

From Strange to Normal: Computational Approaches to Examining Immigrant Incorporation Through Shifts in the Mainstream

Sociological Methods & Research

1–40

© The Author(s) 2022





Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/00491241221122596

journals.sagepub.com/home/smr

Andrea Voyer ¹,
Zachary D. Kline ²,
Madison Danton²,
and Tatiana Volkova³

Abstract

This article presents a computational approach to examining immigrant incorporation through shifts in the social “mainstream.” Analyzing a historical corpus of American etiquette books, texts from 1922–2017 describing social norms, we identify mainstream shifts related to long-standing groups which once were and may currently still be seen as immigrant outsiders in the United States: Catholic, Chinese, Irish, Italian, Jewish, Mexican, and Muslim groups. The analysis takes a computational grounded theory approach, combining qualitative readings and computational text analyses. Using word embeddings, we operationalize the chosen groups as focal group concepts. We extract sections of text that are salient to the focal

¹ Department of Sociology, Stockholm University, Stockholm, Sweden

² Department of Sociology, University of Connecticut, Storrs, Connecticut, USA

³ Department of Computer and Systems Sciences, Stockholm University, Stockholm, Sweden

Corresponding Author:

Andrea Voyer, Department of Sociology, Stockholm University, SE-106 91 Stockholm, Sweden.

Email: andrea.voyer@sociology.su.se

group concepts to create group-specific text corpora. Two computational approaches make it possible to examine mainstream shifts in these corpora. First, we use sentiment analysis to observe the positive sentiment in each corpus and its change over time. Second, we observe changes in each corpus's position on a semantic dimension represented by the poles of "strange" and "normal." The results indicate mainstream shifts through increases in positive sentiment and movement from strange to normal over time for most of the group-specific corpora. These research techniques can be adapted to other studies of social sentiment and symbolic inclusion.

Keywords

word embeddings, sentiment analysis, immigrant incorporation, manners, etiquette

Introduction

Developments in computational social science provide new angles for studying old questions. The study of immigrant incorporation has a long history in the social sciences. However, research has generally focused on the adaptation of immigrant minorities to the societies where they settle. Even though many influential theories of immigrant incorporation posit the importance of the non-immigrant "mainstream" to immigration (see Alba 2005; Alba and Nee 2003; Massey and Sánchez 2010; Portes and Rumbaut 2001; Portes and Zhou 1993), "established" (Jiménez 2017) individuals rarely take center stage in studies of immigrant incorporation. When they do, they tend to be seen as part of the context of immigrant reception or as gatekeepers in the incorporation process.

There is research showing how established individuals change as a result of immigration – including developing comfort in multicultural spaces, rethinking aspects of how they live their everyday lives and conduct social relationships, relativizing their own social and cultural norms, and adopting a more expansive sense of what it means to be a neighbor and a citizen (Jiménez 2017; Voyer 2013b). These social impacts of immigration reshape group boundaries and collective practices, as well as shared meanings and sentiments associated with the "mainstream" (Voyer 2013a). Methodological innovations from computational social science provide new ways to observe immigrant incorporation through such shifts.

This article presents a computational sociological approach to examining immigrant incorporation through mainstream adaptations and adjustments. We draw on different research techniques in a computational grounded theory approach, leveraging qualitative insights and quantitative power (Nelson 2020). The analytical work begins by selecting a theoretically-indicated corpus consisting of 20 editions of Emily Post's *Etiquette*, which we take as a text-based representation of the American mainstream. We use human reading and machine techniques such as topic modeling and keyword-based approaches in exploratory analyses aimed at determining how immigrants and immigrant incorporation are represented in the corpus.

Based on the exploratory work, we conduct two analyses of text changes occurring in light of immigration and in relationship to long-standing groups who once were and may currently still be seen as immigrants and outsiders in the United States: Catholic, Chinese, Irish, Italian, Jewish, Mexican, and Muslim (Dumenil 1991; Portes and Bach 1985; Takaki 1993). We use Concept Mover's Distance, an approach rooted in word embeddings, to operationalize the selected groups as focal group concepts and to identify paragraph-length chunks of text that are salient to each concept within each edition (Stoltz and Taylor 2019). We then aggregate these chunks to create "pseudoeditions." We conduct sentiment analyses, examining changes in positive sentiment across editions. We then create a semantic dimension anchored by the poles of "strange" and "normal" and establish the position of each pseudoedition on that dimension. The computational results are interpreted through triangulation with qualitative analysis of the texts.

