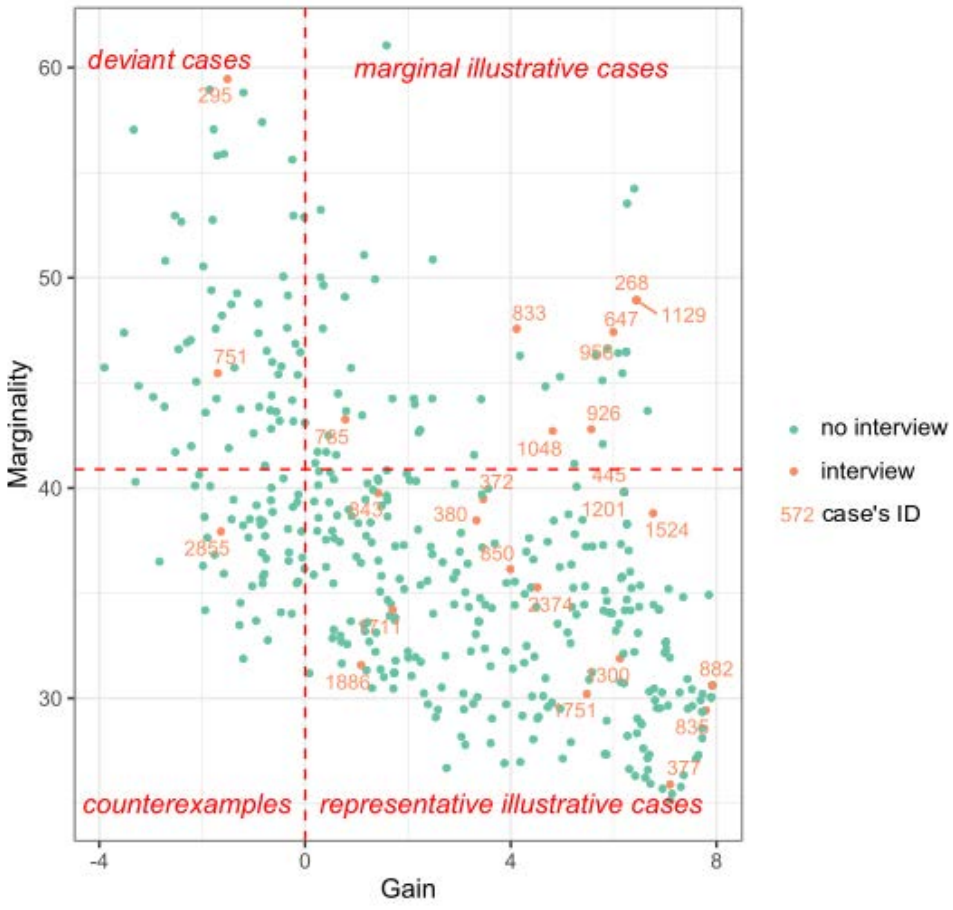
Figure 3: Interpretation of observations by the combination of marginality and gain (discrepancy analysis framework)



Depending on the objectives of the qualitative analysis, cases in each of these quadrants can be of interest and selected for qualitative analysis. We now go through each quadrant before discussing their utility in mixed methods. Table 1 summarizes this discussion. In the next section, we illustrate more practically how these quadrants can be used to select cases, and how they can be linked to qualitative analysis. The lower-right quadrant groups the "representative illustrative cases." These cases have low *marginality* (i.e. residuals) and are thus well represented by the "average" trajectory of their group. They also have high *gain*, which means that the quantitative analysis provided useful information to describe their trajectories. These cases are therefore both typical of their group and illustrative of the quantitative relationship. The cases at the bottom right end of the quadrant are the most typical and illustrative and, as such, they are particularly interesting to select because of their clear-cut position. A qualitative study of these cases is therefore useful to *confirm* the logic of the quantitative association. As pointed out by Seawright and Gerring (2008), they allow qualitatively analyzing how the (usually) cross-case quantitative relationship takes place *within a given case*. Such qualitative analyses might deepen our understanding of how an explanatory factor influences the analyzed trajectories. This quadrant is therefore useful from a confirmatory perspective. Figure 4 presents the scatterplot of the two indicators for our illustrative study, focusing on respondents socialized in Paris to keep the presentation simple. Logically, as the association is statistically significant,

the “representative illustrative” cases are the most common. This should generally be the case if the relationship is strong enough. As seen with the index plots ordered by *marginality* and *gain* (Figure 2), these cases showed “common” homeownership trajectories staying in the most central areas. Qualitative analysis could be used to confirm (or not) the relationship between growing up in Paris and staying in the most accessible locations. The aim of the qualitative analysis is then to understand the social processes underlying this relationship.

Figure 4: Distribution of observations and interviewees according to gain and marginality indicators for people socialized in Paris



Source: Biographies et Entourage survey (INED, 2001)
Note: The dotted lines divide the scatterplot in quadrants as defined above and in figure 3. The vertical line is dra wn at zero and the horizontal line at the overall marginality average.

The upper-right quadrant groups the “marginal illustrative cases.” These cases have high residuals, and thus have uncommon or atypical trajectories. They should therefore not be interpreted as representative of the population (Seawright and Gerring, 2008). However, at the same time, the included explanatory factor has proven useful in understanding their trajectories. These cases can thus illustrate how the logic of the studied quantitative relationship operates for atypical cases, and shed light on the diversity of mechanisms at work. They are useful for *exploring* the diversity of the association. In this sense, these cases can provide an intermediate perspective between the confirmatory and the exploratory approaches. Qualitative analysis can then be used to complement the analysis of representative illustrative cases. It could confirm

the mechanism highlighted for representative illustrative cases or highlight other dimensions at work in the studied relationship.

