Names | A Journal of Onomastics

The Concentration of Personal Names in Late Medieval Florence

Stefano Ciliberti
École Pratique des Hautes Etudes, FRANCE

ans-names.pitt.edu
ISSN: 0027-7738 (print) 1756-2279 (web) Vol. 74 No. 3, Summer 2026 DOI 10.5195/names.2026.2848

Abstract

This study applies quantitative methods to explore the evolution of personal names in late medieval Florence. Using large historical datasets—including the "Libro di Montaperti" (1260), the "Catasto" (1427), and the "Tratte" (1282–1532)—I examine how the concentration of personal names changed over time. After assessing several possible indicators of diversification, I show that many are affected by finite-size effects and selection bias, which can distort historical interpretation. The homonymy rate, defined as the probability that two randomly chosen individuals share the same name, is identified as a robust and scale-independent metric. Applied across the Florentine sources mentioned above, this measure reveals a gradual but significant process of concentration beginning in the early fourteenth century and accelerating strongly around the time of the Black Death. The results provide a quantitative basis for dating and characterizing the long-term evolution of naming practices, illustrating how statistical reasoning can enrich the study of cultural and historical phenomena.

Keywords: anthroponymy, first names, homonymy rate, Middle Ages, statistics, Florence, Italy

1 Introduction

Anthroponymy belongs to the linguistic field of onomastics, yet it has long attracted the attention of scholars from a variety of disciplines, including history. Despite the abundance of medieval sources containing personal names, and the relative ease with which such data can be organized into databases, the interest of medieval historians in anthroponymy has long remained limited. This is particularly striking, since it was during the Middle Ages, roughly between the tenth and fifteenth centuries, that the two-element naming system still in use today first emerged, developed, and became established.

Tuscany, a north-central Italian region with Florence as its capital, offers a particularly illuminating case study within the broader European context. Owing to its early urbanization, its flourishing mercantile economy, and the richness of its archival sources, Tuscany exemplifies the institutional and cultural conditions that led to the stabilization of naming practices. As Bec (1967) has noted, Florentine mercantile writing frequently emphasized family lineage, moral inheritance, and honor, placing particular importance on the continuity of name and reputation. This concern with memory and posterity is directly connected to the emergence of stable anthroponymic systems and helps to contextualize the strong interest in self-identification and naming conventions.

In a foundational article, Herlihy (1988) examined this development in depth, analyzing more than three centuries of onomastic data drawn from a variety of Tuscany sources. These include the list of citizens of Pistoia which swore to uphold a peace with Bologna in 1219; the list of Florentines that marched against Siena in 1260; and lists of family heads of different cities, those buried at one particular church in Florence, and citizens matriculated into one guild or scrutinized for some communal offices. Herlihy showed that the adoption of fixed family names in Tuscany occurred both earlier and more systematically than in many other regions of Europe, due in part to the rise of bureaucratic record-keeping practices and to the growing importance of family identity in civic and economic life. The use of patronymic markers (e.g., *di Giovanni* 'son of Giovanni'), toponymic elements (*da Prato, Borgognoni* 'coming from Prato, Burgundy'), as well as diminutives or hypocoristics (*Cecco, Neri,* short for *Francesco* and *Raineri*, respectively), became common and often solidified into stable hereditary surnames. These forms reflect both the linguistic habits of the Tuscan vernacular dialect and the documentary needs of institutions such as the commune, the guilds, and the notarial chancery.

Molho (2000) has already shown that from the mid-fourteenth century onward, the use of family names tended to become generalized. The relationship between surnames and social status in Renaissance Tuscany contributed to the political visibility and economic consolidation of elite families, whose names came to constitute a genuine form of symbolic capital. However, even at its height, in the second half of the fifteenth century, the dissemination of family names included at most half of the urban population.

As for personal names, the Tuscan data reveal a marked process of concentration around a small number of names, despite an initially broad repertoire. According to Herlihy (1988), this profound transformation in the management of the anthroponymic stock took place between the thirteenth and fourteenth centuries. Saints' names came to dominate by the early fifteenth century: the majority of personal names in all Tuscan cities were of Christian origin, whereas two centuries earlier they had occupied only a marginal place. Similar

DOI 10.5195/names.2026.2848 Vol. 74 No. 3, Summer 2026 ISSN: 0027-7738 (print) 1756-2279 (web)

considerations, though following somewhat different chronological trajectories, largely apply to other European regions, as attested, for instance, by studies on English names (Chetwood 2018), Italian names (Bizzocchi 2014), and French names in a specific region (Barthélemy 1990).

Based on these observations, the traditional narrative identifies the following logical sequence: the growing number of individuals sharing the same personal name, reflecting an overall reduction in the available name pool, led to a steady increase in homonymy, which in turn paved the way for the emergence and widespread adoption of family names. While these facts are well established, the presumed causal relationships between them may nonetheless be reconsidered through more detailed analysis.

In his major contribution to the *Genèse médiévale de l'anthroponymie moderne* (Medieval Origins of Modern Anthroponymy), Chareille (2008) demonstrated how effective a statistical approach can be in addressing problems related to anthroponymy. At the same time, he emphasized that numerical data must always be interpreted with caution. Figures can easily mislead if considered superficially, and extracting reliable information requires situating them within a sound theoretical framework.

The main goal of this paper is to quantitatively evaluate the patterns of the concentration process, described above, around a limited number of personal names in late medieval Florence. Specifically, the goal is to determine its extent, its temporal development, and whether external conditions may have influenced it. To this end, I conducted a numerical analysis of the data and I employed statistical indicators, after first discussing their relevance for this type of study.