The results provide evidence of the incorporation of immigrant groups through mainstream shifts. We find a trend toward more positive sentiment for most focal group concepts. We also find that the texts linked to most focal group concepts move from an association with "strange" toward associations with "normal" over time. However, there is group-by-group variation in the timing and trajectory of these shifts. Shared trends toward both normal-ness and positive sentiment across these different focal group concepts suggest broader social shifts in mainstream diversity and inclusion, and group-by-group differences may reflect particular social and historical circumstances. These findings also demonstrate that positive sentiment and "normal"-ness are independent aspects of incorporation into the mainstream.

Literature and Theory: Mainstream Shifts

Many influential theories of incorporation posit the importance of the mainstream to immigration (c.f. Alba 2005; Alba and Nee 2003; Massey and Sánchez 2010). For example, in segmented assimilation theory (Portes and Rumbaut 2001; Portes and Zhou 1993), immigrant incorporation hinges upon different aspects of the migration experience, including the context of immigrant reception, which consists of attitudes towards immigrants and public practices and policies shaping immigrants' opportunities in the new country. Past work has also identified the symbolic processes underlying immigrant incorporation. Alba's (2005) bright and blurred boundaries suggest that, in some contexts, the boundary between ethnic minorities and the mainstream majority is blurred, allowing some immigrant groups to enter the social mainstream, and, in other contexts, the boundary is bright, maintaining the exclusion of other immigrant groups. Wimmer extended this insight and developed a comprehensive typology of the boundaries between ethnic groups, including various approaches to modifying ethnic group boundaries such as the development of overarching global, civil, and local identities (Kroneberg and Wimmer 2012; Wimmer 2008; Wimmer 2009).

The literature tends to be optimistic regarding the possibilities for immigrant incorporation, which is typically seen as a multigenerational process (Drouhot and Nee 2019). We can point to immigrant "success" stories, such as how Irish, Italian, and Jewish immigrants, who all initially faced hostility and racialized exclusion, over generations came to be seen as white American ethnic groups (Brodkin 1998; Ignatiev 2009; Roediger 2005). Of course, other immigrant groups did not "become white" – including Chinese, Japanese, and Mexican immigrants (Lee 2003; Massey, Durand, and Malone 2002; Takaki 1993). As Roediger (2005) notes, these non-white immigrants and their descendants, even in cases where they achieved economic and educational success, are not always symbolically included in the American mainstream. Instead, their inclusion shifts with the ebb and flow of political and economic tides – from inclusion as racialized Americans to exclusion as potential outsiders who could be exhorted to "go back to their own country." This in-betweenness is reflected in and reinforced by systematic racism, for example, in federal red-lining policies and the development of post-war housing that moved Irish, Jewish, and Italian immigrants into the white suburbs, while leaving Mexican and Asian immigrants in the cities (Roediger 2005). Recent research shows the continuing importance of race to immigration. Contemporary white Americans do not mind having

immigrant neighbors, especially if they are white, speak English fluently, and volunteer in the community. However, they see non-white immigrants as less preferable neighbors and less similar to themselves (Schachter 2016).

We recognize the concept of the American “mainstream” is closely related to the white, middle-class Anglo-Saxon Protestant majority established through colonization, conquest, and the subjugation of native and enslaved peoples (Brown 1970; Takaki 1993). Symbolic boundaries, including racial boundaries, of the mainstream are crucial to the trajectory of immigrant incorporation. There is growing literature extending our knowledge of mainstream shifts in light of immigration. The fundamental insight of such work is that immigrant incorporation is a relational process involving the adjustments and adaptations of immigrants and the established people and practices in the societies where they settle. For example, Voyer (2013b) finds that rapid and substantial Somali immigration into a historical white town prompts a redefinition of community membership that could include black, Muslim immigrant newcomers, a revision of belonging that blurs racial boundaries. Meanwhile, in research among established individuals in Silicon Valley, Jiménez and Horowitz (2013) find that mainstream conceptions of success devalue whiteness compared to Asian-ness. Jiménez (2017) observes that established individuals in at least the third generation in the United States revise their notions of what it means to be an American as they become comfortable and familiar with the languages, foods, and cultural and religious practices of immigrant newcomers. The existing work on mainstream shifts consists mainly of ethnographies and interview studies. Computational methods for text analysis provide an opportunity to extend this research by focusing on the construction of the American mainstream over time and in relation to immigrant groups.