The lower-left quadrant groups the "counterexamples." These trajectories are discordant with the quantitative association, but are nevertheless close to the central trajectory, and therefore typical of their group. A qualitative analysis of these cases can thus shed light on the limitations of the quantitative findings, by suggesting possible alternative mechanisms behind divergent trajectories, and by nuancing the quantitative association. The aim here is again intermediate between the confirmatory and the exploratory. These cases can be used to explore the limitations of the quantitative relationship and potentially invalidate it. Particular attention can be given to covariates that might be missing in the quantitative analysis. Qualitative analysis here could focus on why these cases are discordant with the quantitative association. Such analysis might identify other competing factors, but also the ambiguity or non-univocal nature of the factors highlighted by the illustrative cases. In this sense, these cases may refine the study of the quantitative relationship or point to the limitations of the model. Finally, the upper-left quadrant groups the "divergent cases." These cases are atypical trajectories that are also counterexamples of the quantitative analysis. As such, they are mainly useful for exploring mechanisms ignored by the statistical model from a highly exploratory perspective. Since these cases are highly atypical, they should not be considered as representative of the population (Seawright and Gerring, 2008). A qualitative analysis of these cases could focus on why they are not well described by the statistical association.

Table 1: Guidelines for qualitative analysis by quadrant

Quadrant	**Approach**	**Focus of the qualitative analysis**
Representative illustrative	Confirmatory	Understand the inner logic driving the quantitative association. Confirm the (usually cross-case) quantitative associations and deepen the interpretation based on within-case qualitative analysis.
Marginal illustrative	Confirmatory and exploratory	Understand the inner logic driving the quantitative association and how it operates in a diversity of contexts and cases.
Counterexamples	Confirmatory and exploratory	Understand why the association is not found for some cases and explore possible competing factors
Divergent	Exploratory	Explore the limitations of the quantitative association, such as non-included factors.

The first step for case selection is to choose the relevant quadrants. If one is primarily looking to confirm an association, the "marginal illustrative" and "representative illustrative" are the most interesting. The "counterexamples" one is also useful for documenting the limitations of the quantitative analysis. However, exploratory studies might also be interested in studying the cases in the last quadrant. Once the relevant quadrants have been chosen, the cases themselves can be selected from each quadrant using Figure 4. It is recommended to include at least a few cases at the extreme of the *gain* axis, which provide more information about the quantitative relationship. If the qualitative material has already been collected, as in our illustrative study, the quadrants can be used to label each qualitative case. This labeling can then be used to deepen the interpretation and limitations of the association using the qualitative cases as illustrated in the next section. Such labeling might also be useful to guide the analysis when a large number of qualitative interview is available.

5 Qualitative Analysis of Selected Cases

Once the cases have been selected or labeled, the qualitative material can be analyzed. In this section, we illustrate this procedure using our sample application. To keep our presentation simple, we focus on individuals socialized in Paris. Nevertheless, qualitative cases socialized in the suburbs or outside Paris were analyzed using the same procedure and we report the main conclusions. All the interviews were studied. However, we selected one qualitative case for each quadrant to illustrate the procedure and highlight the identified mechanisms. These cases were selected according to the richness of the interview and the values of the proposed indicators.

5.1 Representative Illustrative Cases

Representative illustrative cases of people socialized in Paris are trajectories in central locations accessing homeownership. The qualitative interviews located at the bottom right of the quadrant of Figure 4 (no 882, 835, 377, 1300, 1751: see Fig. 4) are of special interest as they are the "most representative and illustrative." They all highlight the role of socialization notably through the attachment to a childhood neighborhood and urban lifestyles, but above all the support of family networks anchored in those neighborhoods. The case of Josiane (case 377, see inset), a psychologist, socialized in a Parisian neighborhood, illustrates this finding. Josiane's trajectory was driven by family and characterized by its geographical stability. Practically no location choice was involved. She stayed in the neighborhood where she was born and grew up, and where her family is located. Many cases reveal similar patrimonial and locational family strategies, as observed for people socialized in the suburbs too.

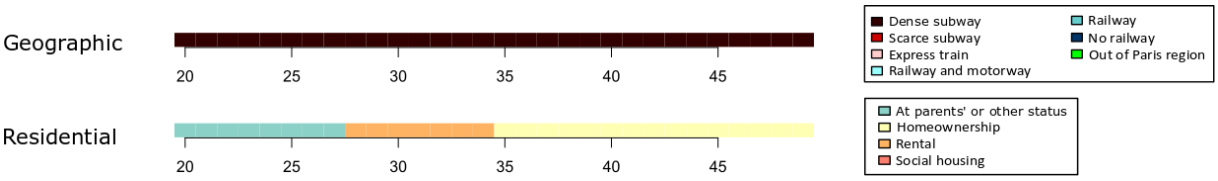
Nevertheless, interviewees in this group expressed varying degrees of agency about remaining in central location: while some cases seemed to be passive, speaking of having "no choice" or "self-evidence," other cases reflect more thoughtful choices and arbitrations, through compromises on the neighborhood or the size of the dwelling.

Individuals socialized in Paris further reject the suburbs, even if they never lived there. Interestingly, this phenomenon is also observed in mirror among individuals socialized in the suburbs.

Paul's (no. 882) representation of the suburbs illustrates this:

“Paris is not a choice. Paris is where I’ve lived, where I was born, so you have all your roots in Paris and you don’t plan to [move to the suburbs]. When a child is born in the country, it’s hard to live in the city. When you’re born in the city, you have difficulty living in the country.”

