

The First World War and the disappearance of surnames in France: A trial estimation based on the Galton–Watson model

Pierre Darlu, Pascal Chareille

TRADUCTION **James Tovey**

DANS **POPULATION 2020/1 Vol. 75**, PAGES 97 À 128

ÉDITIONS **INED ÉDITIONS**

ISSN 1634-2941

DOI 10.3917/popu.2001.0101

Date de mise en ligne : 22/07/2020

Article disponible en ligne à l'adresse

<https://shs.cairn.info/journal-population-2020-1-page-97?lang=en>



Découvrir le sommaire de ce numéro, suivre la revue par email, s'abonner...
Scannez ce QR Code pour accéder à la page de ce numéro sur Cairn.info.



Distribution électronique Cairn.info pour Ined Éditions.

Vous avez l'autorisation de reproduire cet article dans les limites des conditions d'utilisation de Cairn.info ou, le cas échéant, des conditions générales de la licence souscrite par votre établissement. Détails et conditions sur [cairn.info/copyright](https://shs.cairn.info/copyright).

Sauf dispositions légales contraires, les usages numériques à des fins pédagogiques des présentes ressources sont soumises à l'autorisation de l'Éditeur ou, le cas échéant, de l'organisme de gestion collective habilité à cet effet. Il en est ainsi notamment en France avec le CFC qui est l'organisme agréé en la matière.



Pierre DARLU * and Pascal CHAREILLE **

The First World War and the Disappearance of Surnames in France: A Trial Estimation Based on the Galton–Watson Model

In France, every town has a memorial to its inhabitants who perished in the First World War. Numbering in the hundreds of thousands, the names of these individuals were engraved on such monuments to ensure their permanence in the nation's history. However, some names have disappeared with the deaths of their bearers. This article estimates the extent of the disappearance of surnames attributable to the war based on a comparison of two databases: the 'Morts pour la France' database and the INSEE surname file. How many surnames have disappeared? Which regions were most affected?

The death toll of the First World War was particularly high. The losses, though mainly military, were considerable. The number of dead among individuals enlisted in the French army alone was around 1.5 million (Héran, 2014).⁽¹⁾ Their deaths led to the disappearance of their surnames. The scale of this extinction and hence of the decrease in the number of French family names attributable to the conflict has failed to draw the attention of researchers. The issue may admittedly be considered incidental relative to the tragedy constituted by the slaughter of the war.

The disappearance of family names became a preoccupation among demographers and mathematicians as early as the mid-19th century. Research on the subject was motivated by fears of the gradual disappearance of upper-class families to the benefit of families from the demographically more fertile lower classes. The issue was brought to the fore by Sir Francis Galton in his eugenic thinking on racial degeneration. Working with the mathematician Henry

(1) Out of some eight million French individuals mobilized between 1914 and 1918, of whom 1.8 million young people born between 1894 and 1899 (Boulanger, 2003).

* UMR7206, MNHN-CNRS-Université Denis Diderot, Paris.

** Institut national d'études démographiques (INED) and CETHIS (EA 6298), Université de Tours.

Correspondence: Pierre Darlu, Université Denis Diderot, Paris, France. Email: pierre.darlu@mnhn.fr

William Watson (Watson and Galton, 1875), he laid the groundwork for a solution to a problem identified as early as 1845 by Irénée-Jules Bienaymé:

Considerable attention has been paid to the possible increase in the number of men, and various highly curious observations have recently been published on the fatality of the nobility, the bourgeoisie, and the families of illustrious men. It is said that this fatality will ineluctably lead to the disappearance of what is referred to as closed families. (Bienaymé, 1845)⁽²⁾

He responded to the challenge issued in 1873 by Alphonse de Candolle in these terms:

Naturally, all names must die out.... A mathematician could calculate how the reduction of names and titles will occur, consistent with the probability of all-female, all-male, or mixed births and the probability of the lack of births for a given couple. (De Candolle, 1873)

Looking beyond mathematical formulae, common sense suffices to understand that families having just one child have fewer chances of passing on their name than those who have more children.

This article estimates the number of surnames that disappeared because of the First World War, taking account of possible disparities between French departments (*départements*).⁽³⁾ While the focus of most of the work addressing the demographic consequences of the war is on the ‘disruption in individual behavior and practices’ (Faron, 2002), our intention here is not to explore the disruption that may have affected anthroponymic practices, even if considerable benefit may clearly be drawn from socio-anthropological approaches in the analysis of family behaviour and social and cultural practices.⁽⁴⁾ Instead, our focus is on the war’s impact on France’s patronymic stock.

Following a presentation of our data, drawn from the INSEE surname file and the file of ‘Morts pour la France’ (MPFs, or soldiers having died for their country), we will assess the Galton–Watson (G–W) model and the issues involved in applying it to the study of the disappearance of surnames. The model requires several parameters to be taken into account, including the distribution of the number of boys per household, the number of men likely to pass on their surname, and the occurrence of the surnames themselves (as a rare surname has a greater likelihood of dying out than a common one). Based on this model, we propose for each French department (French territory of 1871 corresponding to the current territory excluding Moselle, Bas-Rhin, and Haut-Rhin) and for France as a whole an estimate of the number of surnames

(2) Galton was probably unaware of this text.

(3) The documentation used requires us to limit our focus to the disappearance of surnames relating to the deaths of soldiers or civilians having obtained the distinction of having ‘died for France’. The losses attributable to the deaths of soldiers or civilians not having gained this distinction elude us. They are undoubtedly limited relative to the former. For example, we know that civilians accounted for no more than 12% of deaths (Héran, 2014), including the beginning of Spanish influenza.

(4) For a recent summary of the population impacts of the First World War, see Rohrbasser (2014). On the specific issue of mortality, Vallin remains the reference (1973, 1984).

whose extinction may be considered ‘natural’ (resulting from the simple chance of reproduction, be it socially or geographically differentiated). We then compare this estimate with the number of surnames included in the MPF file but missing from the INSEE surname file. Our discussion will focus on the geography of surname losses, a comparison of these losses with human losses, the sensitivity of the G–W model to various parameters, and the difficulties in assessing disappearances that can truly be attributed to the war.

I. Data used

This study is based on the statistical extrapolation of two computer files: the surname file provided by INSEE and the MPF file provided by the Department of Memory, Heritage, and Archives (DMPA) of the French Ministry of Defence.

1. The INSEE surname file

The INSEE surname file (INSEE, 1985) is a classic reference for addressing questions relating to family names. For each municipality in metropolitan France, it provides the number of births registered by surname for different periods, and notably the periods P1 [1891–1915] and P2 [1916–1940] used for this work.

The INSEE file is the most complete on this subject, but it is not exhaustive. It concerns only those people still alive in 1972, or so says INSEE, which produced the file, in its presentation.⁽⁵⁾ Whatever the case, the file can be used to provide estimates of the number of different names listed by department and the number of registered births.⁽⁶⁾

2. The ‘Morts pour la France’ file

Established by the Act of 2 July 1915 (amended in 1922), ‘Mort pour la France’ status is an official mention added to the civil registry (in the margin of records) and subject to Articles L. 488 and L. 492bis of the French code of military invalidity and war victims’ pensions. It initially concerned only soldiers killed in combat or having succumbed to war wounds. The file of recipients (MPF file) used here includes both the last name and first name of the deceased and their date and department of birth.⁽⁷⁾ While available in

(5) This assertion does not hold up to an in-depth analysis of the data. The number of (live) births in metropolitan France for the P1 period is nearly 20,000,000 (Daguët, 1995), and the mortality tables by generation (Vallin and Meslé, 2001) can be used to estimate the number of survivors in 1971 for the generations born between 1891 and 1915 at around 8,700,000. Yet the number of births listed in the file for this period is 10,500,000. The INSEE file thus surely includes people who were no longer alive in 1972 (certainly including those with MPF status).

(6) These figures are undoubtedly underestimations of the real values owing to the non-exhaustive nature of the INSEE file.

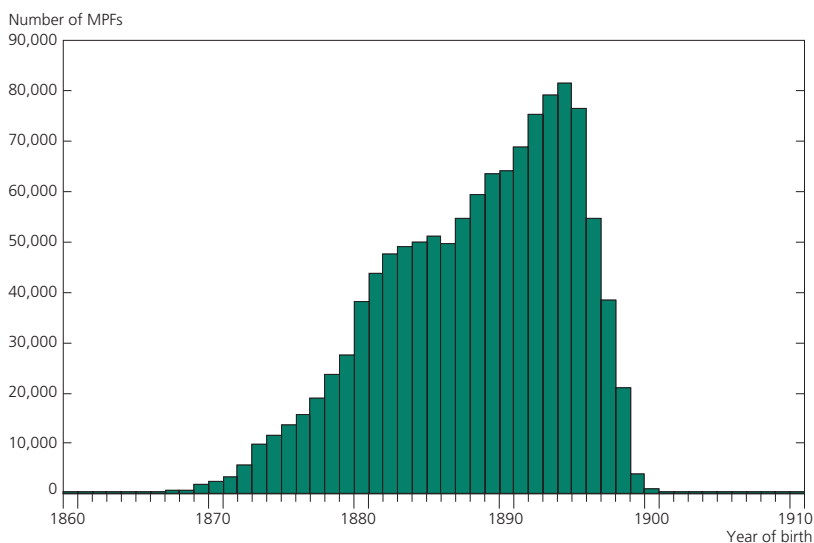
(7) This is still a ‘living’ file, as families may still request the attribution of MPF status for ascendants whom they consider unjustly forgotten or whose death may be directly attributable to the war.

the individual files of recipient soldiers, data on the municipality of birth and the date and cause of death are not systematically computerized; they are thus not used here.

The original file provided by the DMPA comprised 1,343,377 registrations,⁽⁸⁾ but this work takes account only of individuals with MPF status born in metropolitan France (the borders of 1871). The file has also deleted duplications.⁽⁹⁾ Following a ‘clean-up’, the ‘useful’ file comprises 1,211,523 MPFs corresponding to a patronymic stock of 179,037 surnames.

Though nearly 20% of the soldiers mobilized during the war lost their lives, the chronological distribution of MPF births (Figure 1) shows that not all generations paid the same price.

Figure 1. Distribution of MPFs by year of birth



Coverage: MPFs born between 1860 and 1910 in metropolitan France (1871 borders).

Source: MPF file.

The most impacted ‘classes’ are those of the generations born between 1870 and 1900, particularly MPFs born between 1892 and 1895.⁽¹⁰⁾ For these four classes, MPFs account for nearly 30% of mobilized soldiers.⁽¹¹⁾

(8) Status of the MPF file when made available on 19 April 2013.