Data

To uncover mainstream shifts, we analyze a corpus of etiquette books. The purpose of etiquette writing is to bring mundane practices and meanings from the cultural background to the foreground, and we use this practical and symbolic background to represent the social mainstream. Sociologists have long recognized etiquette as a valuable source of data on the social world (Arditi 1999; Bourdieu 1984; Elias 2000; Goffman 1963; Mennell 2007; Wouters 2007). Recent research has used computational analysis of etiquette books to study general social changes (Abrutyn and Carter 2015). To our knowledge, this is the first research to use etiquette literature to study immigrant incorporation.

We constructed an etiquette corpus from 20 editions of the best-selling American etiquette manual, Emily Post's *Etiquette*, which was first published in 1922 and most recently published in 2017.¹ *Etiquette* is considered a "pre-eminent example of American advice literature" (Lees-Maffei 2012: 217). *Etiquette*'s popularity is enduring, with sales at least in the 10s of millions (McHugh 2021:13). Only the bible has been purchased more by schools and libraries (Claridge 2009; Jacobs 2001). In 1995, *Etiquette* was featured in the New York Public Library's "Books of the Century" exhibit, and it was commemorated with a postage stamp in 1998 (Lees-Maffei 2012).

Table 1 displays the *Etiquette* editions and authors. Emily Post wrote *Etiquette* until her death in 1960. Her granddaughter-in-law, Elizabeth Post, then took over authorship. In 1997, Elizabeth was succeeded by her daughter-in-law, Peggy Post. Beginning 2011, three Post-cousins, the great-great grandchildren of Emily Post, joined Peggy in authorship. Two of those cousins, Peter Post Senning and Lizzie Post are currently the main authors of *Etiquette*.

We attach no more social influence or causal power to *Etiquette* and its authors than we would to any other successful authors or widely read books. *Etiquette* provides insight into the American mainstream because presenting the mainstream is the purpose of such texts. Etiquette books are tasked with representing social organization at the time of their publication by keeping pace with the changing social norms of their intended audience. Just as Emily Post revised *Etiquette* in light of social changes large (e.g., the Prohibition and World War II) and mundane (e.g., the rise of the telephone and traveling by plane), so have subsequent authors adjusted to the frontiers of social life in terms of changing gender roles, family structures, and waves of rising racial and ethnic inclusivity and exclusivity, etc. But there is also consistency in form and substance across editions (see discussion of text changes in *Exploratory Analyses*). Research shows that *Etiquette* remains thematically and structurally consistent with its contemporary etiquette manuals (Arditi 1999; Wouters 2007). Our analyses of *Etiquette* suggest that the text has historically been written in reference to the white, middle-class, Protestant-American mainstream (Voyer, Kline, and Danton 2022).

Each edition is 500–1100 pages long and includes photographs, figures, and charts. The first edition is in the public domain. All other editions remain under copyright and are used under fair-use doctrine. A physical copy of each edition was scanned into a text document using optical character recognition (OCR) software. Errors arising

Table 1. Etiquette Editions Included in Corpus.

Edition	Year	Title	Author
1	1922	Etiquette: In Society, in Business, in Politics and at Home	Emily Post
2	1927	Etiquette: The Blue Book of Social Usage	Emily Post
3	1931	Etiquette: The Blue Book of Social Usage	Emily Post
4	1934	Etiquette: The Blue Book of Social Usage	Emily Post
5	1937	Etiquette: The Blue Book of Social Usage	Emily Post
6	1940	Etiquette: The Blue Book of Social Usage	Emily Post
6*	1942	*Wartime supplement	Emily Post
7	1945	Etiquette: The Blue Book of Social Usage	Emily Post
8	1950	Etiquette: The Blue Book of Social Usage	Emily Post
9	1955	Etiquette: The Blue Book of Social Usage	Emily Post
10	1960	Etiquette: The Blue Book of Social Usage	Emily Post
11	1965	Emily Post's Etiquette	Elizabeth L. Post
12	1968	Emily Post's Etiquette	Elizabeth L. Post
13	1975	The New Emily Post's Etiquette	Elizabeth L. Post
14	1984	Emily Post's Etiquette	Elizabeth L. Post
15	1992	Emily Post's Etiquette	Elizabeth L. Post
16	1997	Emily Post's Etiquette	Peggy Post
17	2004	Emily Post's Etiquette	Peggy Post
18	2011	Emily Post's Etiquette: Manners for a New World	Peggy Post, Anna Post, Lizzie Post, Daniel Post Senning
19	2017	Emily Post's Etiquette: Manners for Today	Lizzie Post, Daniel Post Senning