Case 377: Josiane, representative illustrative case of people socialized in Paris



Josiane, a psychologist, was born in 1935 and always lived in Paris except during the first years of the war, which she spent in Lyon with her grandmother. She describes the 15th arrondissement of Paris as a veritable family fiefdom: her grandparents already lived there, as her parents and sister. This “local” family lives in a small area within the capital, helps and sees each other more than once a week. Having left her parents’ home late, at the age of 27, she moved into the studio that her grandfather had rented. She stayed there for five years. Josiane, who wanted to be able to receive visitors and to expand, decided to buy with the support of her parents. Her father, a company director, had regretted not being able to buy an apartment in the 1930s, since co-ownership was rare at the time. “They were almost the ones who pushed me, because I was a bit panic-stricken at the time about what it meant financially.” In 1970, she bought a large one-bedroom apartment off plan, right next door to her parents’ house. She never thought of leaving the neighborhood she loved, because her family was nearby, and because it was a quiet neighborhood with shops and transportation, which she describes as a family neighborhood. Two years later, she met her husband, who came to live in the apartment. In 1975, they decided to buy a two-bedroom apartment together in the same neighborhood.

Note: Josiane’s interview was explored by looking for all the factors and resources that could explain her residential choices in central locations, while having a classic homeownership trajectory.

The qualitative exploration of representative illustrative cases provides better and deeper understanding of the main mechanisms at work. The quantitative framework guided the analysis, revealing the mechanisms linking individuals’ place of socialization during childhood and their subsequent residential choices. Among other factors, it highlights the key role of the individual’s family and social network as a resource.

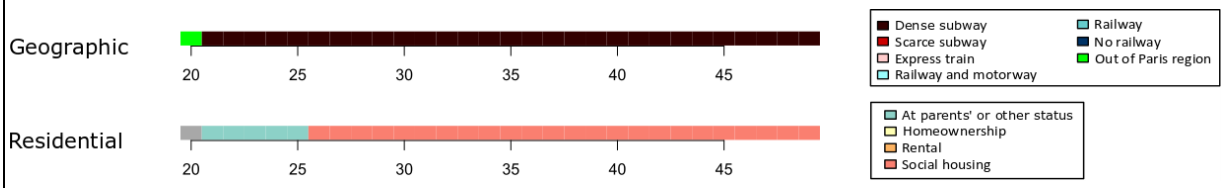
5.2 Marginal Illustrative Cases

Among the marginal illustrative cases of respondents socialized in Paris, we focused on those having high *gain* values (no. 1129, 268, 647, 956, 926, 1048, 1201, 445, 1524, 833, Fig. 6) as they are illustrative of the quantitative relationship even if their trajectories are less typical (relatively low *marginality*). They turn out to be trajectories anchored in central locations through social housing. Among cases with high *gain* and *marginality*, Serge (Case 956) who has always lived in the same neighborhood has a very informative life history. He never tried to become a homeowner, nor to leave Paris. He rather relied on a family strategy to stay in his neighborhoods. His family maintained a minimal occupancy to keep a social housing and exchange dwellings between family members according to the evolution of household sizes. These strategies were identified among several qualitative cases of this quadrant.

In all the marginal illustrative cases explored through interviews, respondents' desire to stay in the neighborhood where they grew up or in the most central locations won over their desire to be a homeowner. The presence of social housing supply in Parisian neighborhoods thus enabled a portion of working-class households to maintain themselves in their childhood places through strategies relying on local social and family networks. Interviewees evoked their attachment to their neighborhoods. However, some cases also felt trapped in social housing.

Generally speaking, the marginal illustrative cases are individuals with few financial resources and low socio-professional status. They are marginal in the sense that they resisted the overall centrifugal trend that would otherwise have carried them, like most of the working classes, out to the suburbs.

Case 956: Serge, marginal illustrative case of people socialized in Paris



Born in 1942, Serge was the youngest of a family of six children. "I have never left my neighborhood for 60 years, I was born in the yard next door." His parents, who lived in a social housing in the 20th arrondissement, were forced to leave Paris when he was 6–7 years old. However, to make sure they kept their Parisian home, they gave it to Serge's older sister, their daughter. "So it was the daughter who kept the apartment, when we came back, we got the apartment back." In 1958, Serge's father, the owner of a bar, died and the family returned to the apartment, which the eldest daughter then vacated. When Serge married in 1963 to a neighbor in the same building, his mother, who was left alone after her last son's departure, offered to keep the apartment. She said, "Well, listen my son, I'm all alone now, here's what I'm going to do, I'll leave you the apartment and then I'll find a little something." But very quickly, with the birth of two children, the flat became too small. Once again, the solution came from the family: this time, in-laws who lived in the apartment across the street. "When my in-laws retired, they said to me: 'Listen, Serge, you're going to do something, now with your two kids, you're going to take over this one. There's still one more room, one more bedroom.' We made a change of contract." At the time, the exchange of housing units was a very common practice in both social and private housing stock. Serge never wanted to move, even to become a homeowner, which would have forced him to change neighborhoods or even leave Paris. "A long time ago if we wanted a house, we would have bought a house. But I didn't want to leave Paris, I didn't want to leave my neighborhood, I didn't want to buy."

Note: In Serge's interview, we looked at how his residential socialization during childhood could explain his following trajectory in central locations. Being marginal due to his residential trajectory in social housing, we also looked for the specific role of this trajectory in his location choices and compared it with processes brought to light in the representative illustrative cases.

