UNLOCKING MIDDLE EASTERN NAMES

WILL HANLEY

Personal names are data repositories. The Islamic scholarly tradition tells us as much: the parsing of names and their placement in genealogical series and scholarly lineages is a durable feature of many genres over many centuries. Read in this way, names are signifiers for persons, and those persons are nodes in the networks that constitute Islamic society. But this prosopographic approach is not the only way to extract meaning from names. Names are also language artifacts that can teach us about language and culture independently of the person that they signify. This paper describes a path of analysis that seeks to understand something of the lives of ordinary people by means of the names they carried. Most names that appear in the historical archive can never be reconciled to any notable person. These names offer clues about the social context of their bearers, however, and these context clues can be read back over history's pool of names in such a way as to enrich our understanding of Middle Eastern pasts.

CURRENT APPROACHES

The *Onomasticon Arabicum* is the most sustained Western scholarly attempt to get a handle on Middle Eastern names. Now available in digital format, after almost a century of contributions, this work is essentially an index of notable individuals in the Islamic scholarly tradition.[1] The *Onomasticon* does not conceive of names as semantic resources. Instead, they function as personal identification, as keys that lead to historical individuals. The *Onomasticon* serves as a finding aid that can help scholars to identify positively individuals referred to in historical works. The destination of search, once the person in question has been identified, is an entry in a biographical dictionary.

Integral to the *Onomasticon* is a categorization schema for names that employs almost three dozen classes. These classes fall into several general groups: personal names (*ism*), genealogical markers, personal markers (*laqab*), and geographical and time markers. The digital version atomizes long names into their constituent units, facilitating search. Each person's name is rendered into a list of single words, and each word is assigned to a class in the schema. In its current iteration, the categorization is a manner of filtering advanced search, rather than a way to view an array of results. The function is to index.

In simplified form, this same schema lies at the heart of Annemarie Schimmel's very different study in her book *Islamic Names* (1989).[2] Schimmel celebrates the meanings of Islamic names in isolation from the persons to whom they are attached. Drawing in part on biographical dictionaries, but much more on the miscellanea of a lifetime of research in the Islamic world—including newspapers from Pakistan, telephone directories from Yemen, and the names of many dozens of people she encountered on her path—Schimmel maps out a range of meaningful naming practic-

es. The book is the product of bubbling curiosity and an alert ear. She suggests a potent and poetic range of meaning invoked by those who give Islamic names to their children and their friends. Ultimately, her study is a sort of etymological dictionary, its objective being understanding of the words themselves. *Islamic Names* exemplifies a qualitative approach; Schimmel understands names through their etymologies, and organizes her exposition of naming around origins. Names derived from God's names are different from those that honor the Prophet Muḥammad, from those that honor his various companions, and from those that honor saints. Children's names are derived from the time of their birth or its circumstances. The book is a catalogue of rules and exceptions. Schimmel is delighted by inconsistencies and seeks to root out their origin, typically by discovering a calque or an elision. The linguistic and geographic range of her research (Tunisia, Egypt, Turkey, Iran, South Asia, and further) offers her the means to produce this account of themes and variations.

This account answers many kinds of questions, but it does not address key issues of interest to social historians. First, Schimmel's account describes individuals, close families, and broad categories of membership by kin, language, and sect. It does not help to conceive of the work that names do to identify and differentiate between members of mixed communities, such as a neighborhood or a city. Second, Schimmel's definitions are meant to help us understand the rich cultural tradition of Islam. The individuals who bear these names are mere ciphers for that tradition. They help us to understand the unity and variegation of the Islamic world, but not the lives of the people who make it up. None of these are faults, but they describe a path for further research.

A third approach, which emphasizes neither individuals nor semantics, is that taken by Richard Bulliet in his landmark but isolated cliometric study *Conversion to Islam in the Medieval Period: An Essay in Quantitative History* (1979).[3] Bulliet set out to discover whether the aggregation of name evidence might reveal patterns of conversion over time. His study is based on the premise that conversion is an act that is marked by a name change. Sometimes this is straightforward, as in the adoption of ʿAbd Allāh as a given name (*ism*) by converts. Sometimes (and this is a special focus for Bulliet) conversion to Islam entails a change in the ethnic coding of the convert's name, especially from Persian to Arabic.

Bulliet worked with biographical dictionaries of the medieval period. Using the technology at his disposal in the 1970s, he coded names according to his variables of interest and produced two dozen charts showing patterns of change over time. The months of handwork behind each chart are impressive—and perhaps help to explain why his methodological example was not taken up by others in his wake. From our present-day perspective, it is noteworthy how contracted this cataloging labor appears. The work of analysis was directed toward a single purpose; it was not possible to generate the data in a way that could be reused. The inflexibility imposed by the technology of the period meant that the data collection led to an analytical culde-sac. While the findings there were of value, clearly Bulliet's method did not elicit imitation.