during the OCR process are managed and corrected within the version control manager GitHub, and HTML mark-up assists with the retention of text features like images, tables, and sidebars. Prior to analysis, the text is tokenized and stop words are removed. Analyses are conducted using text analysis tools based in R and Python programming languages.

Method

This research followed a computational grounded theory approach. Computational grounded theory, which allows for back-and-forth movement between qualitative readings and computational analyses of text, makes it possible to answer theoretical and qualitative sociological questions by developing large-scale and reproducible analyses of the social meanings present in texts (Nelson 2020; Nelson 2021). Ongoing verification and confirmation are central to computational text analysis (Nelson 2020). We employ many different approaches to confirmation and verification. Description of confirmatory analyses is incorporated into the discussion of each aspect of the research and further described in the online appendices. An overview of the research process is depicted in Figure 1.

Exploratory Analysis

Exploratory analysis uncovers patterns that may be investigated through subsequent analysis while also providing a general sense of the corpus. By combining human reading and computational analyses, we observe that mainstream shifts associated with immigrant incorporation are often implicit in the texts. This exploratory work was crucial in determining the direction of the main analysis.

During exploration, we examined the nature of changes in *Etiquette* over time. Many factors contribute to edition-by-edition change in texts (Giuliani et al. 2006). Because each edition of *Etiquette* is written in explicit reference to the previous edition(s), the decision to add or delete topics might reflect social change. At the same time, simple shifts in vocabulary, editorial practices, developments in publishing and printing, and authors' stylistic quirks also shape *Etiquette*'s development. Authorship changes provide a compelling alternative explanation, as opposed to mainstream shifts associated with immigrant incorporation, that could be driving relevant changes. For this reason, we assessed the impact of author changes on text changes. Table 2 shows changes in the corpus over time by plotting the cosine similarity between all editions. Editions generally display

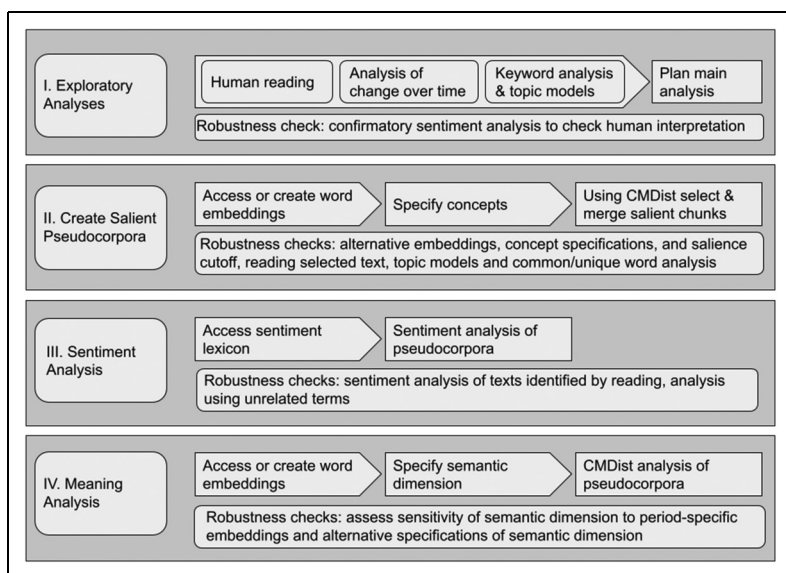


Figure 1. Process overview. Elements of research organized by phase of the analysis (I–IV). Arrow-shaped boxes indicate elements that occur sequentially, leading to outcome of research phase, represented by boxes with sharp corners. Ongoing robustness checks listed in rounded boxes below. Note: elements of exploratory analysis can occur simultaneously.

a high degree of similarity, with edition-by-edition changes rarely exceeding .3. The pace of change is largely independent of author, outside an increase in the rate of change following the death of Emily Post. These results suggest that changes in *Etiquette* do not merely reflect changes in authorship. Given etiquette books' purpose, it is reasonable to assume that at least some of the changes in the texts represent mainstream shifts.