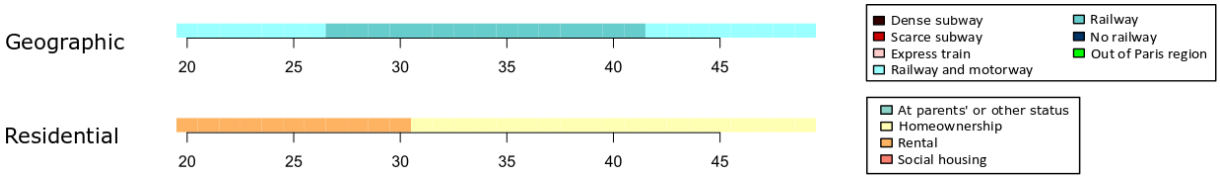
Marginal illustrative cases thus add value to the analysis by accounting for the diversity of mechanisms at work behind particular quantitative associations. Often hidden by major trends, these cases reveal a wider range of behaviors—as for Serge, the "choices" of renting and the family strategies to stay in the childhood neighborhood. All the qualitative cases in this quadrant were from the working or the lower-middle class and living in social housing. Such regularity suggests adding the social class to the quantitative analysis.

5.3 Counterexamples

As a reminder, the cases in the "counterexamples" quadrant do not follow the logic of the quantitative association.

As people socialized in Paris during childhood tended to stay in Paris, counterexamples are people raised in Paris who then settled in the suburbs. Here, we only have one interview. The case of Laurence (Case 2855) is nevertheless enlightening. It shows that residential socialization can forge counter-models. She grew up in an overcrowded Parisian apartment during the post-war housing crisis, and built a negative representation of Paris. She then left Paris very early to live in the suburbs, looking for space and lower density.

Case 2855: Laurence, counterexample of people socialized in Paris



Laurence was born in 1949 in Paris, the second of five children. Until the age of 12, she lived with her parents and siblings in a single room with only one window. It was a caretaker's lodge [her mother's work] in a building near the Montparnasse train station. Despite the good memories of Paris, she was marked by the tough housing conditions she had: "there are good memories and then there are… because with a little bit you could play, with… it's true that Paris wasn't like Paris at the time. And it's true that we have good memories, we have less good memories of this cramped, dark apartment, there's one window, one front door. It's true that it's not ideal, right?"

Her father was also born in Paris, and her paternal grandmother was a caretaker in a building just across the street. At the birth of his fifth child, her father obtained a four-room apartment in Fresnes [southern suburbs] from his employer's social housing stock. "There it's true that there was really space, eh, compared to what we had known, we, the older ones, compared to my brothers who came after… there was really space, eh." Still, as told Laurence, this move uprooted this family of Parisians.

After technical training at the age of 16, she quickly found a job in Paris near the Champs Elysées. At 20, she married and lived on the ground floor without a bathroom in a house in Antony (close to Fresnes). She had three children, in 1969, 1973 and 1975. Her husband, who was a blue-collar worker, was transferred to the north of Paris and obtained social housing in Sarcelles (northern suburbs) through his employer. But she did not like living in this apartment. She couldn't stand the overcrowding, the neighborhood, the smells, the noises. "I didn't like it at all, I was on the verge of depression sometimes."

Four years later, she managed to convince her husband to move, especially since, like many tenants in social housing, they were receiving advertisements for off-plan houses in their mailboxes. They decided to build a six-room house in Persan [a municipality in the far northern suburbs] where there was still free and accessible land. Laurence has been living in Persan for 22 years and is very happy there. She appreciates the space all the more given how she suffered from the lack of it as a child:

"For me, it is space. But it's the fact… I think that must be it, the fact that when I was young, very young, I lived in cramped conditions, we couldn't play too much in the apartment, we had to go play outside. But… I think it can only be that, because I can't stand being in a cramped apartment. I can't stand it anymore. I need some space! (laughing). Space, the garden…"

Note: As Laurence is a counterexample in the sense that she moved to a highly peripheral location despite her socialization in Paris, we looked for the aspects of her life history that could explain this discordant behavior.

Exploring counterexamples can help nuance the quantitative association by bringing out cases that do not fit the quantitative analysis. Laurence's case showed that the role of residential socialization during childhood is not univocal. It can even have a negative effect, for example through bad housing or transport conditions. The "counterexamples" from the suburbs shed light on competing factors: school or work neighborhood may lead to other aspirations and representations. The counterexamples thus may give clues to improving statistical and theoretical models.

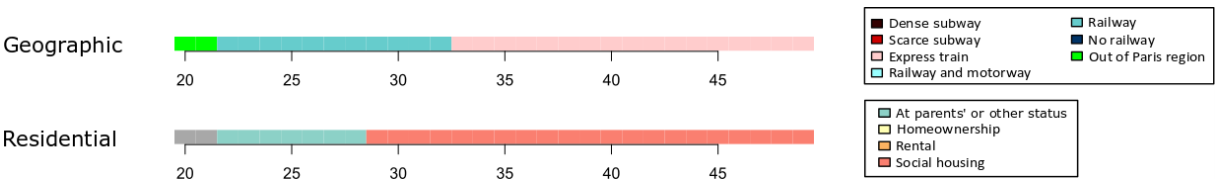
5.4 Divergent cases

The divergent cases are trajectories that do not follow the pattern identified by the quantitative analysis and are far from the typical trajectories of their group.

The trajectories of divergent cases among participants socialized in Paris (no. 295, 751: Fig. 6) mostly left central locations without purchasing a housing. These cases mainly highlight one limitation of the quantitative analysis: it does not take into account the partner's residential socialization, which may dominate the couple's residential choices.[6] For instance, Paule (Case 295), who grew up in Paris in a foster home, had no choice but to follow her husband, who was socialized in the suburbs. Her husband's employer offered them an apartment in a social housing close to her husband's workplace, which she could not refuse. She experienced her departure from Paris as an exile, while her husband could stay near his parents, friends and work. Interestingly, if her husband had been the respondent, the couple would have been classified as "socialized in the suburbs" and had an illustrative trajectory. The partner's socialization place is also observed among divergent cases of people socialized in the suburbs.