The core of the paper is developed in the next section, which is organized into four subsections. The first introduces the "Catasto" database, which will serve as the basis for the numerical experiments presented in the following subsection, aimed at identifying the most appropriate metric for quantifying name concentration. Then, after introducing additional data sources for the city of Florence, I present my results on the temporal evolution of onomastic concentration in late medieval Florence. These findings are summarized in the concluding section, where further research directions are also presented.

2 Measuring Concentration

2.1 The "Catasto" Database

One of the most important sources for the study of the Tuscan population in the late Middle Ages is the well-known "Catasto" dataset, the Florentine land and property tax register compiled between 1427 and 1430. It covers not only the city of Florence, but also the towns and rural areas under its control. The document records personal, familial, and patrimonial data for each household in the region. For the city of Florence alone, the dataset includes approximately 10,000 householders, of whom 8,242 are men. The data were made available in a machine-readable format as part of a French–American project led by Herlihy and Klapisch-Zuber in the 1970s. The reader is referred to their individual works, depending on his/her interest in raw data (Herlihy et al. 2002), an initial introduction (Klapisch-Zuber et al. 2016), or a detailed and complete study (Herlihy & Klapisch-Zuber 1985).

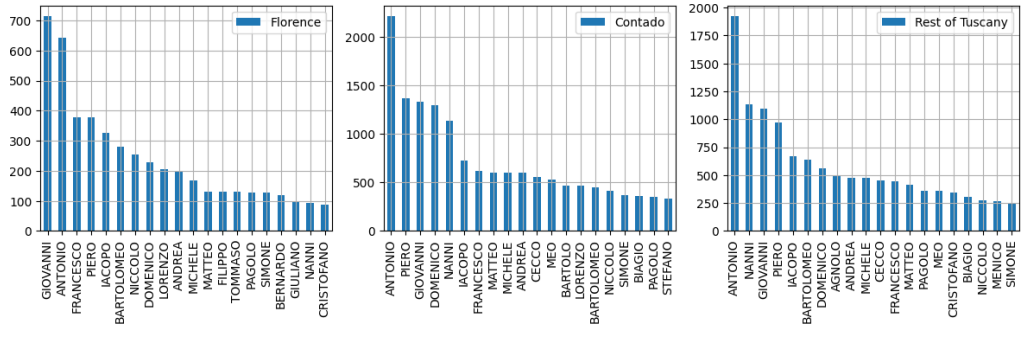

Figure 1: The 20 most popular male names in 1427 in Tuscany.

ISSN: 0027-7738 (print) 1756-2279 (web) Vol. 74 No. 3, Summer 2026 DOI 10.5195/names.2026.2848

The Florentine archives relating to the "Catasto" of 1427–1430 are divided into three series: the "portate", or self-declarations, submitted by individuals and describing their property and household composition; the "campioni", registers in which these declarations were copied and the fiscal assessments calculated; the "sommari", alphabetical lists of tax-payers indicating taxable wealth, household size, and the number of adult males subject to taxation. None of these series is complete. Data extraction relied primarily on the "campioni", while gaps were filled using the "sommari" and "portate", which were also employed to verify incomplete or uncertain information.

It is important to note that the compilers of the database, working from the above archival sources, chose to preserve all orthographic distinctions, even the slightest ones, between personal names. Thus, for example, we find 229 occurrences of *Niccolò*, 41 of *Niccolaio*, 7 of *Niccola*, on top of 2 *Niccoletto*, 1 *Niccolino*, and 1 *Niccoluccio*: these seven versions correspond respectively to 81.49%, 14.59%, 2.49%, 0.71%, 0.36%, and 0.36% of the 281 individuals whose name can be seen as a variant of *Niccolò*. Similarly, we have many *Gherardo* but also some *Gherardino* and *Gherarduccio*; almost as many *Uberto* as *Ubertino*; and so on. This is a legitimate editorial decision and, even if there is often a clear prevalence in these mini lists, it may have a non-negligible effect on statistical outcomes.

Using the list of householders provided in the "Catasto" of 1427, we present in figure 1 the twenty most common male names in Florence, its "contado" (the rural area surrounding the city), and the rest of Tuscany. As men are far more numerous than women in this dataset, we extract a richer statistical description. It is noteworthy that while in Florence there are at least seven times as many occurrences of *Giovanni* as of *Nanni*, these two names are equally represented in the other areas. A similar pattern can be observed for pairs such as *Francesco / Cecco, Bartolomeo / Meo*, and other combinations of names and their diminutive or hypocoristic forms. This discrepancy likely reflects differences in the way names were recorded.

The inconsistent treatment of hypocoristic forms may be due to more meticulous clerks in the main city, who often replaced diminutives with their formal equivalents, or to the systematic absence of "campioni" in rural areas. In any case, this heterogeneity precludes extending our statistical analysis to the whole of Tuscany, since the data would not be comparable. In what follows, we therefore focus exclusively on Florence.

2.2 Choosing the Best Metric

Let N be the size of the population and K the number of distinct names, which we order in some way (for instance alphabetically). We note by $N_1, N_2 \ldots N_K$ the number of individuals bearing respectively names *1, 2,.. K*, so that the sum of all N_k is equal to N. The ratio N/K represents the average number of individuals per name, while $s = 100\ K\ /\ N$ indicates the number of distinct names available per 100 individuals. These two indicators have been widely used in anthroponymic studies, from the pioneering works of 19th-century scholars (Aebischer 1928; Brattö 1953) to the recent interpretations as imbalance index (Zhao et al. 2021) and the attempts to explicitly model the *K-N* relationship (Miyazima et al. 2000).