(9) ‘Duplications’ here are considered as registrations corresponding to MPFs with the same surname, the same first names, and the same date and place of birth as another MPF.

(10) The word ‘class’ is used here in its military sense. A (recruitment) class corresponds to all men having reached age 20 and registered in the census tables. For MPFs, recruitment class and mobilization class may be different. The information relative to this (these) class(es) was recorded in the individual files of the soldiers having received MPF status but has not been computerized.

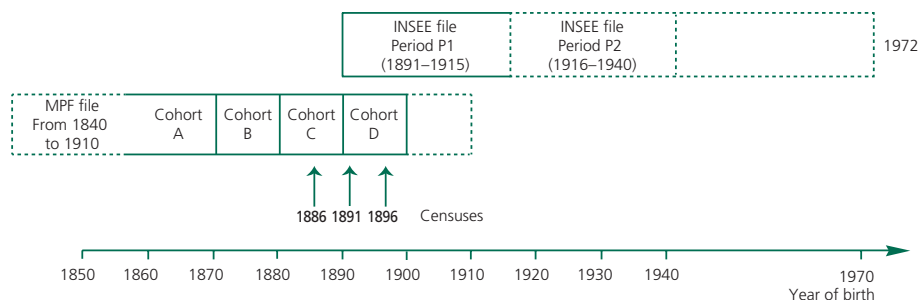
(11) The calculation was made by comparing the number of MPFs in the class in question with the number of mobilized individuals as indicated in *Les archives de la Grande Guerre*, Volume VII, No. 19, Tables B (p. 45) and C (p. 47).

3. Problems inherent in the constitution of the files

Errors in the transcription of names are inevitable when compiling a computer file. Some surnames included in the files have been altered relative to their original written form (random variations). Some names, often in low numbers, must surely have been created artificially, while others have disappeared from the corpus through excessive standardization. Without going through all the original documents (infeasible), it is impossible to assess the exact impact of these inaccurate transcriptions. However, and while this is a major issue when assessing the patronymic stock of a department, no attempt was made to standardize the names other than the standardization imposed by the individuals having produced the two files, who chose to write surnames in capital letters with no accents.⁽¹²⁾ The identity of two transcriptions is never a guarantee of a real patronymic identity, but unless transcription errors or the effects of standardization relating to the lack of accents are unevenly distributed geographically, the assessment of the patronymic stocks of departments, though inaccurate as an absolute value, may be considered correct as a relative value.

The chronological coverage of both the INSEE and MPF files, whose origin and production are different, is incomplete, as shown in Figure 2.

Figure 2. Chronology of the corpuses used



Notes: Of MPFs born in metropolitan France (1871 borders), 265 were born after 1900, and for 127 of them no year of birth is documented. They have not been considered in this study but account for just a minute fraction of the 1,211,523 MPFs in our 'useful' file. The 984 MPFs born before 1861 were included in Cohort A (accounting for just over 10% of the cohort).

Coverage: INSEE surname file (periods P1 [1891–1915] and P2 [1916–1940]) and MPFs born in metropolitan France (1871 borders) before 1901.

Sources: INSEE surname file; MPF file; population censuses of 1886, 1891, and 1896.

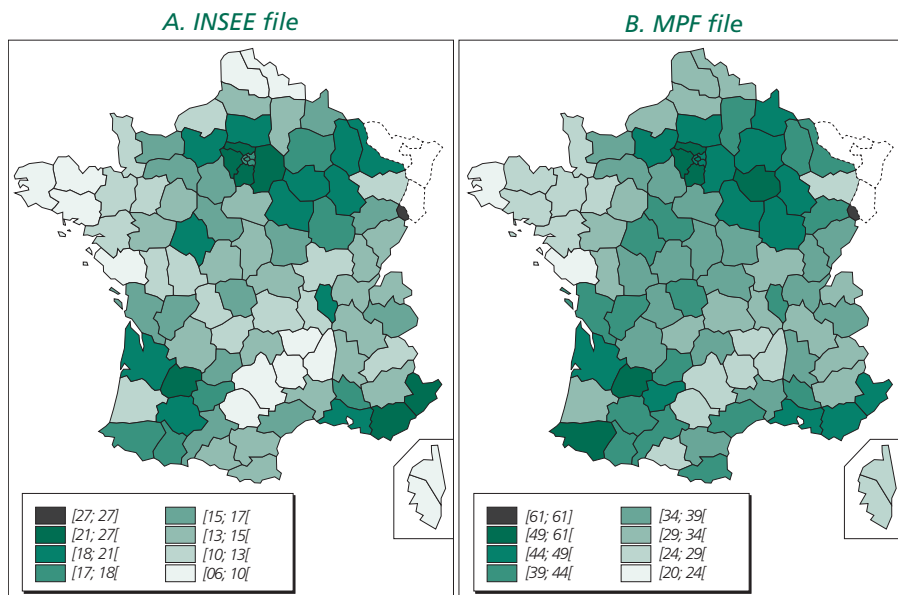
Besides the chronology of the two files, Figure 2 includes the three censuses used to determine the distribution of the number of children per household (necessary for estimating at the departmental level the number of disappeared surnames according to the G–W extinction model).

(12) This reduces the variety of written forms and, hence, mechanically, the diversity of patronymic stocks. The impact of the (other) transcription errors on this diversity is harder to estimate (reduction or increase).

4. Patronymic potential

To assess the representativeness of the patronymic data in the MPF file relative to the reference file constituted by the INSEE surname base, patronymic potential, measured as the number of different surnames per 100 people (Darlu et al., 1997), was calculated by department for the two files (Figure 3).

Figure 3. Patronymic potential by department (number of different surnames per 100)



Note: The patronymic potential is S/N where S is the number of (distinct) surnames and N the number of births recorded.

Coverage: INSEE file (individuals born in P1 [1891–1915]) and MPFs born in metropolitan France (1871 borders) before 1901 and whose department of birth is known.

Sources: INSEE surname file and MPF file.

Whether the calculation is based on the INSEE file for the P1 period [1891–1915] (Figure 3A) or on the MPF file (Figure 3B), the same departmental distribution is observed in the two estimations (linear correction $r = 0.94$).⁽¹³⁾ The representativeness of the MPF file may thus be considered satisfactory, at least from the standpoint of this spatial distribution. In both cases, the number of different surnames is lower in north-west France and higher in north-east, south-west, and south-east France. We know that this spatial distribution indirectly reflects migration flows (for example, Darlu and Ruffié, 1992), as the regions with high patronymic potential are those having experienced substantial immigration flows in the late 19th century, bringing with them new surnames.

(13) Figure 3B is also based on generations common to the two files (1891–1910), with no significant change to its appearance. The correlation between the departmental values of S/N (patronymic potential) based on the selected coverage and those based on common generations is over 0.95.

II. Methods

Where surnames are transmitted from father to child, which is the case in France for the periods we address, the survival of a surname from one generation to the next within a family is possible only if the father has at least one son likely himself to pass on his name to the next generation. But because the chance of survival of the surname of a father having two or more sons is necessarily higher, modeling the disappearance of a surname necessarily requires that the number of sons per father be taken into account. This is the general framework of the approach adopted by Watson and Galton to describe the disappearance of surnames in an iterative manner from one generation to the next. This model, generally used for population dynamics, is related to the theory of branching processes in discrete time. It is characterized by the fact that each individual in the population is born, reproduces, and dies, their descendants (in random number) being in turn and independently subject to the same events.

The disappearance of a surname also depends on its frequency in the reference population. A rare surname is more likely to die out than a common one. The frequency of surnames must thus be considered when modeling the disappearance; we have done so by applying the model to the MPF data.

1. The probability of surname disappearance with the G–W model

General framework

Let $p_0, p_1, p_2, \dots, p_s, \dots, p_q$ be the vector \mathbf{p} of the probabilities of a person (generally a father) transmitting their name to 0, 1, 2, ..., s , ..., q sons, with

$$\sum_{s=0}^q p_s = 1,$$

q being the maximum number of sons through whom the name may be transmitted. To simplify, we refer to this as father-to-son transmission, which was the most common if not the sole situation in France in the late 19th century and early 20th century, such that the probability p_s corresponds to that of having s sons.

Let x_1 be the probability of the disappearance of a surname of the first generation, that of the fathers, to the second, that of the sons.

Consider the case of the surname of a father. Where the latter has no sons, with a probability equal to p_0 , the probability of his surname disappearing is equal to $x_1^0 = 1$. In this case, the probability of the disappearance of the surname is equal to $p_0 x_1^0 = p_0$.

If this father has a single son to whom his name may be transmitted, the name will disappear with a probability equal to $p_1 x_1^1$, the product of the probability of having a single son, p_1 , and that of his name not being transmitted, x_1^1 . If the father has two sons, the probability of disappearance will be equal to $p_2 x_1^2$, the probability of having two sons, p_2 , and that each one of them, independently, does not transmit the surname, $x_1 \times x_1 = x_1^2$. If the father has s

sons, the probability of the name disappearing will be $p_s x_1^s$, the probability of having s sons, p_s , and that none of them transmit their surname, x_1^s . We can thus calculate the probability x_2 of the disappearance of the name in the second generation (from the second generation, that of the sons, to the next):

$$x_2 = p_0 x_1^0 + p_1 x_1^1 + p_2 x_1^2 + \dots + p_s x_1^s + \dots + p_q x_1^q \tag{1A}$$

Whence the following recurrence formula, which gives the probability of disappearance at generation n :

$$x_n = p_0 x_{n-1}^0 + p_1 x_{n-1}^1 + p_2 x_{n-1}^2 + \dots + p_s x_{n-1}^s + \dots + p_q x_{n-1}^q \tag{1B}$$

$$x_n = f(\mathbf{p}, x_{n-1}) \tag{1C}$$

The probability x_n of the extinction of a name thus increases from one generation to the next. But the complete disappearance occurs only in particular population conditions. It may be shown that, if we remain with the situation specified by Watson and Galton (Watson and Galton, 1875), i.e. the case of families with a maximum of two sons, the extinction is complete only on the condition that $\frac{p_0}{p_2} \geq 1$. Otherwise, the probability of extinction converges towards this relation (Bacaër, 2011). The situation is much more complex for a higher number of children per family, as this convergence depends on the distribution of the p_s .

The calculation of an extinction probability thus requires the exact number of ‘family generations’ (father to son) during which the G–W process plays out. It also requires an estimation of the vector \mathbf{p} .

Because we aim to calculate the probability of extinction by department, the values of x_n by department (x_{kn}) must be calculated based on an estimation of the vector \mathbf{p} by department: \mathbf{p}_k .