In recent years, however, digital scholarship has emerged that takes up Bulliet's method using new tools. Maxim Romanov is the foremost practitioner of this work.

Romanov discovered a new use for the massive new online repositories of texts from the Islamic tradition. He scraped content from sites such as Shamela.ws and began to explore it using digital tools. Like Bulliet, he was drawn to biographical dictionaries and the clues that their name data might offer about larger trends in the development of the Islamic world. Romanov draws on the regularity or formulaic nature of the dictionaries to discover patterns. His technique involves parsing the biographical dictionaries using a simple markup format that he developed by means of automatic (though hand-corrected) regular expression search. In this way, he is able to isolate the name that heads each entry and break it down to its component parts. Romanov concentrates especially on the meaning of *nisba*s. Using thirty thousand biographies from al-Dhahabī's fourteenth-century *Ta'rīkh al-islām*, he was able to derive a sense of the shifting centers of gravity in the geography of Islamic scholarship during the medieval period.[4]

Beyond these and other macrohistorical findings, what is notable about Romanov's scholarship is his commitment to open formats. Unlike Bulliet, he is working with tools that encourage and reward imitation and reproduction. He works carefully to document his methods and to encourage other scholars to apply them to other data. To this end, he is a leader of a large "*Open Islamicate Texts Initiative*," which seeks to establish and distribute standard corpuses of the scholarly tradition in plain text versions for computational analysis. This encouraging initiative will surely yield new insights derived from the automated analysis of large volumes of name data.

A fourth recent approach represents a different take on the cliographic tradition. Boğac Ergene has worked for a decade or more with the court records of seventeenth- and eighteenth-century Kastamonu. He decided to pursue quantitative analysis of these records in collaboration with an economist using methods developed in the context of American legal scholarship under the heading of "law and economics."[5] Ergene did careful hand-coding of the law court registers over a long period of time. He used names as a proxy for social position, identifying about a dozen variables of religion, occupation, integration into officialdom, and class standing. He used these identifiers to classify legal actors by social status, then to undertake quantitative analysis of the outcomes of their legal use of the courts.

One notable dimension of Ergene's work is his explicit endorsement of previous studies. Unlike previous quantitative analyses, his does not seek to replace methods such as close reading and microhistory. He wants rather to supplement them. His use of names in some ways resembles that of Bulliet and Romanov: they are proxies or signifiers of what he really wishes to understand, which is socioeconomic location. It remains to be seen whether Ergene's method is reproducible in other settings. Although he describes the classification process, he is not concerned to dwell on its elisions. But other scholars have shown that classification by names involves degrees of arbitrariness.

In sum, previous approaches to names offer a variety of lines of inquiry, each valid and valuable. Methodologically, however, it appears that scholars have had to commit to one or only a few of these lines. Moreover, their work has not been easy to share with scholars adopting different methods in any format but in monograph. At the level of evidence or data, each path exists in solitude.

RULES

A name is a sign. As we have seen, its significance can be read multiple ways. Following Schimmel, we can define the signified as the tradition from which the name was drawn, rather than the person bearing the name her- or himself. Following Bulliet, we can read the signified as socioreligious realignment, indexed by a set of personal markers.

In each of these cases, we make an abstraction of the individual in order to focus on the name, then read our conclusions back onto the individual. The individual bears a name that alludes to Shiʿi history; therefore the individual is (probably) Shiʿi. I wish to suggest that if we do more work on the isolated signifier itself, in the context of all other isolated signifiers, we can bring more context when we work to transfer its meaning back to the person bearing the name. To the extent that we can generate rules, and a system of rules that is open to ongoing refinement, this process can accumulate meaning. For example, Schimmel states that, in general, Islamic names consist of a sequence of elements:

1. *Kunya* Abū l Maḥāsin
2. *Laqab* Jamāl al-Dīn (optional)
3. *Ism* Yūsuf
4. *Nasab* ibn Abī Yūsuf Yaʿqūb
5. *Nisba* al-Makkī
5. *Nisba* al-Ḥanbalī
6. *Laqab* al-Zayyāt

This is not a hard rule, but a template to decode Islamic names. In certain imperial contexts, such as the Ottoman, we must add titles and honorifics to this formula.

The key to progress in this context involves sharing what (former IAS member) Amy Singer calls the "middle product" of research. The middle product is the layer of scholarly interpretation that lies between unprocessed archival manuscript primary sources and polished, published secondary studies. It is the organized data from primary reading. Behind Schimmel's and Bulliet's work lie more extensive datasets.