From an empirical perspective, it has become clear that these two ratios (s and its reciprocal) suffer from important limitations which make them unsuitable for describing changes over time. The two measures clearly convey the same information. Yet the former has the advantage of being expressed as a bounded quantity: the maximum value $s = 100$ corresponds to the situation in which everyone bears a different name, whereas the minimum $s = 100/N$ is obtained when the entire population shares the same name.

However, when N is smaller than 100, the interpretation of s can be misleading: for a (trivial) population of 20 individuals all bearing the same name, s would suggest that there are five distinct names per 100 people. As we shall see, though, the main weakness of this ratio lies in its strong dependence on sample size: s is in practice a decreasing function of N. It should therefore not be used to measure or compare diversification in samples of markedly different sizes.

To investigate the size-effect bias in name-related measurements, I now imagine only having access to a fraction of the "Catasto" list of male names, rather than to the complete document. In a situation typical for historians, the subset preserved is rarely selected at random but rather according to certain (known or unknown) criteria. In what follows, I assume—optimistically—that the sample available to us is free of informational bias. Even under such conditions, what happens to our measure of diversity?

I conducted a numerical test consisting in randomly selecting a subset of n entries (and thus of personal names) from the database, and computing the ratio s for this sample. I repeated this operation several times, each time choosing a different subset of the same size n, and then calculate the empirical mean and standard deviation of s. I subsequently varied n and repeated the experiment.

Figure 2 shows the mean value of s as a function of n, with error bars. The expected analytical behavior, that is, the mathematical expectation of distinct names in a sample of size n, can be computed using standard combinatorial techniques. It is also plotted in the figure: the agreement with the empirical data is perfect.

DOI 10.5195/names.2026.2848 Vol. 74 No. 3, Summer 2026 ISSN: 0027-7738 (print) 1756-2279 (web)

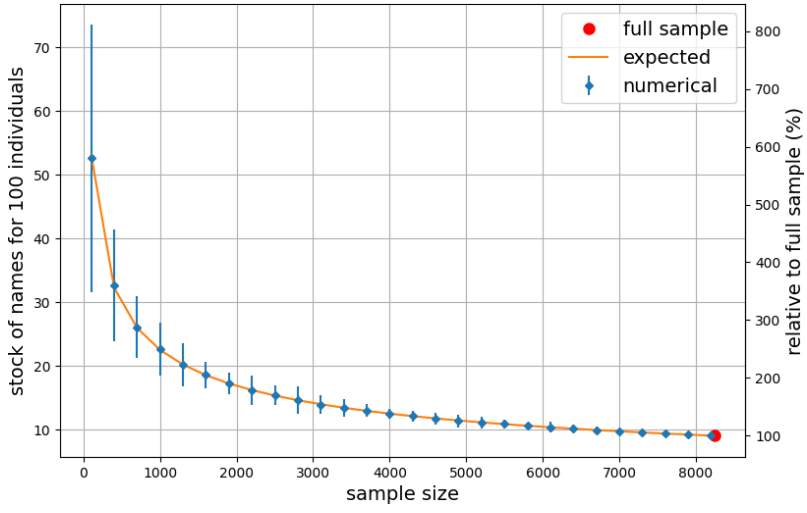

Figure 2: A study of the finite-size effect for *s*.

In the total population of 8,242 male heads of household, 747 distinct personal names are recorded, corresponding to approximately $s \simeq 9$ names per 100 individuals. If I had access to only half of this population ($n \simeq 4{,}100$ in the figure), I would measure for s a value of about 13—roughly 50% higher than the true value! For a sample representing 20% of the total population, the measured value would be about twice the correct value, and up to six times higher for a small sample covering only 1% of the population. In other words, the bias associated with sample size is particularly strong for this indicator. As a consequence, I do not recommend its use when a significant portion of the data is missing, or when one wants to keep track of the diversity of a population whose size changes over time.

Other statistical indicators (see figure 3) can be used to address the same question of concentration while being less sensitive to finite-size effects. I perform the same type of experiment as before for the following quantities:

- The effective number of names, $K_{\text{eff}} = N^2 / \Sigma_k N_k^2$. Widely used in the social sciences and economics to quantify population diversity, this measure is known under various names, and its intuitive meaning makes it a convenient indicator. Its maximum value, $K_{eff} = K$, is reached when each name is borne by the same number of individuals ($N_k = N/K$ for all k), whereas $K_{eff} = 1$ corresponds to a population entirely concentrated on a single name ($N_k = N$ for one value of k, and 0 for all others). As an intermediate example, if more than 90% of the population bears one of three names distributed in roughly equal proportions, and the remainder is spread over many different names (say, 10), the resulting K_{eff} would be between 3 and 4, while $K = 13$.

- The median rank, defined as the minimal number of distinct names required to account for half of the population. This is a particularly stable measure, as it depends on the central portion of the distribution, and this makes it less sensitive to outliers or rare events.