From another perspective, as members of the stable working classes, they, like many workers of the period, found refuge in social housing outside Paris when they formed a family. Here again, introducing social class as a factor would thus have improved the quantitative analysis.

Case 295: Paule, a divergent case of people socialized in Paris.



Born in 1950, abandoned by her parents at birth, Paule spent her entire childhood and adolescence in a series of foster homes in Paris, mostly in the 15th arrondissement. She obtained a diploma in floristry and found work in the 15th arrondissement. One of her colleagues introduced her to a friend who became her husband. He was a mechanic living with his parents in Villennes (distant western suburbs). After the wedding, they found an apartment (social housing) nearby Poissy through her husband's company. Paule found it difficult to move from Paris to the suburban social housing..

[6] Thus, it would be interesting to explore the interactions between the effects of residential socialization and gender.

"Well, they're foreigners, so you know it's not so easy to get used to it when you come from a big city… Well, I came from Paris, I liked life in Paris. It's still changing."

Her husband, on the other hand, didn't change anything in his life.

"He's comfortable, this was his area, he used to come around here all the time. Villennes, you know it's not far, it's 15 minutes by car. He used to come here all the time when he lived with his parents. He had friends here… he had lots of them."

As she repeatedly expressed in the interview, they had no other choice when the children came and started to grow up than to accept the social housing offered by the company. Paule admits that she would have liked to go back to live in the "15th arrondissement where I was in a foster home. That's where I was born too."

Note: In Paule's interview, we looked for factors that could explain her total divergence from the typical behaviors (staying in central locations and buying a dwelling).

The divergent cases mainly highlight the limitations of the quantitative analysis. The exploration of the qualitative interviews suggested several possibilities for improvement. The concept of socialization should also take into account residential socialization of the interviewee's partner, or other forms of spatial socialization, especially through education—which is observed for divergent cases of people socialized in the suburbs.[7]

5.5 Discussion

The mixed method proposed here enabled a deeper understanding of the role childhood socialization's places in their subsequent trajectory. It emphasizes the structuring role of the place where people grew up. In general, people tend to stay in a similar location, and to become emotionally attached to these places. Here, "places" can refer to different scales and forms of attachment: to a neighborhood, a municipality, a *département*, or an urban environment. The qualitative interviews showed the key role of social networks—people tend to stay close to them—, which can even act as a key resource in access to housing (Retière, 2003). The marginal illustrative cases further showed that this structuring role varies between social classes, as exemplified by the inheritance of social housing within Paris.

The counterexamples showed that it might be an ambiguous process. Painful experiences may drive people to reject their childhood neighborhood, because of housing or transport conditions or even because of the social composition of their place of residence (e.g. rejection of the bourgeois milieu). These cases are worth further investigations, as these counter-models are not well documented in the literature.

The counterexamples and divergent cases brought out competing forms of socialization, such as education and work. They are important in the construction of aspirations and representations. The study of residential mobility would thus be enriched by taking into account the geography of other life domains, and not only their place of residence.

The deviant cases reminded us that residential choices are often made within a couple. An important part of the unexplained deviation from the cross-case relationship is due to the omission of the residential socialization of the partner, or sometimes even the ex-partner. The choice seems easier when the partners share a similar social and residential background. It

[7] Cases for people socialized in the suburbs also show that some life events, such as divorce, can produce divergent residential trajectories.

would thus be interesting to better understand how these tradeoffs are made, and who "prevails"—taking a gender perspective—when the partners have differential residential socialization and aspirations.

The interviews showed that people usually try to stay in, or return to, places where they formed important bonds. Nevertheless, exploring cases along the *marginality* axis drew attention to the diversity of the means used to stay in their places of socialization: financial support, housing transfers or provision of housing from the family, or even information and tips from friends and colleagues. This diversity is partly related to diversity in the constraints and resources of different households, in terms of financial and social capital, for example.

These results counterbalance economic models (Evans, 1973) by showing how strongly choices are driven by subjective histories and may result from strategies whose logic is not exclusively economic. Finally, the results support the life course perspective, and the importance to take into account the effect of time, linked lives and the link between life domains. In this perspective, the combination of quantitative and qualitative methods is crucial to identify how individual characteristics, experiences and contexts are related to specific behaviors, but also to understand the variability of the meanings of experiences or transitions for distinct populations (Heinz, 2003), and even the non-linearity of the effect of time on the life course of individuals (Sánchez-Mira & Bernardi, 2000).

6 Generalization and Extensions

In order to keep our presentation as simple as possible, we focused our analysis on the relationship between individual trajectories and a single explanatory factor. However, the proposed tools can also be used with a typology of sequences, or to study several covariates simultaneously.

6.1 Using Multiple Covariates

The methodology proposed here can easily be extended to study several covariates simultaneously. The discrepancy analysis framework offers two methods for this purpose, which can both be used for the proposed case selection technique.

First, one can include several covariates using multifactor discrepancy analysis (Studer, Ritschard, Gabadinho, & Müller, 2011). As shown by McArdle and Anderson (2001), Equation (1) can also be computed with several covariates. The two proposed indicators can thus be computed using the same formulae. However, in this case their interpretation is slightly different: the *marginality* and *gain* indicators now take into account all the covariates taken together, and the specific effect of each covariate cannot be isolated. For instance, by simultaneously including the place of socialization and parents' social class, one would select cases that are typical or marginal of the two factors taken together. The same applies to the *gain* indicator, which would measure the information gain when both the place of socialization and the social class of origin are taken into account at the same time.