We continued exploratory work by looking at the presence and depiction of immigrants in *Etiquette*. We first constructed a list of immigrant keywords by drawing words from contemporary and historical dictionaries, thesauri, and keyword generation tools. These exploratory analyses yielded few explicit mentions and no clear trends. Likewise, exploratory topic modeling and reading explicit mentions uncovered just a few topics, such as foreign travel, hosting foreign guests, choosing between continental or American manners (e.g., the style of holding one's fork and knife), and the correct

Table 2. Document Similarity Matrix, by Edition.

	1922	1927	1931	1934	1937	1940	1942	1945	1950	1955	1960	1965	1968	1975	1984	1992	1997	2004	2011	2017
1922	1.000																			
1927	0.920	1.000																		
1931	0.920	0.990	1.000																	
1934	0.910	0.985	0.994	1.000																
1937	0.856	0.923	0.932	0.934	1.000															
1940	0.866	0.934	0.943	0.944	0.994	1.000														
1942	0.891	0.939	0.944	0.934	0.934	0.947	1.000													
1945	0.853	0.909	0.922	0.922	0.948	0.958	0.946	1.000												
1950	0.842	0.897	0.910	0.912	0.939	0.950	0.931	0.982	1.000											
1955	0.852	0.896	0.903	0.898	0.904	0.917	0.947	0.947	0.955	1.000										
1960	0.819	0.875	0.882	0.889	0.898	0.908	0.906	0.939	0.945	0.967	1.000									
1965	0.759	0.817	0.814	0.807	0.805	0.818	0.861	0.847	0.855	0.897	0.871	1.000								
1968	0.671	0.721	0.740	0.734	0.753	0.769	0.750	0.789	0.797	0.795	0.803	0.859	1.000							
1975	0.685	0.748	0.751	0.745	0.751	0.763	0.779	0.774	0.780	0.813	0.796	0.907	0.843	1.000						
1984	0.648	0.695	0.704	0.705	0.712	0.723	0.732	0.750	0.755	0.774	0.774	0.848	0.815	0.919	1.000					
1992	0.613	0.659	0.661	0.656	0.664	0.674	0.705	0.703	0.709	0.744	0.727	0.845	0.780	0.903	0.927	1.000				
1997	0.569	0.613	0.629	0.630	0.668	0.676	0.644	0.703	0.708	0.692	0.719	0.755	0.803	0.822	0.866	0.899	1.000			
2004	0.437	0.469	0.481	0.488	0.531	0.534	0.498	0.552	0.551	0.545	0.613	0.580	0.625	0.640	0.677	0.687	0.788	1.000		
2011	0.393	0.421	0.422	0.420	0.433	0.440	0.464	0.458	0.460	0.486	0.474	0.566	0.508	0.598	0.611	0.656	0.611	0.664	1.000	
2017	0.374	0.398	0.402	0.399	0.411	0.418	0.445	0.435	0.437	0.458	0.450	0.527	0.483	0.584	0.581	0.618	0.580	0.626	0.961	1.000

Key: □ Author Transition.
Edition by edition similarity. Author Transitions indicated by bordered cells. Higher values and darker shading represent more similarity. Lower values and lighter shades represent less similarity. Indicating less consistency across adjacent editions over time. Substantial by-edition change is not confined to author transitions. Editions beginning with 2004 represent the least consistency with all prior editions.

use of foreign expressions while speaking English. These early analyses led us to conclude that it was difficult to locate explicit evidence of immigrants and immigrant integration in *Etiquette* through keyword-based analyses.

Exploratory analyses also included independent human reading and analysis of *Etiquette*. While the computational exploration turned up few patterns, we determined through human reading that mainstream shifts indicating incorporation are, in fact, present. These mainstream shifts are visible in the inclusion of new religions, foods, cultural practices, and non-Anglo-Protestant names in the texts, among other things. For example, in 1937 Post first instructed readers on how to address and introduce both priests and rabbis, and in 1997 readers were first counseled how to behave at a religious service in a mosque and informed about what to expect at a quinceañera. Bar and bat mitzvahs, appropriate attire for Catholic church services, and what to expect at 100-day parties for babies were added and developed as topics as *Etiquette* was revised in light of its role as an encyclopedic guide to navigating social situations its readers might encounter. So, although preliminary computational exploration focused on immigrant keywords was ineffective at identifying mainstream shifts associated with immigrant integration, immigrant incorporation is visible through changes in the texts that are implicitly or explicitly associated with particular non-Anglo-Protestant immigrants and their descendants.