ISSN: 0027-7738 (print) 1756-2279 (web) Vol. 74 No. 3, Summer 2026 DOI 10.5195/names.2026.2848

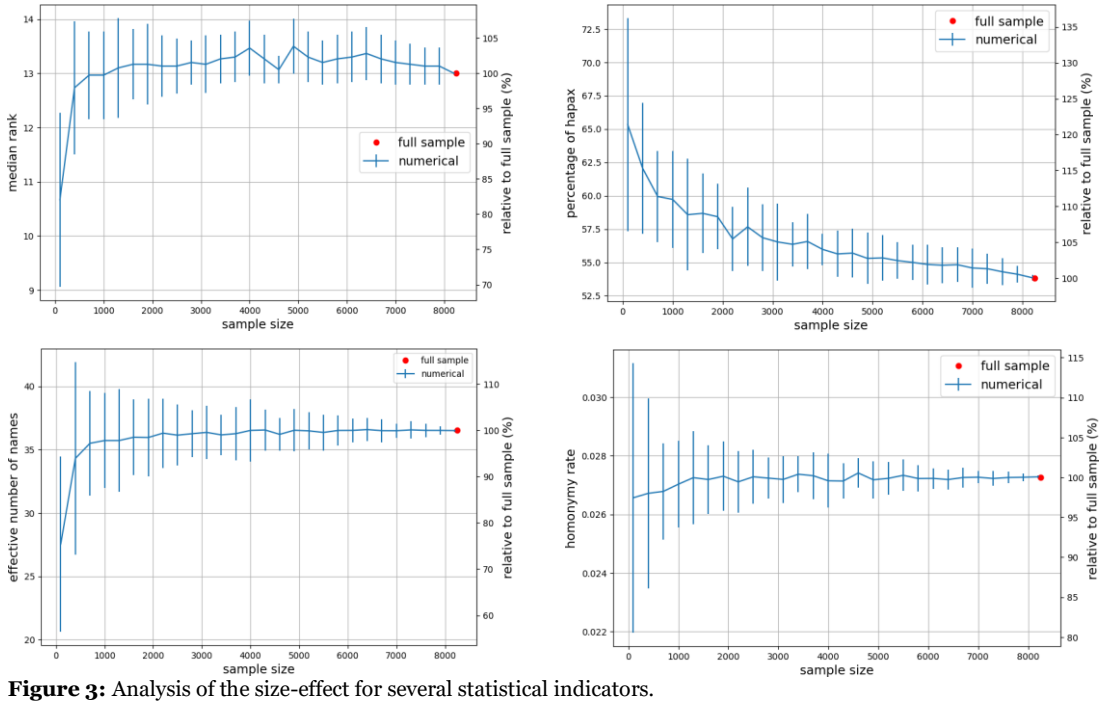

Figure 3: Analysis of the size-effect for several statistical indicators.

- The homonymy rate H, which measures how likely it is that two randomly chosen individuals have the same name. To compute this rate, we consider all possible pairs of people in the population and ask what fraction of those pairs share an identical name. For each name, we count how many distinct pairs can be formed among the people who bear that name. We then add these counts over all names. Finally, we divide this total by the number of all possible pairs that can be formed from the entire population. The resulting ratio is the probability that two randomly selected individuals have the same name. In formulas: $H = \Sigma_k\, C(N_k,2) \,/\, C(N,2)$, with $C(a,b)$ denoting the binomial coefficient. For large values of N and all the N_k, it is easy to show that H goes to $1/K_{eff}$.

- The percentage of "hapax", where a "hapax" is defined as a name appearing only once in the population. The higher this proportion, the more diversified the onomastic repertoire. However, as this indicator focuses on a specific tail property of the distribution, it is numerically less stable and should be used with caution when dealing with incomplete datasets.

The results shown in figure 3 indicate that finite-size effects are numerically less pronounced than those observed in figure 2 for the number of available names. The least robust indicator is the percentage of "hapax", which would be about 10% higher in a sample representing 10% of the total population—whereas, under the same conditions, the increase for s exceeded 200%. Moreover, the graphs show that when working with an exceedingly small sample—say, about 1% of the total population—all these indicators become significantly unreliable, with the notable exception of the homonymy rate. The relative robustness of the latter stems from the fact that its computation depends on the total number of possible pairs within the population, which scales as N^2 instead of just N as for the other indicators: this makes the empirical calculation of the homonymy rate H inherently less sensitive to small-sample effects.

Finally, it should be recalled that, in this experimental protocol, sampling was performed uniformly at random, whereas in practice historical sources almost always exhibit strong selection bias. To assess the potential impact of such bias on our measurements, one should therefore consider the error bars carefully and adopt the most conservative assumption (i.e., the extreme values within each bar). For all these reasons, I consider the homonymy rate to be the most reliable and prudent indicator of name diversity, and this is the measure I shall employ in the following.

DOI 10.5195/names.2026.2848 Vol. 74 No. 3, Summer 2026 ISSN: 0027-7738 (print) 1756-2279 (web)

2.3 The "Tratte" Database

All the indicators presented in the previous paragraph—along with several others—were calculated for two Tuscan sources by Chareille (2008): the "Libro di Montaperti" (Paoli 1889), which lists the names of the 6,204 Florentines involved in the war against Siena in 1260, and the Florentine "Catasto" of 1427, introduced in section 2.2. From this detailed analysis, it emerges that the onomastic diversity of the earlier source had, in fact, been underestimated by a previous study (Brattö 1953), when compared with the later one. In other words, in this particular case, the traditional thesis that there was a decline in the Florentine anthroponymic repertoire and an increasing concentration between 1260 and 1427 is confirmed. The data show that despite the increase in the homonymy rate, the proportion of "hapax" remains essentially unchanged. This observation suggests a strong polarization around a few dominant personal names rather than a simple impoverishment of the available stock—one might say it reflects a new mode of management. In the same vein, the cartulary of the Abbey of Lézat, in Occitania, attests to an intensification of onomastic concentration between the 10th and 12th centuries, without the disappearance of "rare" names (Mousnier 2002).

