Alía Warsco

Migrants, Archives & the State

December 13, 2024

<div align="center">Sketching Indigeneity in Mexican A-Files</div>

Introduction

It is worth stating to begin that identifying Indigeneity can be oxymoronic, since "Indigenous" is a label that people self-identify with based on their own connections to and relationship with Indigenous groups. Indigeneity is deeply rooted in lived experience, historical relationships, and community belonging. In this way, Indigeneity is not something that can be definitively measured or assigned based on data. This complexity is what leads me to propose "sketching" as a potential way to engage with Indigeneity in the A-files of Mexican immigrants. "Sketching" as a metaphor serves to convey the fluid, indirect, and nondefinitive role of limited, biased, and historical data in identifying Indigeneity. A sketch captures shapes, outlines, and traces — approximations, rather than declarations. My goal through sketching is not to label individuals in these A-files as Indigenous through my interpretations of their data, but rather to identify patterns, variables, or connections that may suggest personal ties to Indigenous identity or heritage for the purposes of understanding these people and their relationships with immigration better.

To make informed decisions during my analysis, I ground my work in an understanding of how Indigeneity was framed in post-revolutionary Mexico, which in some ways further complicates this project. The rise of *mestizaje* as a dominant narrative in Mexico served to both create historical ties to and erase explicit markers of Indigenous identity by assimilating them into a generalized national identity. I also draw on census data, geographic trends, and linguistic analysis to further uncover patterns that may suggest Indigenous connection in the data. These elements, when viewed together and through a historical lens, could offer valuable clues for sketching indigeneity even when the data does not explicitly name it.

This project aims to explore how indigeneity might surface in the margins of archival records, and to draw them out. It is not a project of definitive answers, but about uncovering traces, connections, and patterns that could create new opportunities for understanding these stories.

Mestizaje and *Indigenismo* in Post-Revolutionary Mexico

Following the Mexican Revolution (1910-1920), the concept of *mestizaje* begins to play a crucial role in shaping the national identity of Mexican people, which informs the basis of this project. While the revolution brought about significant changes in Mexican government and social spaces, it also became a nation-building project through an imagined historical connection that all Mexicans shared to a broad, precolonial Indigenous past. This formed in response to the history of Spanish colonialism in the country, and a rejection of Eurocentric history and racial hierarchies that previously dominated thought on race and identity (Zhang 2024, 7). By identifying everyone as *mestizo,* or mixed, the *Partido Revolucionario Institucional* (Institutional Revolutionary Party, PRI) government sought to both connect to a vague national indigeneity, and incorporate and integrate Indigenous peoples into this national framework (Zhang 2024; Lewis 2006). Within this national identity, however, the line of who becomes "truly Indigenous," or perhaps "too Indigenous," in the context of a *mestizo* Mexico becomes blurred.

After widespread and successful attempts to encourage *mestizo* identity throughout Mexico, the question emerged of how to incorporate and assimilate Indigenous populations (mostly rural, Indigenous language-speaking, and economically poor) into this national identity. *Indigenismo* describes the ideologies, policies, and practices aimed at addressing this "Indian Problem" in Mexico (Lewis 2006). *Indigenismo* is often characterized as a contradictory approach to Indigenous people — while celebrating art, music and traditions as part of national heritage, *Indigenismo* also implemented policies of assimilation that erode indigenous language and unique practice and ways of relating to the state (Velázquez and Vaughan 2006). The idea of celebrating *Indigenismo*/Indigeneity through art, music, and traditions, while ignoring the material conditions of Indigenous people, were the foundations of mestizaje.

The early decades of the post-revolutionary period were marked by a strong emphasis on the assimilation of Indigenous people into this *mestizo* imaginary. Policies reflecting "spiritual eugenics," which aimed to incorporate Indigenous peoples by encouraging them to abandon their language and cultures are also reflected in later focuses on Spanish-only education and the SEP's efforts to "transform" Indigenous students during the 1930's (Lewis 2006, 179). Facing this discrimination, people were forced to either assimilate into the *mestizo* majority, or reject these definitions in an attempt to stay connected to their specific Indigenous heritage.