From the number of children per family to the number of sons per father

The censuses of the late 19th century provide population data giving the number of children per household by department and for France as a whole.⁽¹⁴⁾ These data can be used to estimate the vector \mathbf{p}_k required for calculating extinction probabilities.⁽¹⁵⁾ Unfortunately, the statistics available for the vector \mathbf{p}_k do not distinguish the number of girls and boys in the number of children. Consequently, to determine the number of boys per household – the only ones likely to transmit their surname – the values of the vector \mathbf{p}_k drawn from the population data were corrected (Brouard, 1989) to obtain the values of the vector \mathbf{p} as in Equation 1C. The correction assumed that the distribution of the number s of boys by sibship size n obeys a binomial law of parameters n and g , where g is the usual proportion of boys, or $g = 105/205$, assumed to

(14) The documents used (see note 16 below) indicate the ‘number of children per family’ while the population censuses report on households; it is thus advisable to refer to the ‘number of children per household’.

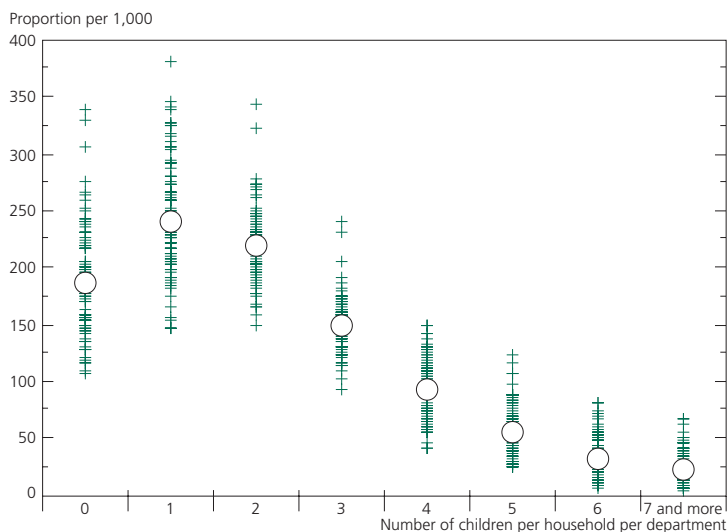
(15) It would be ideal to have the number of children per father (transmitting his surname), but it can be reasonably supposed that this number is not sufficiently different from the number of children per household to cause a major bias in the estimation of the vector \mathbf{p}_k .

be the same regardless of the sibship size. The probability p_s of the number s of boys in a family of n children is thus:

$$p_s = \frac{n!}{s!(n-s)!} g^s (1-g)^{n-s} \quad (2)$$

The censuses selected for this study, those of 1886, 1891, and 1896,⁽¹⁶⁾ specify for the 87 departments of metropolitan France corresponding to its borders of 1871⁽¹⁷⁾ the number of households with no children, one child, and up to seven children or more (the rare households with over seven children are aggregated here with those with seven children). Owing to uncertainties or errors in the censuses, already reported by Bonneuil (Bonneuil, 1989, 1997), we made our calculations based on the average by department of the results of 1886, 1891, and 1896 (Figure 4).⁽¹⁸⁾

Figure 4. Distribution of departments according to the proportion of households with 0 children, 1 child, or up to 7 or more children



Note: Each plus sign (+) corresponds to one of the 87 departments (1871 borders).

The proportion (per 1,000) is calculated based on the average of the three censuses (1886, 1891, 1896).

The average points are denoted by circles.

Coverage: Households listed in metropolitan France (1871 borders).

Sources: Population censuses of 1886, 1891 and 1896.

(16) The data used are accessible in electronic file format on the Centre de Recherches Historiques website of EHESS (see Béaur and Marin, 2011).

(17) The Territoire de Belfort, corresponding to the only part of Alsace and Haut-Rhin remaining in France following the defeat of 1871, was not created until 1922. It is nevertheless possible, based on MPFs' municipality of birth, to reattribute to this department the people who were born there, which enables us to take it into account in the calculations. This *a posteriori* reattribution was made when the data were computerized, since some MPFs are listed as being born in this department.

(18) The only data appearing to have no obvious errors are those of 1886, as reported in the original census lists and reused by Chervin (1888). Those of 1891 and 1896 are undoubtedly inaccurate for several departments, with aberrant values for the Landes department in 1891 and for the Ille-et-Vilaine, Meuse, and Yonne departments in 1896. These values were corrected and replaced by the interdepartmental average.

With these average demographic data, and by applying the G–W extinction model (Equations 1A–1C), we can estimate the probabilities of the extinction of a surname for each department k and calculate them for one or several generations. The same work may be carried out for France as a whole, by using the data on the number of children per household provided for this level of observation.

2. Applying the G–W model to ‘Morts pour la France’ data

The estimation of the probability of extinction as described above considers that, at the start of the iteration procedure, the probability of a father having between 0 and q sons is that of the vector \mathbf{p} .⁽¹⁹⁾ But this is not the case with MPFs. The MPFs born in 1850 had certainly ended their reproductive lives by 1914 and were thus able to have several sons, in which case the use of the vector \mathbf{p} is justified. However, those born in 1898 were only 20 at the end of the war and most certainly did not have the time to have more than two sons. The probability of MPFs having between 0 and s sons thus depends both on their year of birth and their year of death (between 1914 and 1918). This requires the application of the formula (Equation 1A) conditioned by the expectancy of the maximum number of sons, v , that they could have had in their life. To do so, x_2 (the probability of the extinction of the name in the second generation) simply needs to be replaced by:

$$y_2 = f(x_1|v) = \sum_{s=0}^v \frac{P_s}{\sum_{j=0}^v P_j} x_1^s \quad (3)$$

The descendants of these fathers then had the possibility of having between 0 and q sons, such that the formula (Equation 1) serving to calculate the probabilities for the ensuing family generations can be implemented without this truncation.

The question that remains is over how many ‘generations’ the G–W process needs to be pursued, i.e. how many iterations need to be considered. As the file used to identify the surnames of MPFs having disappeared is the INSEE surname file, which supposedly should only concern people still alive in 1972, it appears reasonable to pursue these iterations until that year.⁽²⁰⁾

The estimations of the probability of extinction were thus calculated for each birth cohort h in Table 1, which proposes a correspondence between the

(19) The vector \mathbf{p} changes over the generations, but we did not take this change into account as the lack of sufficiently reliable data meant we were unable to include this information. The selected estimation is based on the average of the three censuses (1886, 1891, and 1896). In addition, the correlations between the three series of departmental values are very high, suggesting a fairly slow change in the vector \mathbf{p} at the end of the 19th century.

(20) Other models could certainly have been devised, taking account of the MPF’s age at death and their age (possible for some, real for others) at paternity; because it was impossible to cover the individual histories of each person, we opted for plausible ‘iterations’.

Table 1. Correspondence (for MPFs) between year of birth, their division into cohorts, and the number of family generations separating them from 1972

Year of birth	Birth cohort h	Number of family generations until 1972	Number of MPFs	Number of surnames
[1826–1870]	A	5	8,672	6,331
[1871–1880]	B	4	168,750	59,835
[1881–1890]	C	4	533,202	117,740
[1891–1900]	D	3	500,507	114,106

Note: Cohort A includes the 984 MPFs born before 1961.
Coverage: MPFs born in metropolitan France (1871 borders) before 1901.
Source: MPF file.

year of birth (discretized in four classes) and the (possible) number of ‘family generations’ (three to five) until 1972.⁽²¹⁾

3. Estimation of the number of disappeared names according to the G–W model

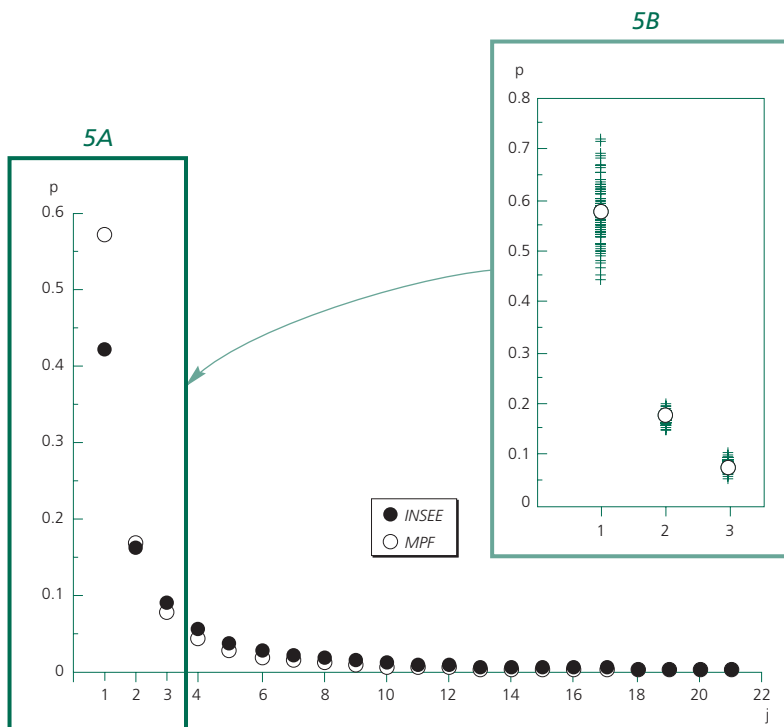
The probability of extinction described above applies to each MPF. It takes account of the MPF’s birth cohort and the number of iterations. However, the disappearance of a surname according to the G–W model also depends on the frequency of the surname, the probability being lower if the name is very common. As such, we need to determine the vector \mathbf{r}_k of the proportion of names represented once, twice, ..., u times in the department k :

$$\mathbf{r}_k = r_{k,1}, r_{k,2}, \dots, r_{k,j}, \dots, r_{k,u} \quad (4)$$

This vector can be estimated using several approaches. The first consists in calculating the vector \mathbf{r}_k based on all the names in the INSEE file for the first period P1 [1891–1915]. It may be assumed that these proportions represent, with no major biases, those of the names in the MPF file. The second consists in considering only the name in the MPF file and retaining, in the INSEE file, the number of occurrences of those names to deduce the vector \mathbf{r}_k . This method, which corresponds more closely to the names in the MPF file, was selected for the rest of the calculations. It shows a considerable and significant difference with the previous method only for the share of hapaxes (names occurring only once in a corpus), of which the average proportion per department is 42% in the INSEE file but 57% in the MPF file (Figure 5).