Can this phenomenon be dated precisely? As mentioned in the introduction, Herlihy (1988) observed that Florentine data indicate the most marked reduction in the number of favored personal names occurred roughly between 1280 and 1320, with the trend continuing throughout the late Middle Ages. For other regions, like the Vendôme in France, we know that the origins of the process are somewhat earlier, between the 11th and 13th centuries (Barthélemy 1993). The emergence of this concentration before the adoption of the two-element anthroponymic system likely varied from one region to another. and was by no means uniform—thus calling for a more nuanced explanation than the simplistic view that concentration itself directly caused the shift to the binomial naming system.

The "Tratte" database (Herlihy et al. 2002) offers a new perspective for advancing this analysis. It records all the highest executive offices of the Florentine Republic (the "Tre Maggiori"), composed of the "Gonfaloniere di Giustizia" and the "Priori", assisted by a notary, as well as two consultative councils: the "Buonuomini", representing the city's neighborhoods, and the "Gonfalonieri di Compagnia", who commanded the urban militias. On average, we have about 100 to 150 recorded names per year, covering the entire period from 1282 to 1532: these data provide a valuable resource for addressing the chronological dimension of the onomastic concentration phenomenon. The main difficulties lie in the relatively small numbers and in the strong heterogeneity of the records over time: the number of "Priori" varies throughout this long period, and the consultative councils are incompletely documented in the database—especially before 1328.

Indeed, although the database records the complete number of Priori from 1282 onward, meaningful statistics are only available beginning in 1329. In fact, an electoral reform enacted in 1328 introduced the selection of men for the highest offices by lot, following periodic screenings of eligible candidates. While the "Buonuomini" and "Gonfalonieri di Compagnia" already existed before that date, their first official registration dates from the term beginning in December 1328 (Herlihy et al. 2002). Prior to this reform, the data are therefore much more limited—generally far fewer than 100 names per year. Also, recalling the results of the previous study concerning very small samples (i.e., 1% of the total population or less), we consider most of the statistical indicators to be biased and unstable before 1328. The most reliable statistical indicator of diversification in these cases, as shown above, is the homonymy rate.

We also note that the editorial choices made in transcribing personal names differ in part from those adopted in the "Catasto". Certain variants were deliberately grouped under a single form: for instance, *Paolo* includes *Pagolo* and *Paolino*; *Niccolò* encompasses *Nicco* and *Niccola* (but not *Niccoletto* nor *Niccolino*); *Francesco* covers *Franco* and *Franceschino*; *Boccaccio* subsumes *Boccaccino*. Conversely, for reasons that are difficult to discern, distinctions are maintained between *Orlando* and *Orlandino*, *Agnolo* and *Agnolino*, and similar cases. These groups typically bring together standard forms with their diminutive, affectionate, or dialectal variants—names that share a common root but differ in suffix or local usage.

Beyond these editorial choices and fluctuations in the number of recorded entries, we believe it is crucial to assess whether the "Tratte" data exhibit a systematic bias. To a degree that varies with the political context, the individuals listed in the "Tratte" belong to the wealthier classes and therefore do not necessarily reflect the composition of the Florentine population as a whole. An illustrative example is provided in figure 4, showing over time the percentage of officeholders without a family name. This proportion is roughly one in ten in the earliest period, rises during the central decades of the 14th century, and then declines steadily from the 1380s through the 16th century. This irregular, non-monotonic and thus somewhat counterintuitive pattern can be explained by shifts in the political class' social composition over the course of the period. In other words, it is not possible to infer the overall frequency of family names in Florence from this chart alone, since

ISSN: 0027-7738 (print) 1756-2279 (web) Vol. 74 No. 3, Summer 2026 DOI 10.5195/names.2026.2848

it primarily reflects the alternation between regimes more or less deliberately open to the "popolo" (lit. "people"), but used here to designate the predominant class of non-aristocratic, economically prominent urban entrepreneurs.

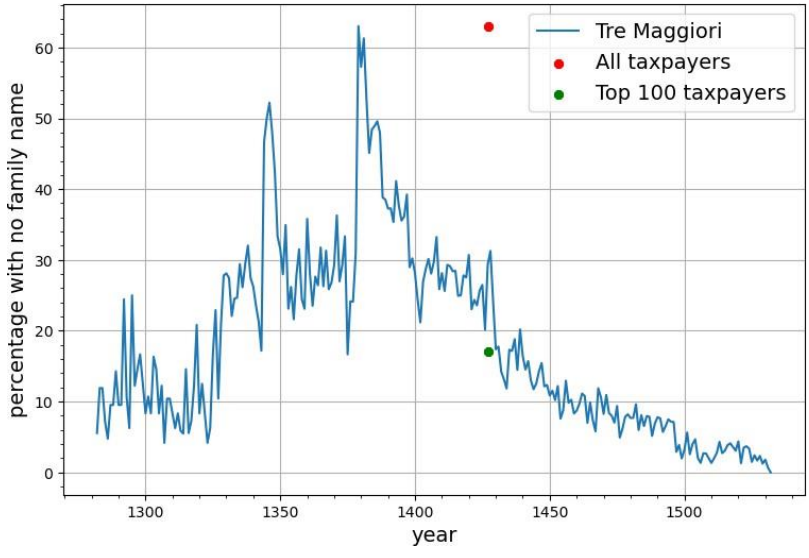

Figure 4: The percentage of individuals in Florence with no family name.