Second, it is also possible to include multiple covariates using sequence regression trees (Studer, Ritschard, Gabadinho, & Müller, 2011). These trees work as follows. First, all the sequences are grouped into a single "root" node. The procedure then splits this node in two according to the values of a covariate. The covariate, and the corresponding binary split, is chosen in such a way that the resulting child nodes differ as much as possible from one another,

or similarly, that the binary split explains the greatest part of the discrepancy of the sequences. The operation is then recursively repeated on each child node until no significant split is found or another stopping criterion is met (typically a minimal node size or a maximum tree depth).

A sequence regression tree highlights the combination of factors that most differentiates the trajectories or, in other words, that best explains the discrepancy of trajectories. Interestingly, it may uncover interaction effects between covariates, i.e. effects of covariates that depend on the value of another variable. For instance, social class may explain the discrepancy of the trajectories of people socialized in Paris, but not those of people socialized in the suburbs. Furthermore, various visual representations can be used to describe how covariates and trajectories are intertwined.

From a statistical perspective, the whole tree can be summarized by a categorical variable storing the terminal node membership. This variable is then used, for instance, to compute the overall fit statistics of a tree. The same "node membership" variable can be used to compute *marginality* and *gain*. Here again, one would select cases that are typical, representative or divergent from the quantitative relationships identified by the tree.

The two above-mentioned strategies allow including several covariates. However, the number of covariates should still be limited, as enough cases should be selected for each combination of covariates. In this respect, sequence regression trees are particularly interesting as they are more parsimonious. They aim to automatically uncover the relevant combination of covariates to study a trajectory and to ignore non-relevant combination.

6.2 Using a Typology of Sequences

As proposed by other authors (see for instance Latcheva & Herzog-Punzenberger [2011] or Verd & Andreu [2011]), one might select cases for a qualitative analysis based on a typology of sequences. In this case, the two proposed indicators can improve the diversity of the selected cases. Selecting cases with lowest *marginality* is roughly equivalent to the usual practice of selection using the medoids. However, selecting cases with "differing *marginality* values might help improve the diversity of the selected cases.

The *gain* indicator is useful to better understand the inner logic of the trajectories identified by the typology. Furthermore, looking at counterexamples and divergent cases is crucial to understanding the limitations of the typology, and therefore validating it (Piccarreta & Studer, 2019).

When using a typology in discrepancy analysis, the association should be very strong, because the typology is built to have groups as different as possible from one another. As a result, high and positive values of the *gain* indicator can be expected in most cases. We therefore recommend using a value greater than zero to distinguish "illustrative" and divergent cases. For instance, the average or median value could be chosen instead.

The proposed indicator can therefore increase the diversity of the qualitative sample by not only looking at the most central individuals, but also at those whose trajectory does not follow the pattern captured by the typology.

7 Conclusion

The mixed-method tools developed in this article offer a novel way to mix quantitative and qualitative approaches. Anchored in an explanatory sequential design, it allows qualitative data to be used to deepen the understanding of quantitative results. It serves the main purpose of mixed methods: integrating the two types of methods to obtain insights that could not be achieved using either type of method on its own (Tashakkori & Creswell, 2007).

The proposed method helps analyze qualitative data by situating them according to the quantitative analysis. These qualitative analyses can be used to improve the understanding of quantitative associations, but also to identify their limitations. As explained by Lieberman (2005), this process allows moving back and forth between the two types of materials to progressively enhance the quantitative analysis.

The guidelines proposed in the section 4.2.3, using typical and atypical cases, illustrative cases or counterexamples, offer various ways to enrich the study of quantitative associations by exploring the diversity of mechanisms at work. One important contribution of this analytical framework is the exploration of atypical cases, which are seldom considered in life course research, and are often hidden by statistical trends. Studying them recovers some of the diversity and complexity of behaviors, and theoretically allows the study of the role of trajectories that deviate from the most frequent patterns in a social phenomenon. Focusing on atypical cases is an opportunity to challenge not only statistical models, but also theoretical ones.

In contrast to the tools already developed for cross-sectional analyses (Seawright & Gerring, 2008), the two proposed indicators have the advantage of directly characterizing trajectories considered as complex objects (multidimensional sequences). Two difficulties related to this complexity are inherent in our analytical framework. On the one hand, the values of *marginality* and *gain* are not easily interpretable in relation to the types of trajectories, especially in our application with multidimensional sequences. On the other hand, the boundaries between divergent cases, counterexamples, representative illustrative cases and marginal cases are fuzzy and sometimes linked to methodological choices (e.g. state definitions in sequences). However, this in no way undermines the contributions of a guided analysis of the qualitative material.

Acknowledgments

We are grateful to Christopher Leichtnam (Ined, translation manager) and Paul Reeve for providing language help. Matthias Studer gracefully acknowledges the financial and research support of the Swiss National Science Foundation (project “Strengthening Sequence Analysis”, grant No.: 10001A_204740).

References

Abbott, A. (1990). Conceptions of Time and Events in Social Science Methods: Causal and Narrative Approaches. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 23(4), 140–150, DOI: 10.1080/01615440.1990.10594204

Abbott, A. (1995). Sequence analysis: new methods for old ideas. *Annual review of sociology*, *21*(1), 93–113.

Aisenbrey, S. & Fasang, A. E. (2010). New life for old ideas: The “second wave” of sequence analysis bringing the “course” back into the life course. *Sociological Methods & Research, 38*, 430–462

Authier, J.-Y., Bonvalet, C., & Lévy, J.-P., eds. (2010). *Elire domicile. La construction sociale des choix résidentiels*. Lyon : Presses Universitaires de Lyon.