Based on qualitative and computational exploration, we determined that we could uncover mainstream shifts by examining *focal group concepts* consisting of the bundles of people, practices, identifications, and objects associated with current or former immigrant groups. We do this by analyzing the semantic contexts in which the focal group concepts appear in *Etiquette* and considering how those contexts change across editions.

We approached the main analysis with two expectations derived from the exploratory work. First, we expect incorporation to be reflected in increased “normal”-ness in the semantic contexts of the focal group concepts. For example, let’s return to the addition of quinceañeras in the 1997 edition. At that time, and again in the 2004 edition, the discussion of quinceañeras presumed that the reader would likely encounter the event as a guest and not as a host (see Table 3). This is clearly visible in comparison with the discussion of the sweet-sixteen celebration, which describes how one should arrange the invitations and the meal. Meanwhile, the discussion of quinceañeras focuses on what a guest who is less familiar with such events needs to know, including appropriate attire and whether or not gifts are expected. In this case, we see the representation of “foreign” practices in *Etiquette*, but not the inclusion of people who would hold a quinceañera in *Etiquette’s* readership.

Table 3. Discussion of Quinceñara and Sweet-Sixteen Tradition, 1997.

<p>Quinceñara The celebration of a Latin American girls' fifteenth birthday, a quinceñara, often simply called a quince, can be either a party or a religious celebration followed by a party, depending on the practice of the Roman Catholic Church in the area...</p>	<p>Sweet-sixteen Parties A sixteenth birthday is a big milestone—really the division between childhood and young adulthood. Therefore it is often celebrated more elaborately than other birthdays, with a sweet-sixteen party...</p>
<p>Not dissimilar in concept to a debutante ball, the celebration commemorates coming-of-age or a young girl's passage into womanhood. A highlight of the party usually includes the tradition of the birthday girl waltzing with her father before general dancing begins.</p>	<p>(...) Invitations may be telephoned, but when the celebration is more formal, they are usually written on decorated, fill-in commercial invitations, which can be found, if you wish, specifically for sweet-sixteen parties...</p>
<p>Dress Very often the party is formal, with guests wearing either black tie or white tie attire, as noted on the invitation, and the young girl wearing an elaborate, formal gown.</p>	<p>All invitations should have "R.s.v.p." on them, followed by a telephone number. Telephoning a response seems easier to phone-prone youngsters than writing a reply, and this will elicit more and quicker answers.</p>
<p>Gifts Gifts are expected, either money or personal items, depending on the tradition of the area, and usually are not opened during the party, which follows the form of a ball or dance rather than a traditional birthday celebration (Post 1997: 566–567).</p>	<p>Refreshments If you are serving a meal—luncheon or dinner—the menu should simply be the favorite food of the birthday girl, although nothing so exotic that it will not appeal to the majority of the guests... (Post 1997: 567)</p>

Quinceñaras are described for someone who is unfamiliar with the event and attending as a guest. Sweet-sixteen parties are described for those who will be hosting such an event.

Table 4 shows a shift in *Etiquette's* mainstream readership. The 2011 edition of *Etiquette* presumes it has readers who need advice on *hosting* a quinceañera. An altered description of the event says a "quince" originates in Latin America, instead of being something that "Latin American girls"

Table 4. Discussion of Quinceañera, 2011.

Invitations are issued by the honoree's parents or by the honoree and her parents.

For an informal quinceañera a phone call, personal note, or informal card are good invitational choices. But for a formal event invitations may be printed or engraved and might include a "Black tie" or "White tie" notation.

A printed, formal invitation to the religious service would also include a reception card inviting guests to the quinceañera party afterward. Since the invitation itself is to a religious service, it does not include an RSVP, but the reception card does.

It's customary to take gifts to a quinceañera party, but presents aren't opened at a formal ball. Religious items such as rosaries or crosses are appropriate and traditional; monetary gifts and personal items for teenage girls are also popular.

The young honoree must send written thank-you notes to all friends and family who gave her a gift. (Post et al. 2011:318).