For instance, the peak visible in the curve coincides with the Ciompi Revolt of 1378, which briefly brought to power social groups traditionally excluded from state governance (Brucker 1968). From the early 15th century onward, the composition of the political elite appears to have stabilized, supporting the hypothesis of an increasingly widespread use of family names (Najemy 2006). In general, however, this empirical pattern arises from two distinct phenomena. The first is the gradual adoption of family names, likely increasing monotonically over time; the second concerns fluctuations in the social composition of the ruling elites. The two processes that are, in principle, difficult to disentangle.

In the particular case of family name, the quantification of this phenomenon remains limited to the upper strata of society, as illustrated by the two data points in the same figure 4. A simple comparison between this figure and the data from the 1427 "Catasto" clearly shows that the sample drawn from the offices of the "Tre Maggiori" represents primarily the wealthiest families (only 17.21% without a family name for the top 100 taxpayers), whereas in the Florentine population as a whole the proportion of individuals without a family name reaches 63.43% (referred to as "All taxpayers" in the same figure).

The picture changes dramatically once we turn our attention to personal names rather than family names. Indeed, as we shall verify in the next subsection, personal names can be freely chosen from a broad, context-dependent set; hence, one may expect them to be less affected by the selection bias inherent to the ruling elites.

DOI 10.5195/names.2026.2848 Vol. 74 No. 3, Summer 2026 ISSN: 0027-7738 (print) 1756-2279 (web)

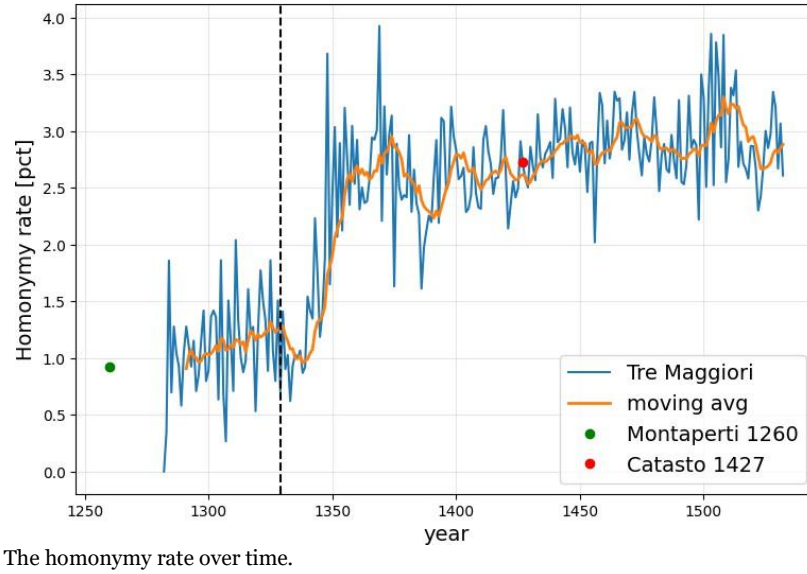

Figure 5: The homonymy rate over time.

2.4 The Evolution Of Name Diversity

In figure 5, we plot the yearly homonymy rate from the "Tratte" database, covering the period 1282–1532. A 10-year moving average is also shown to smooth out short-term fluctuations and emphasize long-term trends. In addition, the homonymy rate is computed for the populations recorded in the "Libro di Montaperti" (1260) and in the "Catasto" (1427). As anticipated, the 1427 value—calculated over the entire Florentine population—aligns perfectly with that measured for the smaller sample of "Priori" from the same period, confirming that this metric is not subject to specific bias. We also note that the 1260 value appears fully consistent with the data from the 1280s and 1290s. The pairwise homonymy rate thus proves to be a robust and informative indicator of population diversification, even when based on relatively small statistical samples.

The curve shows a sharp increase, rising from 1.12% before 1330 to 2.49% by the mid-1350s, and then continuing to grow steadily to around 2.95% by the end of the 15th century. The surge observed in the late 1340s was almost certainly amplified by the Black Death: current estimates (Cohn 2023) suggest that the 1348 plague caused the death of approximately one half, and possibly up to two thirds, of the population of Florence: this effectively acted as a partial demographic reset. The underlying process was likely already underway at a slower pace and was then accelerated by the catastrophic mid-century event. From a quantitative perspective, we can conclude that the concentration increased by a factor 2.5 or, equivalently, the diversification has been reduced to 40%.

It is important to note that this change in the homonymy rate is quite substantial. To gauge what might reasonably be expected, I performed numerical simulations for a population of N=10,000 individuals whose name frequencies follow Zipf's law (Zipf 1949): in layman's terms, this is a highly skewed distribution in which a small number of names are very common, while most names are rare; more precisely, the frequencies decrease approximately in inverse proportion to their rank. I varied the number of distinct names K between 100 and 3,000. The resulting pairwise homonymy rates range from about 2% to 6%, corresponding respectively to the maximum and minimum values of K used in the simulations. Although Zipf's law may not perfectly model this specific population, and the absolute values of the homonymy rates should therefore be interpreted with caution, their order of magnitude is nevertheless meaningful. In this light, the observed increase from 1% to 3% represents a remarkably large shift.