This context reveals the difficulties in identifying Indigenous people at all in post-revolutionary Mexico. If the nation's identity relied on a European/Indigenous mixture, the concept of Indigeneity thus became fluid and contested. Some definitions of "Indigenous" were broad, encompassing bilingual (Spanish and Indigenous Language) rural Mexicans, or even all impoverished *campesinos*, regardless of language or cultural ties to Indigenous groups (Lewis 2006, 178). Others take into account a combination of factors like Indigenous languages, dress, religion, social organization, consciousness, and self-definition (Lewis 2006, 178). But in the view of the state in its assimilative policy building, Indigenous, *indígena,* and *indio* became labels that reflected more material realities than self-identification. On the other hand, identification with them *mestizo* national identity came with benefits. The PRI's land distribution system, while attempting to address historical injustices, favored *mestizo* farmers. Access to education, employment opportunities, and social mobility were also often tied to an assimilation in the *mestizo* mainstream (Zhang 2024, 16; Mackinlay 1996).

Despite official celebrations of Indigenous pasts through *mestizaje,* indigenous people continued to face discrimination and prejudice in many areas of Mexican society. The association of "Indianness" with poverty, backwardness, and lack of education created a social stigma that some wanted to distance themselves from or reject. Alfonso Caso offered a definition that was ultimately adopted by the National Indigenous Institute (INI): "An *indio* is someone who feels he belongs to an indigenous community." (López Caballero 2018, 215). In my attempts to identify indigeneity, I hold this definition as central to my analysis while holding the context of how Indigenous people were identified in documentation. During

this time, *Indigena* and *Indio* start to take a socioeconomic and linguistic meaning rather than a racial one. This directly informs my methodologies in searching for Indigenous language features, characterizations of race (through both specific race labels and skin tone classifications), as well as regional distinctions, literacy, and occupations. I hope this does not reflect my argument as supporting these factors as Indigeneity determiners. They are, however, variables that could be classified consistently in official documentation within A-files and are historically associated with Indigenous peoples under this context.

Methodology

Along with the historical context of *mestizaje* and *indigenismo* in revolutionary and post-revolutionary Mexico, census data largely informs many of the initial assumptions I made in gathering and analyzing data. The 1921 Census gives a few insights into Indigenous populations in Mexico at the time. In 1921, Indigenous people represented 29% of the population (4,179,449 total persons), and *raza mezclada (*mixed race; worth noting, not *mestizo*) represented 59% of the population, while white and other unlisted races made up the little rest (9.8% and 1%, respectively). Additionally, about 14% of the total population listed an Indigenous language as their native language. Of these, Nahuatl (472,690), Yucatec Maya (234,675), and Zapotec (214,586) were the most commonly spoken languages (INEGI 1921, "Razas"). This contextualizes the position of Indigenous people around the time most applicants in this dataset were born or were children during, since the files available are only ones who list a birth date of over 100 years ago.

By comparison, the 1970 census reveals the changing landscape of Indigeneity after the rise of *mestizaje* and *Indigenismo*. Specifically, the Mexican government stopped recording racial classifications after 1921, declaring instead that all Mexicans were now *mestizos.* This pushed the analysis of race, specifically Indigeneity (or Blackness, whiteness, or any other race) to be inferred from language-based, regional, or socioeconomic census data. Most of the immigrants in the A-files I analyzed migrated during the 1960s and 1970s, when the ideologies of *Indigenismo* and *mestizaje* were clearly fully incorporated

into the national Mexican narrative and imaginary. However, there is still linguistic data for languages

other than Spanish, and detailed regional and socioeconomic data that may be helpful. The 1970 census

reveals Nahuatl (799,394 speakers), Maya (454,675), and Zapotec (283,345) still represented the 3 most

commonly spoken languages, but at lower overall rates (6% of the total population spoke an Indigenous

language, or about 3.1 million people).[1] States with highest populations of Indigenous language speakers

were Oaxaca, Veracruz, Yucatan, Puebla, and Chiapas (INEGI 1970, "Lengua Indígena"). Using this

data, I decided to rely mainly on linguistic features to try to establish a connection to Indigeneity, but to

keep factors such as region and socioeconomic factors in mind while exploring the data.

Phase 1: Data analysis

The three Indigenous languages identified as most commonly spoken in both sets of census data

(Nahuatl, Yucatec Maya, and Zapotec) formed the basis for searching the total A-file dataset (Mexico

ALL OR YYYY.csv). These languages offer unique linguistic features that can help to quickly

distinguish them from other languages, particularly European languages. This analysis focuses on

identifying these linguistic features within the applicant's first and last name or their parents' names.