(21) The cohort [1891–1900] was defined *a priori* to respect the chronological milestones of the period P1 in the INSEE file [1891–1915] (see Figure 2). Other discretizations were attempted, which modify the results only marginally.

Figure 5. Average and departmental proportion of the number of names present r times in the MPF file and the INSEE surname file



Notes: Figure 5A: Average proportion p (calculated for the 87 departments) of the number of names attested to j times (based on the MPF file: symbols \circ ; based on the INSEE surname file (period P1 [1891–1915]): symbol \bullet). Figure 5B: Proportion of the number of names attested to $j = 1, 2,$ or 3 times in the MPF base for each of the 87 departments (symbol $+$) and the corresponding average value (symbol \circ).

Coverage: Individuals born in P1 [1891–1915] and MPFs born in metropolitan France (1871 borders) before 1901 and whose department of birth is known.

Sources: INSEE surname file and MPF file.

The number of names likely to disappear according to the G–W process is then calculated by department k as follows:

Let S_k be the number of surnames of MPFs in department k . Only a part of this number is likely to have disappeared under the G–W process. Let $D_{k,h}$ be this number per department k and for the cohort h :

$$D_{k,h} = S_k \times [r_{k,1} \times y_{k,h}^1 + r_{k,2} \times y_{k,h}^2 + \dots + r_{k,u} \times y_{k,h}^u + \dots] = S_k \times \alpha_{k,h}, \quad (5)$$

with $y_{k,h}^u$ the probability of extinction per department k and cohort h for the names in the MPF file calculated according to the procedure detailed above. The first term $r_{k,1} \times y_{k,h}^1$ is thus the probability of disappearance where the surname is borne by only one person in the department k , the second term $r_{k,2} \times y_{k,h}^2$ where it is borne by two people, and so on.

For a high-quality comparison between the INSEE and MPF bases, only Cohort D should be selected, i.e. MPFs born after 1891. As they died for France,

and thus were not alive in 1972, their birth should not in theory be reported in the INSEE file.

The sum extended to the different cohorts provides an estimate of the total number of names that could have disappeared according to the G–W process:

$$D_k = S_k \sum_{h=a}^d \alpha_{k,h} \quad (6)$$

III. Results

An IT procedure was implemented to search for surnames in the MPF file that are not included in the INSEE file. The 12,489 surnames extracted from this comparison have thus disappeared. They account for a little over 7% of the 179,037 surnames belonging to MPFs born in metropolitan France (1871 borders). But these disappearances have multiple causes, since if no MPF should be listed as an individual in the INSEE file,⁽²²⁾ their surname may well figure in the file. This is true where the surname belongs to possible descendants⁽²³⁾ and/or several collateral kin alive in 1972, or if it is borne by simple homonyms. If a name attested to in the MPF file is not included in the INSEE file, then it has disappeared:

- either by the G–W process, with a number of iterations separating the age at the possible paternity of the MPF and the date of 1972 (or between $n = 3$ and 5 according to the MPF's birth cohort; Table 1). The model described earlier can be used to estimate the number of this type of extinction;
- or because of the war, the disappearance of the MPF's name resulting from his death;
- or by the disappearance before 1972 of all possible people with the same surname as the MPF, whether related to him or not.⁽²⁴⁾

Only Cohort D in the MPF file (births between 1891 and 1900) covers the chronology of the INSEE file (births between 1891 and 1915). The estimates of the names having disappeared for this cohort, which is strongly represented (41% of MPFs), are in theory less biased than those of the cohorts corresponding to births before 1891. We nevertheless preferred to make the calculations while taking account of an estimation established on an average of the extinction probabilities calculated for each cohort (weighted by the corresponding number of surnames) and to apply this average probability

(22) See above and note 5.

(23) Because the children born during the war to an MPF father were aged between 54 and 58 in 1972, they could still have been alive at that date, and their name may thus be included in the INSEE file.

(24) It is difficult to make the distinction between surnames having disappeared owing to the First World War and those having disappeared owing, for example, to epidemics. While Spanish influenza (1918–1920) may have had an impact, its effect on the disappearance of surnames remains difficult to assess for France, particularly at the departmental level (on this epidemic, see Darmon, 2000; Lahaie, 2011; Vagneron, 2015). The same difficulty applies to typhoid fever in 1914–1915.

to the total number of surnames. This solution takes account of the number of different surnames observed in all the cohorts of a department, whereas a calculation based on each cohort would include surnames present in several cohorts.

1. Geography of human losses

The assessment of the human losses attributable to the First World War has been addressed by in-depth research. The oldest such studies attest to the early attention paid to the issue (Huber, 1931), while others have contributed to renewed general interest sparked by the centenary celebrations (Sangoï, 1997; Winter, 2004; Prost, 2008, 2014; Héran, 2014; Guillot and Parent, 2018). Most of this work has focused on the (direct and indirect) demographic consequences of the war.⁽²⁵⁾ Regarding France, however, research addressing the issue through an analysis of mortality rates by department remains relatively scarce. And research attempting to estimate the scale of departmental losses (Festy, 1984; Gilles, 2010; Gilles et al., 2014; Loez and Mariot, 2014) is hindered by a major problem: the MPF file does indeed mention the birth department of the recipients, but the calculation basis for estimating the relative weight of these deaths is a delicate choice. Should the number of MPFs born in a department be compared with the population of that department at the time of the 1911 census? For the male population only?⁽²⁶⁾ Or with the population of the enlisted men from the department? The choice is neither easy nor neutral. This is a sensitive issue as it is at the root of heated debate on the supposed sacrifices of certain regions that made a particularly strong contribution to the war effort. For example, the Bretons and Corsicans are said to have paid a heavier tribute than others.⁽²⁷⁾ Each calculation method has its merits, but none is ideal. We have selected two methods for this work. The first consists in simply considering the number of MPFs by department (Figure 6A). Thus assessed, the losses appear to be evenly distributed across France, though with slightly more deaths in Brittany, the Atlantic coast, in the north, and along a Gironde–Rhône axis.

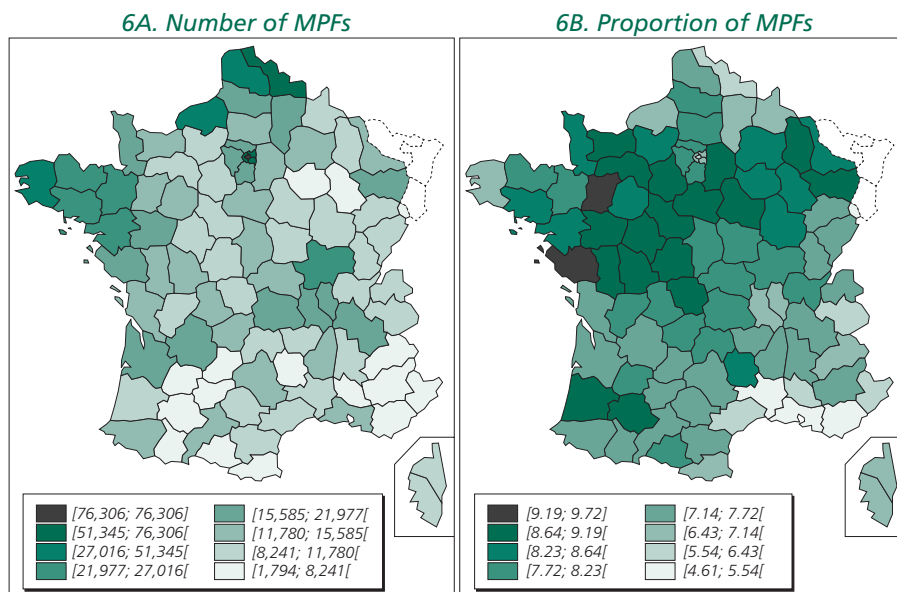
The second option ‘corrects’ the gross weight of these deaths by comparing it with all the male births corresponding to the most war-impacted generations (Figure 1). For this calculation, we chose as our departmental reference value the number of male births recorded between 1867 and 1899, corresponding

(25) See note 4.

(26) In the MPF file in our possession (as of 19 April 2013), 35 women are listed (at least). Obviously, these women were not soldiers because women were not to be admitted to the French army before the Act of 11 July 1938. These women, some of whom nurses, were among the civilian victims given MPF status, but their number is incidental relative to that of men.

(27) For debates over the unequal regional contribution to the war effort, see in particular Gilles et al. (2014) and Loez and Mariot (2014).

Figure 6. Number and proportion of MPFs by department (France, 1871 borders)



Note: Figure 6A: number of MPFs (department of birth); Figure 6B: proportion of MPFs (born in the department) relative to the total number of births registered between 1867 and 1899.

Coverage: MPFs born in metropolitan France (1871 borders) before 1901 and whose department of birth is known, and male births (same territory) between 1867 and 1899.

Sources: MPF file and *Annuaire statistique de la France* for the years 1867 to 1899.

to the mobilization of the soldiers of the 'classes' 1887 (called up between March and August 1916) to 1919 (April 1918).⁽²⁸⁾

This second option reveals stronger territorial contrasts. According to the department, losses account for between 9.7% in Vendée and 4.6% in Bouches-du-Rhône of the male births of the generations concerned (Figure 6B).⁽²⁹⁾

Relative to the number of births per department,⁽³⁰⁾ the (relative) losses were the highest along a line extending from Loire-Atlantique to Meurthe-en-Moselle following the lower part of the Loire and a good half of its middle part.

(28) For most of the departments corresponding to the borders of 1871, we have data covering the entire period thanks to the *Annuaire statistique de la France* (ASF) and/or the files provided on the websites of CRH and INSEE (note 16). We have made a few minor corrections (relative to the inconsistencies observed or incorrect values detected by making comparisons of the files or with the ASF) and estimated a few missing values through various adjustments.

(29) Without departmental mortality tables for the generations of the late 19th century, it is not possible to compare the number of MPFs born in a given department with the estimated number of men still alive in the department when the war broke out. The calculated values (%) are thus considerably lower than the weight of the real losses in the population of 'survivors' in 1914. According to Hérán (2014), at age 20, 72% of the male generation born in 1894 had survived infant and juvenile mortality, while 25 years earlier, the generation born in 1869 had already been reduced by 37% [at the same age].

(30) The choice of this option is not neutral. We see it as more coherent regarding the question of surname losses and the materials available for answering that question (MPF file and INSEE surname file).