Other indicators, which we deem reliable only from 1329 onward, show similar behaviours. The effective number of names goes from 60 to 35 in the transition years 1330–1350, then slowly down to 30 in the following decades. The median rank, which is another "bulk" metric, has a very similar trend, being reduced by almost 50% in the same time period. By contrast, the decline in the fraction of "hapax", a "tail" metric, is much more limited: from 0.71 before the plague, it falls to 0.60 and then 0.55 in the following decades, corresponding to a total reduction of barely 20%. These results therefore confirm that the concentration around a

ISSN: 0027-7738 (print) 1756-2279 (web) Vol. 74 No. 3, Summer 2026 DOI 10.5195/names.2026.2848

limited number of personal names is both real and quantitatively significant (a reduction of a factor between 2 and 2.5 in the diversification according to different metrics); nonetheless, that change is not accompanied by a proportional contraction in the overall stock of available names.

What about the chronology? My results, particularly those shown in figure 5, do not support Herlihy's hypothesis of a sharp decline in the diversity of personal names between the 13th and 14th centuries. Instead, the evidence points to a gradual transformation: one that began in the early 14th century, remained incomplete through the 1330s, and accelerated sharply in the two decades around the Black Death in mid-century.

3 Summary and Conclusion

In this study, I have applied a reasoned statistical approach to address a question traditionally situated within the domain of the humanities: how did European societies evolve from the early Middle Ages—with their high degree of name diversification—to the late Middle Ages, when naming practices exhibited much greater concentration?

By examining how to measure diversification in a corpus of personal names, I have shown that many intuitive indicators can be heavily distorted by finite-size effects or by the inevitable selection bias affecting historical data. The same analysis allowed me to identify the homonymy rate as a particularly robust indicator: its mathematical definition makes it inherently less sensitive to variations in sample size, and its interpretation remains stable even when dealing with heterogeneous datasets. Once this measure was validated, it could be applied confidently to distinct sources. Then it enabled a consistent reconstruction of the century-long process of name concentration in Florence. The results show that the concentration of personal names increased more than twofold between the late thirteenth and the fifteenth centuries. Even if the process appears to have begun well before the mid-fourteenth century, it was sharply accelerated in the decades surrounding the Black Death, which likely acted as a demographic and cultural catalyst.

It would be valuable to gain a deeper understanding of the mechanisms underlying this phenomenon. Among the various factors that may have contributed to it, the growing influence of the Church undoubtedly played a central role. As Beck (2002) has shown for rural France, the spread of infant baptism encouraged parents to favor the names of the apostles over those of local saints, leading to a progressive contraction of the stock of early medieval names.

The role of immigration also deserves careful consideration. In a dynamic, trade-driven, and mercantile center such as late medieval Florence, this factor can hardly be regarded as negligible. Laliena (2002), for example, has demonstrated how naming practices were shaped by the migration of populations from southern France into medieval Aragon, suggesting that population movements may have significantly affected the composition of the local name pool.

A natural extension of our quantitative approach would be to examine the full distribution of personal names using mathematical models. In modern societies, such distributions often follow scale-free power laws, and a preliminary inspection of our data points to a similar pattern. Zanette and Manrubia (2001) have successfully applied a simple stochastic process, based on Simon's model (Simon 1955), to account for the distribution of family names. A comparable, albeit more refined, framework could be developed to model the distribution of personal names, shedding light on factors such as immigration dynamics, the persistence of family traditions, and preferences for originality in naming practices. We leave this line of inquiry for future research.

AI Disclosure Statement

No AI Tools or Technology were used to conduct the research or write this article.

Acknowledgments

I'd like to thank Stéphane Péquignot, Alma Poloni and Paolo Tomei for discussions and comments on this work. Thanks also to Prof. I.M. Nick for her support in the final editing stage of the manuscript.

References

Aebischer, Paul. 1928. *Essai sur l'anthroponymie catalane du XIe au XIIe siècle*. [Essay on the Catalan Anthroponymy from 11th to 12th century]. Barcelona: Publications de l'Oficina Romanica.

Barthélemy, Dominique. 1990. "Vendômois: le système anthroponymique (Xe - milieu XIIIe siècles)". [Vendôme: The Anthroponymic System (10th–Mid 13th Centuries.] *Genèse médiévale de l'anthroponymie moderne. Tome I*. [Medieval Origins of Modern Anthroponymy. Volume I]. Edited by Monique Bourin, 35–60. Tours: Presses universitaires François-Rabelais.

Barthélemy, Dominique. 1993. *La société dans le comté de Vendôme, de l'an mil au XIVe siècle*. [The Society in the Vendôme County, from the Year 1,000 to the 14th Century]. Paris: Fayard.

Bec, Christian. 1967. *Les marchands écrivains : Affaires et humanisme à Florence, 1375–1434*. [Merchants Writers: Business and Humanities in Florence, 1375–1434]. Paris: Mouton.

Beck, Patrick. 2002. "Personal Naming Among the Rural Populations in France at the End of the Middle Ages". *Personal Names Studies of Medieval Europe: Social Identity and Familial Structures*. Edited by George T. Beech, Monique Bourin, and Pascal Chareille, 143–156. Kalamazoo: Medieval Institute Publications.

Bizzocchi, Roberto. 2014. *I cognomi degli Italiani. Una storia lunga 1000 anni*. [The Family Names of Italians. A 1,000-Year Long Story]. Bari: Laterza.