Discussion assumes the reader may be hosting the event.

do. Furthermore, the text offers advice related to hosting the event. The book also includes a sample quinceañera invitation in the appended resources. In the case of the quinceañera, we observe mainstream shifts reflecting increased "normal"-ness attached to cultural practices associated with immigration.

We also expect increased positive sentiment to coincide with mainstream shifts indicating incorporation. Sentiment may be accessible through human reading. For example, there is an underlying negative view of immigrants in this quotation from the 1937 and 1942 editions: "[A child] must never be allowed to hold his fork immigrant fashion, perpendicularly clutched in the clenched fist, and to saw across the food at its base with his knife" (Post 1937:745). Computational tools can follow qualitative insights by exploring latent sentiment in texts (Prabowo and Thelwall 2009). During exploratory research, we conducted a basic sentiment analysis of the excerpts presented in Tables 3 and 4.² These results confirm our qualitative interpretations: there is more positive sentiment in the texts around quinceañeras in 2011 than in 1997. However, the positive sentiment in the 1997 excerpt about Sweet 16 celebrations is higher than in either quinceañera excerpt.³

In planning the main analysis, we determined that we would consider mainstream shifts related to long-standing group categories that were and may still be seen as immigrant outsider categories in the United States: Catholic, Chinese, Irish, Italian, Jewish, Mexican, and Muslim. We selected these particular focal groups because we had a sense from reading that they were present in the texts. Moreover, they are associated with different historical waves of migration from both European and non-European contexts, and,

as discussed in the theory section, they have different trajectories when it comes to incorporation through “becoming white”.

Analysis and Results

Computational social scientific analysis of text data extends the possibilities for research on mainstream shifts. A computational approach can include a larger volume of text than is feasible for hand-reading, subjects qualitative interpretations of texts to additional scrutiny through the application of quantitative methods designed with interpretation in mind, and leads to the development of research techniques that can be applied to other empirical questions and text data (Nelson 2020). We came to the final results through theory-based interpretation, examination of the texts in their historical moments, and robustness checks. This interpretive and confirmatory work is presented along with the presentation of results.

Two points must be emphasized concerning the analysis. First, our study examines immigrant incorporation in an interpretive and symbolic sense. It does not include measures or estimates of immigrant incorporation in a material sense (e.g., rates of intermarriage or citizenship acquisition). Second, immigrant incorporation is mainly implicit in *Etiquette*. Politics and public policy, including the topics of immigration and citizenship, are rarely addressed. Instead, etiquette books describe everyday manners and offer counsel for behavior during key moments. However, as Jiménez’s (2017) investigation of established individuals shows, mainstream adaptations in light of immigration occur precisely within the mundane world of everyday life: workplace relationships, food and fashion, norms of communication, familiarity with different religions, and the rituals that punctuate the life course from birth to death. The *Etiquette* corpus is well-suited to revealing the scope of immigrant incorporation in relation to these elements of the mainstream.

Identifying Focal Group Concepts Through the Creation of Pseudocorpora

For the analysis, we created a pseudocorpus for each focal immigrant group by sampling texts from *Etiquette* that were more related to the group. To do this, we used the word embedding method Concept Movers Distance (CMDist; Stoltz and Taylor 2019). Word embeddings leverage the connection between particular words and their linguistic contexts (Ellis 2019; Firth 1957; Garvin 1962; Lenci 2018; Stoltz and Taylor 2019). The underlying premise is that language can be conceptualized as a multidimensional

map of words that reflects all the contexts in which words are used. Word embeddings are trained over very large text corpora. Through that training, they develop a detailed account of the interrelationships of words. Words used in similar contexts are presumed to have a similar meaning, which is quantified as cosine similarity. Even when words do not appear together but are used in similar ways, they are assumed to have similar meanings. CMDist uses cosine similarity scores from word embeddings to estimate the salience of a focal concept. CMDist does this by calculating the distance between all words in a document and the word or words denoting that focal concept (Stoltz and Taylor 2019). CMDist makes it possible to estimate the salience of implicit concepts that do not appear directly in the texts, making it particularly useful to this research since the exploratory analyses turned up few explicit references to immigration and immigrant groups but many examples of implicit immigrant integration through mainstream shifts.