Bernardi, L., Huinink, J., Settersten R. A. (2019). The life course cube: A tool for studying lives. *Advances in Life Course Research*, 41, 100258, DOI: 10.1016/j.alcr.2018.11.004.

Bidart, C., Cacciuttolo, P. (2013). Combining qualitative, quantitative and structural dimensions in a longitudinal perspective. The case of network influence. *Qual Quant*, 47, 2495–2515. https://doi.org/10.1007/s11135-012-9667-6

Billari, F. C. (2005). Life Course Analysis: Two (Complementary) Cultures? Some Reflections With Examples From the Analysis of Transition to Adulthood (pp. 267-88). In R. Levy, P. Ghisletta, J.-M. Le Goff, D. Spini, & E. Widmer (Eds), *Toward an Interdisciplinary Perspective on the Life Course*. Amsterdam, the Netherlands: Elsevier.

Bonvalet, C., & Brun, J. (2002). Etat des lieux des recherches sur la mobilité résidentielle en France. In J.-P. Lévy, & F. Dureau, *L’accès à la ville : les mobilités spatiales en questions* (pp. 15-64). Paris : L’Harmattan.

Bonvalet, C., & Dureau, F. (2000). Les modes d’habiter : des choix sous contraintes. In F. Dureau, V. Dupont, E. Lelièvre, J.-P. Lévy, & Lulle T. (coord.), *Métropoles en mouvement: une comparaison internationale* (pp . 131-153). Paris: Anthropos, Collection Villes.

Bonvalet, C., & Fribourg, A.-M., dir. (1990). *Stratégies résidentielles : actes du séminaire organisé par Catherine Bonvalet et Anne-Marie Fribourg (Paris, 1988)*. Paris : Ined.

Bonvalet, C., & Gotman, A., eds. (1993). *Le logement, une affaire de famille : l’approche intergénérationnelle des statuts résidentiels*. Paris : L’Harmattan.

Bonvalet, C., & Lelièvre, É. (2016). *Family beyond household and kin.* Paris: INED.

Clark, W. A. V. (2009). Changing Residential Preferences across Income, Education, and Age: Findings from the Multi-City Study of Urban Inequality. *Urban Affairs Review*, *44*(3), 334–355.

Clark, W. A. V., & Dieleman, F. M. (1996). *Households and Housing: Choice and Outcomes in the Housing Market*. New Brunswick, NJ: Center for Urban Policy Research.

Courgeau, D., & Lelièvre, É. (1991). The event history approach in demography. *Population an English Selection*, 63–79.

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage.

Dureau, F., Dupont, V., Lelièvre, E., Lévy, J. P., & Lulle, T. (Eds.) (2000). *Métropoles en mouvement: une comparaison internationale*.

Duvoisin, A. (2019). *Les origines du baby-boom en Suisse au prisme des parcours féminins*. Bern, Suisse: Peter Lang CH.

Elder, G. H. (1975). Age differentiation and the life course. *Annual review of sociology*, *1*(1), 165–190.

Elliott, J., Gale, C., Kuh, D. & Parsons, S. (2011). The design and content of the HALCyon qualitative study: a qualitative sub-study of the National Study of Health and Development and the Hertfordshire Cohort Study. *CLS Working Paper 2011/5*, London: Centre for Longitudinal Studies.

Evans, A. W. (1973). *The economics of residential location*. Springer.

Gauthier, J. A., Widmer, E. D., Bucher, P., & Notredame, C. (2010). Multichannel sequence analysis applied to social science data. *Sociological methodology*, *40*(1), 1-38.

Giele, J. Z., & Elder, G. H. (1998). *Methods of life course research: Qualitative and quantitative approaches*. Thousand Oaks, CA: Sage, doi : 10.4135/9781483348919

GRAB (2009). *Fuzzy states and complex trajectories: Observation, modelization and interpretation of life histories*. Paris : Ined, Méthodes et savoirs, 6.

Hägerstrand T. (1970). What about people in regional science? *9th European Congress of the Regional Science Association, in Regional Science Association Papers*, Vol. XXIV.

Heinz, W. R. (2003). Combining methods in life-course research: A mixed blessing. *Social dynamics of the life course. Transitions, institutionalism and interrelations*, 73–90.

Heinz, W. R., Kelle, U., Witzel, A., & Zinn, J. (1998). Vocational training and career development in Germany: Results from a longitudinal study. *International Journal of Behavioral Development*, *22*(1), 77–101.

Hollstein, B. (2014). Mixed methods social networks research: An introduction. In S. Domínguez & B. Hollstein B. (eds.), *Mixed methods social networks research: Design and applications* (pp. 3–34). Cambridge University Press.

Ivankova, N.V., Creswell, J.W., & Stick, S.L. (2006). Using Mixed-Methods Sequential Explanatory Design: From Theory to Practice. *Field methods*,18, 3–20. DOI:10.1177/1525822X05282260

Johnson, R.B., Onwuegbuzie, A.J., & Turner, L.A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research*, 1, 112–133.

Laborde, C., Lelièvre, É. & Vivier, G. (2007). Trajectories and Landmark Events: How People Narrate their Lives Analysis of Life Course Facts and Perceptions. *Population [English]*, vol. 62, (3), 489–505. doi:10.3917/popu.703.0567.

Latcheva, R., & Herzog-Punzenberger, B. (2011). Integration trajectories: A mixed method approach. In M. Wingens, M. Windzio, H. de Valk, C. Aybek, eds., *A life-course perspective on migration and integration* (pp. 121–142). Dordrecht: Springer.

Legewie, N. M., Tucci, I. (2021). Studying turning points in labour market trajectories – benefits of a panel-based mixed methods design. *Longitudinal and Life Course Studies*, 12 (1), 41-62.