These linguistic features include specific consonant clusters and other morphemes found in the three

identified languages.

These linguistic features include:

- Nahuatl: *tl, tz, ch, xo*, *qui*[2]
- Maya: *bak, cab, cal, dz'*[3]*, ch, k', b', ii, aa*
- Zapotec: *dz, xh, ty, zua*[4]

---

[1] The higher speaker totals may reflect better campaigning efforts and accessibility to rural Indigenous populations. These totals include both people who also speak Spanish or are monolingual Indigenous language speakers, which was not recorded in 1921. However, the increase in speakers may also reflect the isolation of certain Indigenous populations, and the ability to continue speaking and teaching their language.

[2] Some clusters, like *qui*, are common in other languages as well (Quintanilla is of Spanish origin, so it is not counted). For these cases, where the origin is not obvious, I tended to conduct a simple web search to determine if the name is of European language or potentially Indigenous language origin

[3] The apostrophes denoting glottal stops or plosives in Maya/Zapotec are not recorded in the records, potentially to not interrupt the encoding schema of the CSV file.

[4] Identified consonant clusters were collected & adapted from Suárez 1983.

Ancestry websites were also helpful in identifying specific last names associated with these languages, such as *Oaxaca, Maal, Poot, Tun, Moroyoqui*, and *Ixtla*. These were also recorded and considered in this analysis. With more time, this search could be expanded to include more of the most spoken Indigenous languages and unique features or last names associated with them. There is also a possibility of misspellings within the files that cannot be completely accounted for with these methods.

This search of the data identified 119 records that include these linguistic markers and last names. With these A-files separated, I tried to find any of them in the full A-file OCR texts we have available. I was unable to find any, which necessitated a shift in approach when working through the A-files' full OCR. Further analysis of these 119 A-files is included in the "Key Findings" section.

Phase 2: Text analysis

Since none of the 119 identified records were available with in the collection of 58 full-text, OCR-processed A-files, I varied my approach to this phase of analysis. I found that I was also unable to use the same search strategy for specific linguistic features as they yielded fewer relevant, if any, results. Given these challenges, I shifted my rationale to searching for potentially related and more indirect factors that were consistently documented in the A-files, like racial classifications and markers, regional or geopolitical space, and socioeconomic factors.

This analysis was conducted using the program Voyant Tools, an open-source text analysis program that allows for easier searching with additional context and relationships within large corpuses of text. My search terms initially consisted of words like race (*raza),* Indigenous *(Indígena),* mixed (*mestizo/a, mezclada),* and white (*blanco/a).* While mestiz* appeared in 10 A-file documents, *Indígena* appeared only once. What I quickly realized as a problem within the data was that race was inconsistently documented and recorded in the files. The same applicant could be classified as white on some documents, "Mexican" on others, and *mestizo* on even more. Primarily, I found that most applicants were

usually classified as white on official documentation in the US immigration process, even if they were

classified otherwise elsewhere.

Fig 1: Trinidad Gomez (A34261850), listed as "MEX" race.

Fig 2: Rafael Sanchez Vega (A13073806), listed as "W" race

Fig 3: Optional Form FS-510 (*Solicitud de visa de inmigrante y registro de extranjero)* from 1969 for Edwina Arias

Gonzalez (A18961437), nothing her *tez* (complexion) as both "*moreno"* (brown/dark-skinned) and "fair"

In conducting these searches and doing close readings of the context surrounding certain words, I

was able to identify related markers that were consistently documented within A-files, including skin

complexion descriptions, states of origin, literacy rates, and occupations. These struck me as potentially

important indirect factors in identifying Indigeneity within these A-files. To explore these factors, I

identified and recorded aspects of each, such as:

- States of origin: Applicants from Mexican states with historically high Indigenous populations,
  such as Oaxaca, Veracruz, Yucatán, Puebla, and Chiapas. Though these states were not highly
  represented in these files, they may be a contextual link to Indigeneity based on the previously
  identified census data.

- Skin complexion descriptions: Skin complexion (tez) was often recorded in both US and Mexican documents. I imagine these were subjective calls made by immigration agents. Descriptors like fair/white (*blanco*), brown/dark-skinned (*moreno*), and olive might reflect racial perceptions tied to physical appearance, and may provide clues on Indigenous identity.