This line continues east and north-east of the Loiret, in the Yonne and Seine-et-Marne departments, and, more broadly, those corresponding to the most intense combat zones (notably Meuse and Vosges). When considering the ‘Clémentel regions’⁽³¹⁾ of 1919, the regions of Nantes and Caen (roughly corresponding to today’s Pays de la Loire and Basse-Normandie) also show substantial human losses. This calculation method highlights a sharp contrast between the north and the south, the departments lying north of an axis between La Rochelle and Mulhouse having suffered much heavier losses, although the death toll is high in the Landes and Gers in the south-west. Mortality rates relative to births are considerably lower in the departments of south-east France, notably those on the Mediterranean.

2. Geography of surname losses

Losses by department

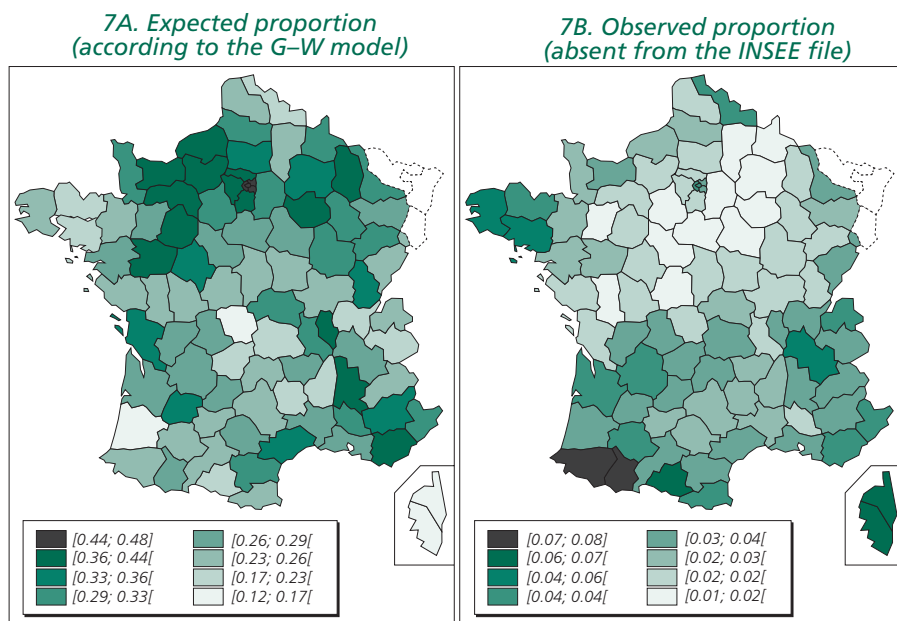
We have several indicators for each department that can be used to estimate expected surname losses and real surname losses (Appendix Table).

As the number of names belonging to MPFs changes considerably from one department to the next, the number of disappeared surnames was compared with the total number of MPF surnames per department. This proportion is much higher for surnames whose extinction is expected under the G–W model (Figure 7A) than for the surnames of MPFs absent from the INSEE file (Figure 7B). The comparison was made both for the cohort of MPFs whose birth should have been registered in the INSEE file (Cohort D, births from 1891 onwards) and for all the cohorts weighted by their size (Figure 7A), with no substantial changes in the differences between departments (linear correction $r = 0.99$).

Assessed solely on the basis of the MPF names effectively absent from the INSEE file (Figure 7B), the disappearances of surnames represent just 1% to 8% of the stock, depending on the department. A comparison of Figures 6B and 7B also shows that the geography of human losses differs considerably from that of the effective disappearances of surnames. While human losses are the highest along the middle and lower stretches of the Loire, surname losses (observed by comparing the two files) are very low in this central region of the country. Effective surname losses are substantial in Brittany and Corsica, with their low ‘patronymic potential’ and average human losses. In the Basque Country, with its high immigration levels in the late 19th century and, hence, high ‘patronymic potential’, human losses are at the national average, while surname losses are considerable compared with the rest of the country.

(31) The First World War was at the root of the creation of the French regions. Regional economic groupings, referred to as the ‘Clémentel regions’, were introduced in 1919 largely on the basis of chamber of commerce groupings.

Figure 7. Expected and observed proportions of disappeared surnames by department (1871 borders)



Note: The proportions (expected and observed) of disappeared surnames are calculated relative to the total number of surnames of MPFs from the department.

Coverage: MPFs born in metropolitan France (1871 borders) before 1901 and whose department of birth is known.

Sources: MPF file and INSEE surname file for the identification of disappeared surnames.

The application of the G–W model (Appendix Table) shows that expected surname disappearances are much higher for the most recent cohorts (births after 1881) than for previous cohorts (births before 1881). This is logical, as the older generations lived longer and thus had more opportunity to transmit their surnames. According to the G–W model and as a departmental average, Cohort D lost 618 surnames compared with just 124 for the Cohort A. These differences apply to the MPF surnames that have effectively disappeared from the INSEE file, though at a more modest level, with average surname disappearances for the two cohorts totaling 59 and 13, respectively.

Equation 3 expresses surname disappearance probabilities while considering that the death of MPFs put an end to their possibility of transmitting their name (their death having occurred before the end of their reproductive lives). The effect of this correction is more emphatic for Cohort D (the youngest MPFs, i.e. those born after 1890 and whose descendance was probably cut short).⁽³²⁾ Depending on the department, the estimated number of expected surname

(32) The comparison of the results between departments shows that the impact of the correction (application of Equation 3) is low, albeit significant ($p < .001$); the correlation between departmental measurements with and without correction remains very high ($r = 0.99$).

extinctions is between 1.1 and 1.3 times what it would have been if the MPFs had not died and had thus had every opportunity to transmit their name. For example, the expected number of surname extinctions for the Eure department is estimated at 1,347 when the correction (Equation 3) is applied (1,257 without the correction). This number is 344 for Corsica (273 without the correction) and 10,178 for France as a whole (8,929 without the correction).

Figure 7A also shows that the expected disappearance of surnames (according to the G–W model) varies substantially from one department to the next (Appendix Table). The highest values are in the northern regions of France, particularly around Normandy, as well as in the south-east. This variability cannot be attributed to the model itself but to contrasted population realities because the total number of different surnames, the distribution of the number of children per household (particularly the proportion of households without sons or with only one son), and the expectancy of the number of sons possible for a father according to his years of birth and death vary considerably between departments.

These expected disappearances are directly related to the construction of the assessment, their number being strongly correlated to the frequency of households with no children (Figure 8).⁽³³⁾ This result is logical since the surnames of households without children (or with a low number of children) are the most likely disappear.

But surprisingly, a comparison of the surname extinctions expected under G–W with the number of surnames in the MPF file effectively absent from the INSEE file shows that the number of effective extinctions (surnames missing from the INSEE file) is much lower for almost all departments than the G–W model predicts (Appendix Table). Several explanations can be posited:

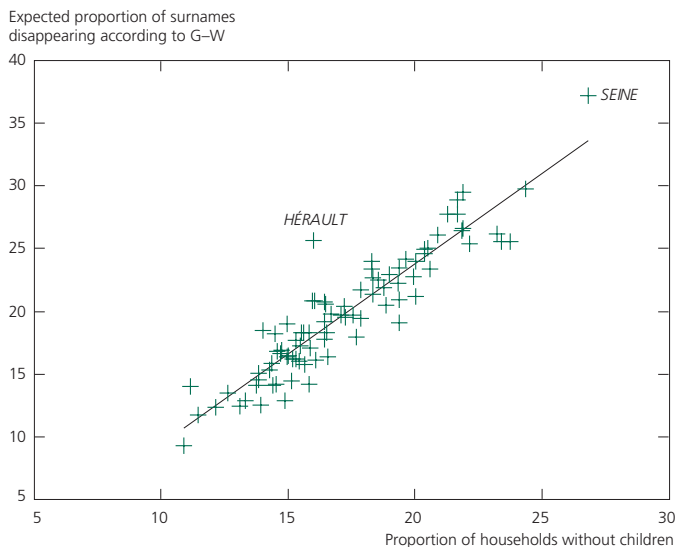
1. As we mentioned earlier, more births are listed in the INSEE file for the 1881–1915 period than expected. It is thus not impossible to imagine that MPFs, and thus their surnames, figure among this surplus.

2. The estimation of the vector r_k (Equation 4) of the frequency of names represented 1, 2, ..., u times (Equation 5) in the population of MPFs is possibly biased by the overestimation $r_{k,1}$ of the names occurring just once among MPFs (Figure 5).⁽³⁴⁾ The weight of the names borne by a low number of individuals – and notably that of hapaxes – is substantial in the estimation of the number of extinctions expected under the G–W model.

(33) This frequency corresponds to the p_0 of Equation 1. The linear correlation between the values per department of the proportion of expected surname disappearances (according to the G–W model) and the values of p_0 is $r = 0.93$. The correlation is $r = 0.56$ if we consider the frequency of households having just one child.

(34) The over-representation of hapaxes in the MPF file remains difficult to explain. It may result from a large number of transcription errors in the initial data entry of the surnames. For example, the proportion of double surnames is much higher in the MPF file than in the INSEE file. The electronic file we used also includes MPF surnames that are evidently not real surnames.

Figure 8. Relationship between the proportion of MPF surnames whose disappearance is expected according to the G–W model and the proportion of households without children (%), by department



Coverage: MPFs born in metropolitan France (1871 borders) before 1901 and whose department of birth is known, and births (same territory) between 1867 and 1899.

Sources: MPF file and *Annuaire statistique de la France* for the years 1867 to 1899 (see note 28).

3. The number of expected disappearances under G–W and the number of surnames in the MPF file missing from the INSEE file are assessed by department. A name may disappear from a department, however, without disappearing from the national corpus. We can assume that some surnames expected to disappear according to the G–W model in one department remain present in others, whether or not the people with that name are related (for example, brothers, cousins, uncles, or nephews). Migration, which concerned a considerable part of the population at the beginning of the 20th century, may have led to the reintegration of some names in departments where their extinction was expected according to the G–W model.

In any case, the departmental distribution of MPF surnames missing from the INSEE file (Figure 7B) is enlightening. While based on a low number of surnames, it reveals a closer relationship between the number of different surnames S_k and the number N_k of MPFs per department.⁽³⁵⁾ Surname extinctions thus appear to have been largely proportional to the patronymic stock and/or total number of MPFs. Pyrénées-Atlantiques and Hautes-Pyrénées (as well as, to a lesser extent, Ariège and Corsica) are the only departments with an exceptional level of disappearances, undoubtedly attributable to the anthroponymic

(35) Linear correlations of $r = 0.92$ and $r = 0.86$, respectively. The Seine department has a strong impact on these correlations, as they fall to $r = 0.76$ and $r = 0.69$, when it is excluded.

particularities of regions with a high number of surnames, combining names of Basque origin with those of Occitan origin (Darlu and Oyharçabal, 2006).