Something that scholars in the humanities do not do well—and something that diminishes the quality of shared data—is documenting how they create their middle product, and sharing it in a way that is open to those who do not share their interpretive premises. A notable recent example is the ten-volume series of court case summaries from seventeenth-century Istanbul edited by Timur Kuran.[6] Kuran, like Ergene, classifies litigants and cases for aggregate analysis of the workings of Ottoman-Islamic justice. He directed a team of researchers who prepared summaries of every court case they read, and these summaries form the core of this large publication. The result is certainly intriguing, but it is not clear how other researchers might repurpose this material for their own uses. The summaries are organized according to themes ascribed by the researchers (e.g., guilds, credit, interest). The language of the case record itself—notably, the names of the litigants—is not available. Also, the summaries are available only in print format, so automated reading is not possible. (It is important to note that Kuran plans to make the raw text available once his research team has published its results.)

Scholars are explicit about the ambiguity of naming rules. They do not pretend that this framework fits all cases. But we could do more to pursue the opportunity that ambiguity offers for discovery of new knowledge. The key step is to circulate the middle product of our research in as raw a form as it can be rendered. Often, this means exposing vulnerability, because the interpretation of ambiguity that is necessary in order to draw conclusions also risks criticism. Nevertheless, there is rather more to be gained than lost in showing our work.

It is not enough merely to share the raw data, however. It is also important to share the interpretive layer, which contains our own scholarly gloss on the raw data. So, in a list of names, it is important to make clear how each is being interpreted, what corrections have been inserted, and so on.

One of the weaknesses of the existing approaches surveyed above is that they are not designed for openness to reuse and reinterpretation. Schimmel, for instance, offers rules but no extended data—just an anecdote in support of each rule. Bulliet offers the abstraction of summary aggregate data, but no specifics—for instance, it is not possible to reproduce his work on a new dataset. If a scholar wishes to test a hypothesis about other terms that do not appear, she is out of luck. Their hypotheses about names are not falsifiable. Ergene's work is rather more specific about the meaning of the key terms that he uses for social classification, and another scholar could reproduce his analysis on a counterpart body of material. Scholars cannot ask new or different questions of the Kastamonu names, however, as they are not shared. In this sense, it is Romanov's approach that is most attractive. His data series is publicly available, his methodology is well documented, and he works to produce still more open texts for analysis. Most importantly, he sees the value in accumulating name data that have been lightly annotated (marked up) and can be analyzed in ever-larger pools. His annotation system is a kind of rule, but it is a translational language of interpretation open to reuse and adaptation in other contexts. Accumulating more and more data, from various historical contexts, offers a path towards unlocking the meaning of Middle Eastern names.

AN EXAMPLE

My own effort to unlock Middle Eastern names is located in the legal records of turn-of-the-twentieth-century Alexandria. These records are replete with terse references to individuals who can never be definitively identified: witnesses, bystanders, and others identified by a single name or by no name at all. These people represent the majority of the population of the city, and any social history must take seriously their claim to importance, but their names are recalcitrant. As a result, I have become interested in understanding how I might read them in context, and how I might generate context out of these names. I want to try to answer basic questions. For example, when I encounter the name Ibrāhīm, should I think that I'm dealing with a Jew or a Muslim? Both are possible, but which is more probable? I do not know the answer to this question, or even if it has one. I want to discover under what conditions it might be answered.

The most fruitful material I have found to approach this question are the lists of names maintained by various authorities in the city. I have lists of property own-

ers, lists of notable individuals from an *Indicateur Egyptien*, consular lists of subjects, and the like. In the paragraphs that follow, I will discuss my findings working with a set of lists of 303 Tunisian subjects in Alexandria in 1882. In that year, following the French invasion of Tunisia, the protection of these subjects in Egypt transferred from the agent of the bey of Tunis to the French consul. This transfer was managed by these two authorities with reference to the Egyptian interior ministry. I thus have three lists of these individuals, generated by three different authorities (the bey's agent, the French consulate, and the Egyptian interior ministry). Two lists are in Arabic, one in French. The Egyptian list is only names, but the other lists often contain other identifying material. Correlating the three impressions allows me to understand certain resonances of these names.

After digitizing these raw names, my first step was to break them down into their component parts in much the same manner as Maxim Romanov prescribes. This allowed me to isolate each element and consider it in the context of the elements that surround it. The first finding that strikes me, in considering the list that results, is that the classification framework on which the *Onomasticon* and others depend fits very loosely, if at all, on this set of evidence. Every individual has a given name (*ism*), but beyond this there is no standard rule. It is especially interesting that there would be such variation in a set of names developed only for identification purposes. Each of these names is meant to refer to a unique, known individual in an unambiguous manner, so that this individual's administrative status is clear. It would seem to make sense, in this context, to standardize the system of reference, but the evidence shows that this was not done.