- Literacy rates: Visa application forms commonly record languages that are spoken, read, and written by applicants. Notably, no Indigenous languages appear on these documents.[5]

- Occupations: Laborers and farmers were the most commonly recorded occupations of male applicants, while most women were housewives. These occupations, which are commonly associated with lower socioeconomic status, may be a contextual clue based on the associations of Indigeneity with rural, *campesino* classes.
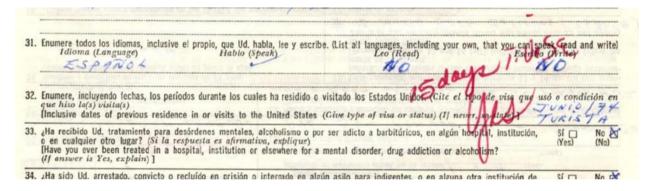

Fig 4: Optional Form 230-B (previously FS-510[6]) from 1975 for Abundio Cornejo Cuevas (A35938326), noting his illiteracy.

Throughout the files, reference to specific Indigenous languages was entirely absent. However, I believe this review may reveal descriptors and factors that have been previously identified as tied to Indigenous identity. Race is documented in unclear and inconsistent ways throughout the documents, which while frustrating, may reveal patterns in the way Indigeneity was encoded through a combination of these more indirect factors and contextual evidence. These alternatives may help researchers further

---

[5] Given the specific ways Indigeneity was tied to low literacy rates, which in part inspired the *Indigenismo* educational policies of Spanish-only education, I thought this was an important variable to include.
[6] Forms 230-B and its predecessor FS-510 were consistent and helpful forms frequently referenced for the collection of literacy, skin complexion, and occupation data. This form may be useful for further analysis.

insights into Indigenous immigration in a *mestizo*-standard Mexican context. Further findings from the

OCR text analysis are also in the next section.

Summary of findings

Below is a summary of the findings I made during my research process. In no way are these

findings meant to be conclusive for identifying Indigenous representations within A-files. Instead, they

represent patterns I identified in the limited data and potential opportunities for further exploration and

questioning.

Of the identified records from Phase 1 (n=119):

- San Ysidro (SYS) and El Paso (ELP) represent about 65% of the POEs for records who
  have that data.
- There seems to be a trend that many of the immigrants identified entered during the
  1960s, which may point to a relationship with labor programs like the Bracero program.[7]
- Almost all the records identified did not specify sex.

Of the records analyzed during Phase 2 (n=58):

- "Mestiz*" appears in 9 files (15.5%). "Indigena" appears only once. Race classifications
  on English documents tend to be "White" or "Mexican," regardless of other racial
  identifiers within the application.
- Indigenous languages very rarely (if ever) enter official documentation.
- Skin complexion (tez) is often reported by agents. Of the 58 files, the breakdown of skin
  complexion is as follows:
    - Fair/Blanca: 28.8%
    - Medium/Moreno[8]: 48.8%
    - Olive: 2.2%
    - Dark: 15.5%

---

[7] Labor programs, like the Bracero program, directly impacted Indigenous people. In fact, local governments often deliberately allocated bracero contracts to municipalities with large Indigenous populations for the sake of "modernization" (García 2021).

[8] As seems to be the trend with this project, *moreno* specifically can often mean either brown or darker-skinned in this context.

- Thirty (51.7%) files had literacy information via the 230-B or FS-510 form. Of those, 80% reported literacy (reading + writing) in Spanish. This is consistent with, if not higher than literacy rates in Mexico broadly in the 60s-70s, when most immigrated. 75% (n=3) of those who were not literate were housewives, and 2 of those housewives were specifically classified as mestiza within documentation.
- Most women (81%) were homemakers (designated as "housewife" or "hogar" in paperwork). Those who were not were laborers, teachers, and housekeepers.
- Most men worked as laborers (33%) or farmers (18.52%). The forms did not specify the type of labor. Other occupations included mechanics (7.4%), merchants (7.4%), service jobs (7.4%), "professional" (3.7%, or 1 applicant) or unemployed (3.7%, or 1 applicant). Men classified as laborers or farmers were more likely to be classified as Moreno (50%) or Dark (25%).
- There may be further correlations between literacy, skin complexion, occupation, and language.

The full data collected is attached in the Appendix as a Google Sheet.