In addition, an analysis of the names having effectively disappeared (i.e. those included in the MPF file but not in that of INSEE) shows that over 93% of them were hapaxes (from the MPF corpus).⁽³⁶⁾ From an anthroponymic standpoint, the war merely served to precipitate the disappearance of extremely rare surnames, which MPFs were probably often the only people of reproductive age to bear.⁽³⁷⁾ None of the disappeared surnames is attested to more than 3 times in the MPF corpus,⁽³⁸⁾ confirming that surname losses were not a result of the disappearance of large sibships which were the only bearers of a rare anthroponymic heritage.⁽³⁹⁾ If the disappeared surnames are sorted into two groups, belonging or not belonging to MPFs born in the same department, almost all the extinctions (99%) concern geohapax names, i.e. names present exclusively in a department.⁽⁴⁰⁾ There is a strong correlation ($r = 0.95$) between the number of MPF surnames missing from the INSEE file and the number of geohapaxes per department.⁽⁴¹⁾ The local deep-rootedness of surnames thus weakens their chances of survival.

Lastly, Figure 9 shows the considerable contrast between the proportion of surnames that may be expected to disappear according to the G–W model and the proportion of surnames that have effectively disappeared from the INSEE file. While the average of this relation is between 10% and 15% (meaning that there are 10 to 15 times fewer surnames absent from the INSEE file than expected), the percentage is between 30% and 47% for a few departments in the south-west, Brittany, and Corsica.

Losses for France as a whole

The G–W model can be applied to metropolitan France in its entirety. Although relinquishing the departmental approach tends to erase the differences

(36) These disappeared surnames include a high proportion of double surnames. The latter account for just 5% of the corpus of MPF names and 6% of the names in the INSEE file but account for nearly 15% of the extinctions.

(37) In any case, other possible contemporary homonyms, related or not to these MPFs, certainly did not have a descendance (or a descendance that did not survive beyond 1971); otherwise, their surname would be included in the INSEE file.

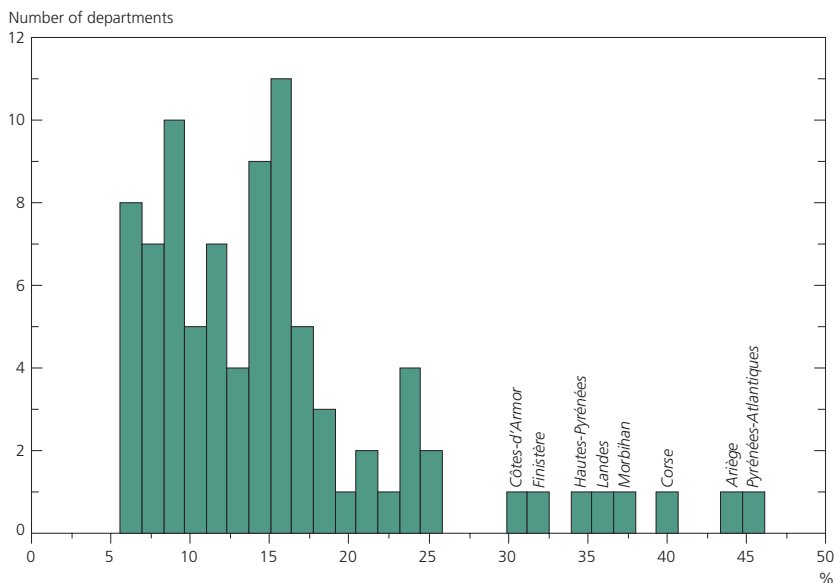
(38) Those having disappeared, while borne by three MPFs, total 20: Boucaine, Codaccionni, Gamee, Goanach, Gossegin, La Bouere, Le Dicabel, Le Porcq, Leblevec, Legouvello de la Porte, Lesolec, Luzj, Marchetay, Martruc, Pasquau, Prost Toulland, Rohfrisch, Roumas Bertranine, Schoonhere and Sourseau. There are also uncertain extinctions, as the written forms 'Codaccioni', 'Le Blevec', and 'De la Bouere' are attested to in the INSEE file.

(39) Those who were killed shared their surname with other men in the population. These certainly exist, even if their exact number has yet to be determined because patronymic identity is not sufficient proof of a direct genealogical link, including for births registered under the same name in the same department.

(40) On the concept of geohapax, see Chareille and Darlu (2013).

(41) There is also a strong correlation between the number of extinctions expected by the G–W model and the number of hapaxes per department ($r = 0.97$) or geohapaxes per department ($r = 0.80$). Disappearances estimated by G–W are correlated to these indicators.

Figure 9. Number of departments according to the relation (as a %) between the number of MPF surnames absent from the INSEE file and the number of MPF surnames expected to disappear according to the G–W process



Interpretation: Five departments show a relation between the number of MPF surnames missing from the INSEE file and the number of MPF surnames expected to disappear according to the G–W process of around 10%. For 11 departments, this percentage is around 16%; for a single department (Pyrénées-Atlantiques), it exceeds 45%.

Coverage: MPFs born in metropolitan France (1871 borders) before 1901.

Sources: MPF file (and INSEE surname file for the identification of extinct surnames).

whose importance has been stressed (in the estimation of the vector \mathbf{p}_k of the number of children per household or in that of the vector \mathbf{r}_k of the number of surnames present n times in department k), it serves to assess complete and definitive surname extinctions (expected according to G–W or observed through comparisons of the MPF and INSEE files) at the national level. This leaves aside the problem posed by the possible reintegration (following a migration) of a surname in a department where its disappearance may be expected according to G–W.

Applied to France as a whole, the implementation of the surname extinction model relies on estimations of the vectors \mathbf{p} and \mathbf{r} made on this scale. While the average proportion of hapaxes per department in the MPF file was 57%, the proportion of hapaxes in metropolitan France is just 16% of the MPF surname corpus (some surnames that are hapaxes in one department may be present in another and are thus not hapaxes for metropolitan France). The difference is considerable and greatly affects the estimation of extinctions expected according to the G–W model, serving to reduce it.

For France as a whole, the number of MPF surnames expected to disappear under the G–W model is 10,178, whereas the number of surnames in the MPF

file missing from the INSEE file is 12,837. The difference means we can estimate at around 2,600 the number of surnames whose disappearance may be attributed to the First World War. This accounts for 1.4% of all the surnames belonging to MPFs. At the national level, the impact of the war thus appears modest, but at the departmental level it is more considerable and, as we have seen, sharply contrasted. These findings underline the problems posed when applying the G–W extinction model to different territorial scales.

Conclusion

The geographical distribution of human losses resulting from the First World War is contrasted. The strong contribution made by Brittany and northern France to the war effort changes when taking account of the birth rates in the decades preceding the conflict. The ‘human sacrifice’ turns out to be greater to the north of a La Rochelle–Strasbourg axis and more modest in the south-east.

But the patronymic losses truly attributable to the First World War remain difficult to assess. If we limit ourselves to the losses highlighted through a simple comparison of the MPF and INSEE files, the surnames of MPFs that no longer occur in the INSEE file (1891–1940) represent between just 1.3% (Eure-et-Loir department) and 8.4% (Pyrénées-Atlantiques department) of the total number of different surnames of MPFs. According to the surname extinction model proposed by Galton and Watson, the number of expected surname disappearances should have been much higher and the geographical distribution substantially different from that obtained by the comparison of the MPF and INSEE data.

The differences observed between the two approaches result, on the one hand, from the imperfect chronological coverage of the two files, which makes them difficult to compare, and, on the other hand, from the implementation at the departmental level of the G–W surname extinction model. Though necessary from our geopatronymic perspective, this approach generates an overestimation of the number of hapaxes in the MPF file. The probability of the extinction of these hapaxes is strong, and their weight in the implementation of the G–W model is essential. Yet certain names that are hapaxes in one department are attested to elsewhere. Their extinction in a department may thus be expected according to the G–W model but is not actually the case (having possibly been reintroduced in the department following migration).

But one thing is sure: France’s patronymic stock resisted strongly to the impact of a war that marked the country deeply and on a lasting basis. After the First World War, the department-to-department mobility of the population and, hence, the mobility of surnames increased considerably, and immigration served to introduce new surnames. Consequently, the country’s stock of surnames, by

department and at the national level, has continued to grow, definitively removing any risk of the impoverishment of a heritage in perpetual motion.

Acknowledgements: The authors would like to thank Laurent Veysseyre, Head of the Cultural Heritage Delegation of the Department of Memory, Heritage and Archives (DMPA) of the French Defence Ministry, as well as his colleague Sandrine Aufray, who helped us to obtain the electronic database and the corresponding 'licence to reuse public information' (Contract of 27 May 2013 between the Defence Ministry and the UMR 7323 and 7206 mixed research units).

This work also benefitted from a subsidy from the French National Research Agency (ANR) (ANR-18-CE02-0011, MathKinD).