The next finding concerns the designation of descent in names (*nasab*). Almost all of these individuals are identified with the names of their fathers. For 218 of the individuals, the particle *ibn/bin/bint* precedes the father's name; for 80 it does not. There is no clear pattern in this usage. Whereas in an Ottoman or Syrian context, for example, the particle *walad* is used for non-Muslims, this is not the case here.

Moving on to other indicators, I find in this set that titles are not used consistently. It is notable that they do not transpose predictably from one administrative system to another. While the Egyptians, the French, and the Tunisians all use "Hājj," they do not always use it for the same individuals. This suggests that using an approach such as Ergene's, which uses titles as a proxy for status, is more complex in such a multi-administrative setting. Not every clerk or administration read status in the same way; it was not an objective determination, and not consistently designated in naming practices.

A third indicator concerns the translation and transliteration of names. Like the converts whom Bulliet studied, these Tunisian French subjects engaged in various forms of passing, and this passing can be read in the orthography of their names. Fifteen of the sixteen *Yūsuf*s were Youssef, but one became Giuseppe; *Mūsā* became Moise and Moussa, *Dāwūd* became David and Daoud. The Jewish name Abraham was designated with the addition of an *alif* in Arabic—but not always. Here, too, there is no unified set of rules, just a range of practices. The correlation of the practices to other factors—class, profession, age, birthplace, and so on—is possible, but it demands a much longer series of data.

Another difficult-to-interpret finding is the spelling "errors" surfacing in the transition from one system to another. The Arabic records of these names quite consistently exhibit orthographic inconsistency or errors. It is not easy to explain this finding. Again, a larger data series might be necessary. It could be the result of a passing translation of some sort.

The findings that I have outlined here are preliminary and anecdotal. What sets them apart from previous research on names is the fate of the "middle product" on which these findings are based. I have shared this data, in a lightly formatted open format, via Github.[7] Included in this repository is documentation of the system that I used to generate and describe the data. This sharing, if repeated often and by many scholars, is the essential step towards the accumulation of a critical mass of data that will be able to yield broader insights into the meaning of Middle Eastern names.

What methods might we use to understand such a pool of information? Here data science offers intriguing tools in the form of ontology. Arp, Smith, and Spear define ontology as "a representational artifact, comprising a taxonomy as proper part, whose representations are intended to designate some combination of universals, defined classes, and certain relations among them."[8] In historians' terms, it is a clearly articulated hypothesis about structures of meaning that can be tested against a pool of evidence. Open-format, shared data yielded by historical research are amenable to such tests, at least in theory. But it will not be a straightforward process for historians to begin to incorporate such work into their research lives. First is the need to share "middle product" data, trusting that it is a contribution to knowledge even if it does not figure in short-term monographic analysis. The incentive structures of academic life in the humanities do not yet support this kind of sharing. Second is the need to establish habits of documentation. Scholars of the Middle East, accustomed as we are to idiosyncratic contexts and sources, do not easily describe our source use in transparent language, but this is a form of scholarly communication that is necessary if we are to generate common pools of evidence. Finally, scholars will benefit from using technology to pursue ontological thinking, which will allow us to test our hypotheses rigorously against actually existing data. In this way, we may manage to generalize about the meaning of Middle Eastern names in a manner that transcends anecdotal contexts.

---

[1] *Onomasticon Arabicum*, http://onomasticon.irht.cnrs.fr/.

[2] Annemarie Schimmel, *Islamic Names*, Edinburgh: Edinburgh University Press, 1989.

[3] Richard W. Bulliet, *Conversion to Islam in the Medieval Period: An Essay in Quantitative History*, Cambridge, MA: Harvard University Press, 1979.

[4] Maxim Romanov, *Al-Raqmiyyāt: Digital Islamic History*, https://maximromanov.github.io/.

[5] Boğaç A. Ergene and Metin Coşgel, *The Economics of Ottoman Justice: Settlement and Trial in the Sharia Courts*, Cambridge: Cambridge University Press, 2016.
[6] Timur Kuran (ed.), *Mahkeme Kayıtları Işığında 17. Yüzyıl İstanbul'unda Sosyo-Ekonomik Yaşam*, 10 vols., Istanbul: Türkiye İş Bankası, 2010.
[7] https://github.com/whanley/names-data/tree/master/tunisians-egypt-1881
[8] Robert Arp, Barry Smith, and Andrew D. Spear, *Building Ontologies with Basic Formal Ontology*, Boston: MIT Press, 2015.