Lelièvre, E. & Trabut, L. (2019). Approches qualitatives de questionnaires d'enquêtes quantitatives : analyse des réponses non-désirées et portraits. Presentation to *Les Rencontres de Statistique Appliquée* "*Articuler analyse quantitative et analyse qualitative sur des données d'enquêtes*", September 19th.

Le Roux, G., Imbert, C., Bringé, A., Bonvalet C. (2020). Social Transformation of the Paris Agglomeration in the 20th Century: A Longitudinal and Cohort Approach to Inequalities of Access to the City. *Population* [English], 75(1), 69–96.

Anonymized

Liao, T.F., Fasang, A.E. (2021). Comparing Groups of Life-Course Sequences Using the Bayesian Information Criterion and the Likelihood-Ratio Test. *Sociological Methodology*, 51(1), 44-85. doi:10.1177/0081175020959401

Lieberman, E. S. (2005). Nested analysis as a mixed-method strategy for comparative research. *American political science review*, *99*(3), 435–452.

Mayer, K. U. (2009). New Directions in Life Course Research. *Annual Review of Sociology*, 35, 413–33. https://doi.org/10.1146/annurev.soc.34.040507.134619.

McArdle, B. H. and Anderson, M. J. (2001). Fitting Multivariate Models to Community Data: A Comment on Distance-Based Redundancy Analysis. *Ecology 82*, 290-97.

Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and policy in mental health and mental health services research*, *42*(5), 533–544.

Pollock, G. (2007). Holistic trajectories: a study of combined employment, housing and family careers by using multiple-sequence analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *170*(1), 167-183.

Ohnmacht, T., Maksim, H., & Bergman, M. M. (Eds.). (2009). *Mobilities and inequality*. Ashgate Publishing, Ltd.

Østergaard, J., & Thomson, R. (2020). Thinking through cases: articulating variable and narrative logics on a longitudinal analysis of drug use and school drop out. *International Journal of Social Research Methodology*, 23(4), 423-436. DOI: 10.1080/13645579.2020.1719616.

Patton, M. Q. (2002). Qualitative research and evaluation methods (3rd ed.). Thousand Oaks, CA: Sage.

Piccarreta, R., & Studer, M. (2019). Holistic analysis of the life course: Methodological challenges and new perspectives. *Advances in Life Course Research*, *41*, 100251.

Pinçon, M., & Pinçon-Charlot, M. (2008). *Sociologie de Paris*. Paris: La Découverte.

Portes, A. & Fernández-Kelly, P. (2008). No Margin for Error: Educational and Occupational Achievement among Disadvantaged Children of Immigrants. *The Annals of the American Academy of Political and Social Science*, 620 (1), 12–36.

Remillon, D., & Lelièvre, E. (2018). Positions et transmissions socioprofessionnelles des femmes et des hommes au sein de lignées franciliennes. *Travail et emploi*, (154), 43-70.

Retière, J.-N. (2003). Autour de l'autochtonie. Réflexions sur la notion de capital social populaire. *Politix*, 16 (63), 121-143. DOI : https://doi.org/10.3406/polix.2003.1295

Rossi, P. H. (1955). *Why Families Move: A Study in the Social Psychology of Urban Residential Mobility*. Glencoe, IL: Free Press.

Sánchez-Mira, N., & Bernardi, L. (2020). Relative time and life course research. *Longitudinal and Life Course Studies*. doi: 10.1332/175795920X15918713165305.

Sandelowski, M., Voils, C. I., & Knafl, G. (2009). On Quantitizing. *Journal of mixed methods research*, *3*(3), 208–222. https://doi.org/10.1177/1558689809334210

Seawright, J., & Gerring J. (2008). Case selection techniques in case study research: A menu of qualitative and quantitative options. *Political Research Quarterly* 61(2), 294–308.

Shanahan, M. J. (2000). Pathways to Adulthood in Changing Societies: Variability and Mechanisms in Life Course Perspective. *Annual Review of Sociology*, 26, 667–692. http://www.jstor.org/stable/223461.

Sharland, E., Holland, P., Henderson, M., Zhang, M., Cheung, Y.S. & Scourfield, J. (2017). Assembling life history narratives from quantitative longitudinal panel data: What's the story for families using social work?, *International Journal of Social Research Methodology*, 20 (6), 667-679. doi: 10.1080/13645579.2017.1279915.

Small, M. L. (2009). How many cases do I need? On science and the logic of case selection in field-based research. *Ethnography*, 10(1), 5–38. https://doi.org/10.1177/1466138108099586

Studer, M., & Ritschard, G. (2016). What matters in differences between life trajectories: A comparative review of sequence dissimilarity measures. *Journal of the Royal Statistical Society: Series A* (Statistics in Society), 179(2), 481–511.

Studer, M., Ritschard, G., Gabadinho, A., & Müller, N. S. (2011). Discrepancy Analysis of State Sequences. *Sociological Methods & Research* 40(3), 471–510.

Tashakkori, A., & Creswell, J. W. (2007). The new era of mixed methods. *Journal of Mixed Methods Research*, 1(1), 3–7.

Tashakkori, A., & Teddlie, C. (1998). *Mixed Methodology: Combining Qualitative and Quantitative Approaches*. Thousand Oaks, CA: Sage.

Verd, J. M., & Andreu, M. L. (2011). The rewards of a qualitative approach to life-course research. The example of the effects of social protection policies on career paths. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 12(3).

Weymann, A. (1999). From education to employment: Occupations and careers in the social transformation of East Germany. In W. R. Heinz, ed., *From education to work: Cross-national perspectives* (p. 87–108). Cambridge University Press.