APPENDIX

Appendix Table. Detailed data per department

Department number	Total no. of births [1867–1899]	Total no. of MPFs	Total no. of MPFs / Total no. of births [1867–1899] (%)	Total no. of different surnames for MPFs	Expected no. of surnames disappearing under G–W	Expected % of surnames disappearing under G–W	No. of MPF surnames absent from INSEE file	% of MPF surnames absent from the INSEE file	% of MPF surnames absent from the INSEE file / expected % of surnames disappearing under G–W	Total no. of surnames / total no. of births [INSEE P1]	Total no. of surnames / Total no. of MPFs	Proportion of households with no children
01	141,601	11,482	8.109	4,153	811	0.195	123	0.030	0.152	0.144	0.362	0.1726
02	225,459	15,585	6.913	6,327	1,440	0.228	95	0.015	0.066	0.132	0.406	0.1996
03	174,176	13,782	7.913	4,882	968	0.198	135	0.028	0.139	0.142	0.354	0.1671
04	56,454	4,095	7.254	1,240	290	0.234	37	0.030	0.128	0.138	0.303	0.2059
05	58,284	4,092	7.021	1,271	208	0.164	38	0.030	0.183	0.124	0.311	0.1496
06	112,745	6,503	5.768	2,913	649	0.223	124	0.043	0.191	0.218	0.448	0.1935
07	189,854	13,794	7.266	3,484	436	0.125	93	0.027	0.213	0.089	0.253	0.1391
08	127,164	8,593	6.757	3,923	883	0.225	67	0.017	0.076	0.164	0.457	0.1858
09	95,817	7,534	7.863	1,976	279	0.141	122	0.062	0.437	0.139	0.262	0.1376
10	88,026	7,291	8.283	3,736	1,035	0.277	58	0.016	0.056	0.213	0.512	0.2129
11	128,121	9,576	7.474	3,440	716	0.208	110	0.032	0.154	0.138	0.359	0.1596
12	200,438	14,992	7.480	3,607	508	0.141	89	0.025	0.175	0.085	0.241	0.1438
13	286,882	13,213	4.606	6,302	1,305	0.207	202	0.032	0.155	0.196	0.477	0.1644
14	157,382	13,643	8.669	4,884	1,300	0.266	147	0.030	0.113	0.157	0.358	0.2191
15	100,886	8,241	8.169	3,017	502	0.166	88	0.029	0.175	0.134	0.366	0.1470
16	135,479	10,529	7.772	4,519	859	0.190	130	0.029	0.151	0.153	0.429	0.1498
17	167,625	12,933	7.715	5,200	1,215	0.234	97	0.019	0.080	0.152	0.402	0.1828
18	149,632	11,780	7.873	4,298	680	0.158	79	0.018	0.116	0.142	0.365	0.1435
19	158,749	12,659	7.974	4,154	562	0.135	120	0.029	0.214	0.117	0.328	0.1264
20	141,935	9,436	6.648	2,461	344	0.140	140	0.057	0.407	0.083	0.261	0.1115
21	127,562	10,816	8.479	4,773	1,118	0.234	103	0.022	0.092	0.182	0.441	0.1939
22	320,183	25,607	7.998	6,276	806	0.128	250	0.040	0.310	0.083	0.245	0.1331
23	108,675	9,641	8.871	3,916	459	0.117	72	0.018	0.157	0.152	0.406	0.1146
24	211,981	16,200	7.642	6,109	1,185	0.194	237	0.039	0.200	0.138	0.377	0.1786
25	136,344	9,974	7.315	3,786	654	0.173	107	0.028	0.164	0.145	0.380	0.1551
26	128,048	9,272	7.241	3,446	912	0.265	104	0.030	0.114	0.143	0.372	0.2187
27	122,661	10,246	8.353	4,529	1,347	0.298	90	0.020	0.067	0.191	0.442	0.2439
28	111,647	10,259	9.189	3,840	805	0.210	49	0.013	0.061	0.158	0.374	0.1937
29	420,018	28,854	6.870	5,678	854	0.150	269	0.047	0.315	0.059	0.197	0.1382
30	195,464	10,820	5.536	3,858	706	0.183	95	0.025	0.135	0.135	0.357	0.1563
31	164,183	12,126	7.386	4,729	838	0.177	134	0.028	0.160	0.165	0.390	0.1531
32	82,067	7,434	9.058	3,059	468	0.153	114	0.037	0.244	0.184	0.411	0.1428
33	276,460	20,172	7.297	9,282	2,031	0.219	342	0.037	0.168	0.191	0.460	0.1877
34	186,195	10,896	5.852	4,436	1,139	0.257	146	0.033	0.128	0.160	0.407	0.1600
35	292,136	23,061	7.894	6,286	1,012	0.161	170	0.027	0.168	0.105	0.273	0.1610
36	120,476	10,649	8.839	3,774	612	0.162	66	0.017	0.108	0.127	0.354	0.1519
37	113,184	9,998	8.833	4,257	1,047	0.246	91	0.021	0.087	0.184	0.426	0.2037
38	230,879	16,880	7.311	5,662	1,117	0.197	261	0.046	0.234	0.143	0.335	0.1709
39	114,834	8,802	7.665	3,318	795	0.239	64	0.019	0.081	0.141	0.377	0.2002

Appendix Table (cont'd). Detailed data per department

Department number	Total no. of births [1867–1899]	Total no. of MPFs	Total no. of MPFs / Total no. of births [1867–1899] (%)	Total no. of different surnames for MPFs	Expected no. of surnames disappearing under G–W	Expected % of surnames disappearing under G–W	No. of MPF surnames absent from INSEE file	% of MPF surnames absent from the INSEE file	% of MPF surnames absent from the INSEE file / expected % of surnames disappearing under G–W	Total no. of surnames / total no. of births [INSEE P1]	Total no. of surnames / Total no. of MPFs	Proportion of households with no children
40	129,668	11,500	8.869	3,487	323	0.093	116	0.033	0.359	0.110	0.303	0.1088
41	111,061	9,944	8.954	3,884	746	0.192	65	0.017	0.087	0.153	0.391	0.1645
42	283,699	19,201	6.768	5,903	1,249	0.212	132	0.022	0.106	0.115	0.307	0.2004
43	152,431	11,246	7.378	2,919	479	0.164	81	0.028	0.169	0.079	0.260	0.1659
44	265,891	22,960	8.635	6,045	1,031	0.170	160	0.026	0.155	0.111	0.263	0.1588
45	154,447	13,817	8.946	5,025	919	0.183	85	0.017	0.092	0.147	0.364	0.1653
46	97,292	7,212	7.413	2,842	524	0.184	84	0.030	0.160	0.150	0.394	0.1403
47	92,111	7,483	8.124	3,929	985	0.251	113	0.029	0.115	0.226	0.525	0.2051
48	74,549	6,181	8.291	1,620	209	0.129	39	0.024	0.186	0.080	0.262	0.1489
49	184,467	16,472	8.930	5,400	1,373	0.254	131	0.024	0.095	0.135	0.328	0.2218
50	204,354	17,360	8.495	4,804	917	0.191	128	0.027	0.140	0.110	0.277	0.1940
51	176,144	14,491	8.227	6,501	1,572	0.242	106	0.016	0.067	0.177	0.449	0.1965
52	89,352	7,558	8.459	3,326	753	0.226	66	0.020	0.088	0.184	0.440	0.1836
53	142,023	13,227	9.313	3,804	683	0.179	63	0.017	0.092	0.115	0.288	0.1768
54	169,159	14,426	8.528	6,129	1,471	0.240	206	0.034	0.140	0.197	0.425	0.1830
55	108,151	9,449	8.737	3,937	1,026	0.261	74	0.019	0.072	0.190	0.417	0.2089
56	274,123	23,324	8.509	6,139	759	0.124	283	0.046	0.373	0.090	0.263	0.1216
58	137,032	11,028	8.048	4,130	695	0.168	81	0.020	0.117	0.161	0.375	0.1456
59	888,558	51,345	5.778	16,379	2,380	0.145	583	0.036	0.245	0.094	0.319	0.1384
60	157,752	12,753	8.084	5,709	1,422	0.249	118	0.021	0.083	0.188	0.448	0.2037
61	119,955	10,429	8.694	3,672	938	0.255	70	0.019	0.075	0.162	0.352	0.2374
62	447,957	32,000	7.144	9,206	1,512	0.164	204	0.022	0.135	0.090	0.288	0.1501
63	210,433	16,537	7.859	5,092	635	0.125	158	0.031	0.249	0.120	0.308	0.1310
64	188,907	13,894	7.355	6,879	1,256	0.183	579	0.084	0.461	0.177	0.495	0.1447
65	85,472	6,114	7.153	2,596	534	0.206	187	0.072	0.350	0.177	0.425	0.1646
66	104,030	6,909	6.641	2,754	465	0.169	113	0.041	0.243	0.144	0.399	0.1473
69	291,283	18,736	6.432	8,044	2,326	0.289	235	0.029	0.101	0.191	0.429	0.2170
70	116,960	8,825	7.545	3,491	747	0.214	64	0.018	0.086	0.151	0.396	0.1835
71	269,521	21,977	8.154	6,457	1,019	0.158	146	0.023	0.143	0.104	0.294	0.1566
72	158,033	13,282	8.405	4,299	1,099	0.256	82	0.019	0.075	0.126	0.324	0.2341
73	149,663	9,290	6.207	3,162	450	0.142	101	0.032	0.224	0.153	0.340	0.1453
74	130,377	9,812	7.526	2,958	429	0.145	109	0.037	0.254	0.128	0.301	0.1515
75	1,243,566	76,306	6.136	34,411	12,806	0.372	1,387	0.040	0.108	0.174	0.451	0.2684
76	413,164	27,016	6.539	8,167	2,141	0.262	203	0.025	0.095	0.114	0.302	0.2325
77	136,328	12,331	9.045	5,674	1,304	0.230	79	0.014	0.061	0.214	0.460	0.1899
78	234,812	18,942	8.067	9,332	2,749	0.295	190	0.020	0.069	0.236	0.493	0.2189
79	140,159	12,424	8.864	3,822	612	0.160	54	0.014	0.088	0.110	0.308	0.1544

Appendix Table (cont'd). Detailed data per department

Department number	Total no. of births [1867–1899]	Total no. of MPFs	Total no. of MPFs / Total no. of births [1867–1899] (%)	Total no. of different surnames for MPFs	Expected no. of surnames disappearing under G–W	Expected % of surnames disappearing under G–W	No. of MPF surnames absent from INSEE file	% of MPF surnames absent from the INSEE file	% of MPF surnames absent from the INSEE file / expected % of surnames disappearing under G–W	Total no. of surnames / total no. of births [INSEE P1]	Total no. of surnames / Total no. of MPFs	Proportion of households with no children
80	218,211	17,066	7.821	5,655	1,157	0.205	134	0.024	0.116	0.128	0.331	0.1884
81	140,531	10,314	7.339	2,610	463	0.178	66	0.025	0.142	0.089	0.253	0.1643
82	71,804	5,379	7.491	2,498	458	0.183	64	0.026	0.140	0.176	0.464	0.1551
83	114,246	5,908	5.171	2,799	776	0.277	110	0.039	0.142	0.231	0.474	0.2169
84	100,208	6,237	6.224	2,457	533	0.217	52	0.021	0.098	0.177	0.394	0.1786
85	195,765	19,021	9.716	3,789	537	0.142	78	0.021	0.145	0.068	0.199	0.1583
86	137,494	11,969	8.705	4,238	688	0.162	97	0.023	0.141	0.126	0.354	0.1530
87	175,646	14,371	8.182	5,125	937	0.183	152	0.030	0.162	0.118	0.357	0.1584
88	179,922	15,815	8.790	4,387	863	0.197	105	0.024	0.122	0.115	0.277	0.1756
89	116,078	10,311	8.883	4,500	918	0.204	61	0.014	0.066	0.189	0.436	0.1722
90	35,264	1,794	5.087	1,103	230	0.209	37	0.034	0.161	0.268	0.615	0.1605

Sources: MPF file, INSEE surname file, data from the study 'La Statistique générale de la France' (volumes of *Annuaire statistique de la France*). Authors' calculations.