Conclusion

My project set out to sketch Indigeneity within the A-files of Mexican immigrants. I knew going into this project that it had inherent challenges: identifying indigeneity is not a simple checkbox, the context of post-revolution Mexico complicates Indigeneity in the entire country, and Indigenous people themselves may want to distance from Indigenous identity to avoid discrimination. In conducting my research, I found several other challenges — a lack of evidence of Indigenous languages within the data, limited findings on specific states or regions, and an inconsistent way of classifying race throughout documents even for single applicants. This, combined with the inherently limited and biased nature of historical records, amounted in a challenge that I was not confident I could reckon with.

I did not discover a way to definitively identify Indigeneity within A-files. However, I did identify specific and consistent variables that could be used to sketch an Indigenous identity with the proper context. These variables include Indigenous language or linguistic features, race identifiers, states with historically high Indigenous populations, skin complexion, literacy, and occupation. I don't believe

my difficulties in sketching Indigeneity in these files suggest that fewer Indigenous people were migrating. Rather, they may point to the systemic erasure or obfuscation of Indigeneity in official documentation.

I still believe there is merit in certain methodologies I used during this project. Some direct variables, such as linguistic features found in Indigenous languages, specific race identifiers (like mestiz* or indígena), and states with high indigenous population strike me as potentially very telling and useful markers for identifying Indigenous connections. Other factors, like skin complexion, literacy rates, and occupation, are varied in their consistency but could provide further clues in how Indigenous people were encoded and affected by immigration records. With access to the full A-files of the 119 records identified in Phase 1, deeper analysis could yield more insights into the intersections of race, language, and identity; and could reveal more factors in identifying these relationships, trends, and classifications more concretely. Additionally, more research could be conducted with the data points I've identified through this project. Cross-referencing and comparing variables such as race or skin tone with different occupations and literacy rates might reveal more patterns in how Indigeneity was lived, and represented or erased in these documents. Close readings of individual A-files could provide valuable insights into the lives, potential interactions, and relationships to Indigeneity that may not surface in the higher-level analysis that I conducted.

Ultimately, I think the project was successful in underscoring the nuance and complexities of identifying Indigeneity within data. Of course, data does not define Indigeneity — people do. But by considering linguistic, geographic, and sociohistorical factors together, we may be able to suggest, or sketch, connections to Indigenous groups or language that reveal deeper insights into this critical group. The factors I've identified may collectively begin to trace Indigenous narratives that have been hidden or obscured in official record. By applying these digital humanities techniques to the two datasets, the project hopes to support further research into these intersections of race, power, identity, and representation.

Appendix

- [Corpus (Voyant Tools)](#)
- Migrants and the State Github:
    - [A-file Aggregator](#)
    - [Full OCR text (all A-files)](#)
- [Google Sheet with all data from text analysis](#)

References

García, Alberto. 2021. "Regulating Bracero Migration: How National, Regional, and Local Political

    Considerations Shaped the Bracero Program." *Hispanic American Historical Review* 101 (3):

    433–60. https://doi.org/10.1215/00182168-9051820.

INEGI (National Institute of Statistics and Geography). 1921. "General Census of Inhabitants, 1921."

———. 1970. "IX General Census of Population 1970."

Lewis, Stephen E. 2006. "The Nation, Education, and the 'Indian Problem' in Mexico, 1920-1940." In

    Eagle & the Virgin, 176-95. Duke University Press.

López Caballero, Paula. 2018. "Anthropological Debates around the Indigenous Subject and Alterity,

    1940–1948." In *Beyond Alterity: Destabilizing the Indigenous Other in Mexico*, edited by

    Ariadna Acevedo-Rodrigo and Paula López Caballero, 199–221.

Mackinlay, Horacio. 1996. "La CNC y El" Nuevo Movimiento Campesino"(1989-1994)." *Neoliberalismo

    y Organización Social En El Campo Mexicano*, 165.

Suárez, Jorge A. 1983. *The Mesoamerican Indian Languages*. Cambridge [Cambridgeshire] ; New York :

    Cambridge University Press. http://archive.org/details/mesoamericanindi0009suar.

Velázquez, Marco, and Mary Kay Vaughan. 2006. "Mestizaje and Musical Nationalism in Mexico." *The

    Eagle and the Virgin: Nation and Cultural Revolution in Mexico, 1920-1940*, 95–118.

Zhang, Qingren. 2024. "The Integrated Mexican Nation-State Building in the 20th Century."

    International Journal of Anthropology & Ethnology 8 (1): 1–20. https://doi.org/10.1186/s41257-

    024-00119-1.