Brattö, Olof. 1953. *Studi di antroponomia fiorentina. Il Libro di Montaperti (An. MCCLX)*. [Studies of Florentine Anthroponymy. The Book of Montaperti (Year 1260)]. Göteborg: Elanders Boktryckeri Aktiebolag.

Brucker, Gene A. 1968. "The Ciompi Revolution". *Florentine Studies. Politics and Society in Renaissance Florence*. Edited by Nicolai Rubinstein, 314–356. London: Faber.

Chareille, Pascal. 2008. *Genèse médiévale de l'anthroponymie moderne. Tome VI : Le nom, histoire et statistiques*. [Medieval Origins of Modern Anthroponymy: Volume VI: The Name, History, and Statistics]. Tours: Presses Universitaires François-Rabelais.

Chetwood, James A. 2018. "Re-Evaluating English Personal Naming on the Eve of the Conquest". *Early Medieval Europe* 26, no. 4: 518–547.

Cohn, Samuel. 2023. "The Black Death. Collapse, Resilience, and Transformation". *How Worlds Collapse: What History, Systems, and Complexity Can Teach Us About Our Modern World and Fragile Future*. Edited by Miguel Centeno, Peter Callahan, Paul Larcey, and Thayer Patterson, 192–206. New York: Routledge.

Herlihy, David, and Christiane Klapisch-Zuber. 1985. *Tuscans and their Families: A Study of the Florentine Catasto of 1427*. New Haven: Yale University Press.

Herlihy, David. 1988. "Tuscan Names, 1200–1530". *Renaissance Quarterly* 41, no. 4: 561–582.

Herlihy, David, Christiane Klapisch-Zuber, Robert Burr Litchfield, and Anthony Molho, eds. 2002. *Online Catasto of 1427, Version 1.3*. Florentine Renaissance Resources/STG: Brown University, Providence, R.I. Accessed on March 24, 2026. https://cds.library.brown.edu/projects/catasto/

Herlihy, David, Robert Burr Litchfield, Anthony Molho, and Roberto Barducci, eds. 2002. *Online Tratte of Office Holders, 1282–1532*. Florentine Renaissance Resources/STG, Brown University, Providence, R. I. Accessed March 24, 2026. https://cds.library.brown.edu/projects/tratte/

Klapisch-Zuber, Christiane, Béatrice Marin, and Nicolas Veysset. 2016, November 30. "Le catasto florentin de 1427–1430: Présentation générale de l'enquête et du code". [The Florentin Catasto of 1427–1430: General Presentation of the Research Project and of the Code]. *L'Atelier du Centre de Recherches Historiques. Les Enquêtes Collectives du CRH*. Accessed March 24, 2026. http://journals.openedition.org/acrh/7458.

Laliena, Carlos. 2002. "Personal Names, Immigration, and Cultural Change: Francos and Muslims in the Medieval Crown of Aragon". *Personal Names Studies of Medieval Europe: Social Identity and Familial Structures*. Edited by George T. Beech, Monique Bourin, and Pascal Chareille, 121–130. Kalamazoo: Medieval Institute Publications.

Miyazima, Sasuke, Youngki Lee, Tomomasa Nagamine, and Hiroaki Miyajima. 2000. "Power-Law Distribution of Family Names in Japanese Societies". *Physica A* 278, nos. 1–2: 282–288.

ISSN: 0027-7738 (print) 1756-2279 (web) Vol. 74 No. 3, Summer 2026 DOI 10.5195/names.2026.2848

Molho, Anthony. 2000. "Names, Memory, Public Identity in Late Medieval Florence". *Art, Memory, and Family in Renaissance Florence*. Edited by Giovanni Ciappelli and Patricia Lee Rubin, 237–252. Cambridge: Cambridge University Press.

Mousnier, Mireille. 2002. "Dépendents du Languedoc Occidental (milieu XIe–fin XIIIe siècle) : aspects anthroponymiques". [Dependents in Western Languedoc (Mid 11th–End 13th)]. *Genèse médiévale de l'anthroponymie moderne. Tome V-2 : Intégration et exclusion sociale, lectures anthroponymiques*. Edited by Monique Bourin and Pascal Chareille, 59–88. Tours: Presses universitaires François-Rabelais.

Najemy, John M. 2006. *A History of Florence 1200–1575*. Oxford: Blackwell Publishing.

Paoli, Cesare, ed. 1889. *Il Libro di Montaperti*. [The Book of Montaperti]. Firenze: G.P. Vieusseux.

Simon, Herbert A. 1955. "On a Class of Skew Distribution Functions". *Biometrika* 42, no. 3: 425–440.

Zanette, Damian H., and Susanna C. Manrubia. 2001. "Vertical Transmission of Culture and the Distribution of Family Names". *Physica A* 295: 1–8.

Zhao, Ziming, Xiaomeng Li, and Qinghua Chen. 2021. "The Use of a Matching Preference Index to Empirically Examine Distribution Imbalances in Beijing Citizens' Names". *Names* 69, no. 3: 41–56.

Zipf, George Kingsley. 1949. *Human Behavior and the Principle of Last Effort*. Boston: Addison-Wesley Press.

Notes on the Contributor

Stefano Ciliberti holds a PhD in Physics and an MA in History. His current research lies in quantitative history, with a focus on the application of statistical and mathematical models to historical data in order to investigate topics such as onomastics, economics, demographic change, and social mobility.

Correspondence to: stefano.ciliberti@gmail.com

DOI 10.5195/names.2026.2848 Vol. 74 No. 3, Summer 2026 ISSN: 0027-7738 (print) 1756-2279 (web)