REFERENCES

- BACAËR N., 2011, *A short history of mathematical population dynamics*, London, Springer.
- BÉAUR G., MARIN B., 2011, La Statistique Générale de la France – Présentation, *L'Atelier du Centre de recherches historiques*, La Statistique Générale de la France. Retrieved from <http://acrh.revues.org/2891>
- BIENAYMÉ I.-J., 1845, De la loi de multiplication et de la durée des familles, *Société philomatique de Paris, Extraits des procès-verbaux des séances, 5^e série*, 10, 37–39 (Reprinted in *L'Institut*, 589, 13, 131–132).
- BONNEUIL N., 1989, Cohérence comptable des tableaux de la SGF: Recensements de 1851 à 1906, mouvements de la population de 1801 à 1906, *Population*, 44(4–5), 809–838.
- BONNEUIL N., 1997, *Transformation of the French demographic landscape, 1806–1906*, Oxford, Clarendon Press.
- BOULANGER P., 2003, La conscription en France au XX^e siècle: de l'armée de masse à l'armée professionnelle, *Revue internationale d'histoire militaire*. Retrieved from http://www.institut-strategie.fr/RIHM_83_26.htm
- BROUARD N., 1989, *L'extinction des noms de familles en France: une approche*, Paris, INED, Documents de travail, No. 27.
- CANDOLLE A. de, 1873, *Histoire des sciences et des savants depuis deux siècles: suivie d'autres études sur des sujets scientifiques en particulier sur la sélection dans l'espèce humaine*, Geneva, H. Georg.
- CHAREILLE P., DARLU P., 2013, Définition, identification et localisation de géohapax dans le corpus des patronymes français: comparaison de sources anciennes et récentes, *Rivista Italiana di Onomastica [RION]*, 19(2), 633–658.
- CHERVIN A., 1888, Sur la natalité en France, *Bulletins et mémoires de la Société d'anthropologie de Paris*, 1(1), 540–578.
- DAGUET F., 1995, Un siècle de démographie française, *Insee Résultats*, 434–435, December.
- DARLU P., DEGIOANNI A., RUFFIÉ J., 1997, Quelques statistiques sur la distribution des patronymes en France, *Population*, 52(3), 607–634.
- DARLU P., OYHARÇABAL B., 2006, Distribution des patronymes et structure des populations dans les Pyrénées occidentales, *Lapurdum*, 11, 137–165.
- DARLU P., RUFFIÉ J., 1992, L'immigration dans les départements français étudiée par la méthode des patronymes, *Population*, 47(3), 719–734.
- DARMON P., 2000, Une tragédie dans la tragédie: la grippe espagnole en France (avril 1918 – avril 1919), *Annales de démographie historique*, 2000-2 [Famille et parenté], 153–175.
- FARON O., 2002, Guerre(s) et démographie historique, *Annales de démographie historique*, 2002-1 [La population dans la Grande Guerre], 5–9.
- FESTY P., 1984, Effets et répercussions de la Première Guerre mondiale sur la fécondité française, *Population*, 39(6), 977–1000 [Republished in Rohrbasser J.-M. (ed.), 2014, *Bouleversements démographiques de la Grande Guerre*, Paris, INED, 201–236].

- GILLES H., 2010, La guerre 1914-1918: les morts par départements et par régions, in Denis G. (ed.), *Mémoire et trauma de la Grande Guerre: Bretagne, Catalogne, Corse, Euskadi, Occitanie*, Rennes, Centre de Recherche Bretonne et Celtique, Université de Rennes 2, 275–323.
- GILLES H., GUIRONNET J.-P., PARENT A., 2014, Géographie économique des morts de 14-18 en France, *Revue économique*, 65(3), 519–532.
- GUILLOT O., PARENT A., 2018, “Farewell life, farewell love”: Analysis of survival inequalities among soldiers who “Died for France” during World War I, *Population*, 73(3), 411–444.
- HÉRAN, F., 2014, Lost generations: The demographic impact of the Great War, *Population & Societies*, 510, April.
- HUBER M., 1931, *La population de la France pendant la Guerre*, Paris, Presses Universitaires de France.
- INSEE, 1985, *Registre français des noms patronymiques*.
- LAHAIE O., 2011, L'épidémie de grippe dite ‘espagnole’ et sa perception par l’armée française (1918-1919), *Revue historique des armées*, 262, 102–109.
- Les pertes des nations belligérantes au cours de la Grande Guerre, 1921, *Les archives de la Grande Guerre*, 7(19), 37–61.
- LOEZ A., MARIOT N., 2014, Tous égaux devant ‘l’impôt du sang’? L’approche spatiale de la mort de masse en 1914-1918 et ses enjeux. Un commentaire de ‘Géographie économique des morts de 14-18 en France’, *Revue économique*, 65(3), 533–542.
- PROST A., 2008, Compter les vivants et les morts: l’évaluation des pertes françaises de 1914-1918, *Le Mouvement Social*, 222(1), 41–60.
- PROST A., 2014, The dead, in Winter J. (ed.), *The Cambridge history of the First World War*, Vol. 3, Cambridge, Cambridge University Press, 561–591.
- ROHRBASSER J.-M. (ed.), 2014, *Bouleversements démographiques de la Grande Guerre*, Paris, INED.
- SANGOÏ J.-C., 1997, La guerre de 1914-1918 et l’évolution démographique française, in Caucanas S., Cazals R. (eds.), *Traces de 14-18*. [Actes du colloque de Carcassonne], Carcassonne, Les Audois, 147–155.
- VALLIN J., 1973, *La mortalité par génération en France depuis 1899*, Cahier no. 63, Paris, INED.
- VALLIN J., 1984, *Tables de mortalité du moment et par génération, 1899-1981 (Mise à jour provisoire des tables annexes du cahier n° 63)*, 3 Vols., Paris, INED.
- VALLIN J., MESLÉ F., 2001, *Tables de mortalité françaises pour les XIX^e et XX^e siècles et projections pour le XXI^e siècle*, Paris, INED.
- VAGNERON F., 2015, *Aux frontières de la maladie. L’histoire de la grippe pandémique en France (1889-1919)* (Doctoral dissertation), Paris, EHESS.
- WATSON H. W., GALTON F., 1875, On the probability of the extinction of families, *The Journal of the Anthropological Institute of Great Britain and Ireland*, Vol. 4, 138–144.
- WINTER J., 2004, Victimes de la guerre: morts, blessés et invalides, in Audoin-Rouzeau S., Becker J.-J. (dir.), *Encyclopédie de la Grande Guerre*, Paris, Bayard, 1015–1024.

Pierre DARLU, Pascal CHAREILLE • THE FIRST WORLD WAR AND THE DISAPPEARANCE OF SURNAMES IN FRANCE: A TRIAL ESTIMATION BASED ON THE GALTON–WATSON MODEL

The possible reduction of the stock of surnames attributable to the First World War (1914–1918) in France has yet to be explored. This article offers an estimate of the extent of this reduction by taking into account the random effects of the disappearance of surnames according to the Galton–Watson process, the distribution of the number of children per household, and the proportion of surnames of a given number of bearers. Estimates have been based on the French Ministry of Defence's electronic file, which includes French soldiers killed during the First World War, and on INSEE's surname file. While the human losses attributable to the war are unevenly distributed throughout the country, with a higher number north of a La Rochelle–Mulhouse line, a different geographical pattern emerges with greater extinctions in the Pyrénées, Corsica, and Brittany. The extinction of surnames actually represents only a small proportion of all the surnames of those who died for France (1.4%). By department, this observation is at odds with the Galton–Watson model, which predicts a larger proportion of extinctions due to the constitution of the two files and the formulated demographic hypotheses.

Pierre DARLU, Pascal CHAREILLE • LA GUERRE DE 1914-1918 ET LA DISPARITION DES NOMS DE FAMILLE EN FRANCE : UN ESSAI D'ESTIMATION À PARTIR DU MODÈLE GALTON-WATSON

Aucune étude en France ne s'est intéressée à l'éventuelle réduction du stock patronymique imputable à la guerre de 1914-1918. L'objet de ce travail est d'estimer l'ampleur du phénomène en prenant en compte les effets aléatoires de la disparition des noms selon le modèle d'extinction de Galton-Watson, la distribution du nombre d'enfants par ménage et la proportion de patronymes selon le nombre de porteurs. Les estimations, sur la France entière et par département, sont établies sur la base du fichier des « Morts pour la France » et du fichier Insee des noms de familles. Si les pertes humaines imputables à la Grande Guerre sont inégalement distribuées sur le territoire, avec un nombre de morts plus élevé au nord d'une ligne La Rochelle-Mulhouse, les pertes patronymiques sont géographiquement différemment réparties et plus nombreuses dans les Pyrénées, en Corse et en Bretagne. Pour autant, les patronymes effectivement disparus ne représentent qu'une proportion faible (1,4 %) de l'ensemble des patronymes des MPF. Par département, cette proportion est très variable et souvent en contradiction avec le modèle de Galton-Watson qui prévoit davantage de disparitions. Plusieurs explications sont avancées : elles tiennent au mode de constitution des deux fichiers et aux hypothèses démographiques formulées.

Pierre DARLU, Pascal CHAREILLE • LA GUERRA DE 1914-1918 Y LA DESAPARICIÓN DE APELLIDOS EN FRANCIA: UNA TENTATIVA DE ESTIMACIÓN A PARTIR DEL MODELO WATSON-GALTON

En Francia no ha habido ningún estudio sobre la reducción eventual del stock patronímico imputable a la guerra de 1914-1918. Nuestro objetivo es estimar la importancia del fenómeno tomando en cuenta los efectos aleatorios de la desaparición de apellidos según el modelo de extinción de Watson-Galton, la distribución del número de hijos por hogar y la proporción de apellidos según el número de portadores. Las estimaciones, para el total de Francia y por departamento, han sido hechas a partir del fichero "Morts pour la France" y del fichero de apellidos del INSEE. Si las pérdidas humanas imputables a la Gran Guerra están desigualmente repartidas geográficamente y son más importantes al norte de una línea La Rochelle-Mulhouse, las pérdidas patronímicas están diferentemente repartidas y son más numerosas en los Pirineos, en Córcega y en Bretaña. Sin embargo, los apellidos efectivamente desaparecidos representan sólo una pequeña proporción (1,4 %) del conjunto contenido en el fichero MPF. Por departamento, esta proporción varía mucho y frecuentemente en contradicción con el modelo de Watson-Galton, que prevé más desapariciones. Varias explicaciones son propuestas, que se refieren al modo de constitución de los dos ficheros y a las hipótesis demográficas.

Keywords: First World War, France, surnames, Morts pour la France, Galton-Watson process

Translated by James Tovey