To select parts of *Etiquette* in which various immigrant focal groups are most salient, we first parsed every edition into paragraph-length 150-word chunks. We then specified each focal immigrant group as a concept. For clarity, we will use italics when referring to the focal groups (*Catholic, Chinese, Irish, Italian, Jewish, Mexican, and Muslim*) when they are operationalized as focal group concepts in CMDist. CMDist assigns each chunk a standardized score for each focal group concept. These CMDist scores estimate how salient the focal group concept is to the text in the chunk (Stoltz and Taylor 2019). Scores generally range between -3 and 3 . Chunks with scores exceeding 1.65 – corresponding to roughly 5 percent of the text – were then aggregated into concept-specific pseudoeditions to be used in subsequent analyses. Table 5 presents the count of salient chunks for each immigrant group. The proportion of salient chunks in the corresponding edition is shown in parentheses.

Embedding selection, including considerations of historical period and the style and genre of the texts used to train the embedding, is important (Stoltz and Taylor 2021). Since CMDist takes embeddings as its foundational building block, CMDist results are only as strong as the embeddings used to create the model. The researcher is responsible for justifying the applicability of their embeddings in their particular context (Stoltz and Taylor 2019; Stoltz and Taylor 2021). We use a combination of historical word embeddings from Google N-Grams and FastText to select the chunks for the pseudocorpora. For editions published prior to 2000, we use decade-specific Google N-Gram embeddings (Timm 2019). Google N-Gram embeddings are trained using books that Google has digitized. While limited by a variety of selection factors (Hamilton, Leskovec, and Jurafsky 2016), historical

Table 5. Count and Proportion of Chunks Selected from Full Corpus That Compose Pseudo Corpus for Each Immigrant Group.

Edition	1922	1927	1934	1937	1940	1942	1945	1950	1956	1960	1965	1969	1975	1984	1992	1997	2004	2011	2017
Italian	21 (.043)	31 (.047)	37 (.055)	44 (.053)	43 (.051)	44 (.046)	41 (.058)	29 (.042)	35 (.046)	37 (.049)	32 (.045)	41 (.061)	54 (.056)	48 (.052)	46 (.054)	53 (.061)	64 (.064)	54 (.060)	53 (.060)
Jewish	34 (.070)	39 (.059)	21 (.031)	33 (.040)	39 (.046)	39 (.041)	32 (.045)	28 (.040)	40 (.052)	47 (.063)	42 (.058)	40 (.059)	34 (.035)	36 (.039)	37 (.043)	41 (.047)	52 (.052)	44 (.049)	45 (.051)
Irish	25 (.051)	39 (.059)	34 (.050)	41 (.050)	50 (.059)	61 (.064)	46 (.065)	49 (.070)	35 (.046)	40 (.053)	36 (.050)	39 (.058)	54 (.056)	55 (.060)	68 (.080)	71 (.082)	58 (.058)	46 (.051)	45 (.051)
Catholic	27 (.055)	37 (.056)	38 (.056)	48 (.058)	45 (.053)	46 (.048)	36 (.051)	41 (.059)	33 (.043)	41 (.055)	41 (.057)	42 (.062)	54 (.056)	53 (.057)	54 (.063)	55 (.064)	42 (.042)	48 (.053)	49 (.056)
Mexican	24 (.049)	28 (.043)	13 (.019)	23 (.028)	39 (.046)	54 (.057)	37 (.052)	29 (.042)	38 (.050)	35 (.047)	30 (.042)	37 (.055)	45 (.047)	51 (.055)	43 (.050)	47 (.054)	54 (.054)	57 (.063)	53 (.060)
Chinese	31 (.064)	32 (.049)	26 (.039)	30 (.036)	47 (.056)	58 (.061)	43 (.061)	30 (.043)	37 (.049)	33 (.044)	30 (.042)	33 (.049)	50 (.052)	46 (.050)	57 (.067)	59 (.068)	50 (.050)	53 (.059)	48 (.055)
Muslim	29 (.059)	35 (.053)	38 (.056)	47 (.057)	29 (.034)	34 (.036)	23 (.032)	32 (.046)	28 (.037)	30 (.040)	38 (.053)	40 (.059)	40 (.041)	34 (.037)	46 (.054)	54 (.062)	50 (.050)	53 (.059)	47 (.054)

embeddings may better reflect implicit meanings in texts from their particular period, increasing accuracy when selecting chunks that are salient to each focal group concept. For twenty-first century editions, pseudoeditions are created using FastText embeddings. FastText embeddings are pretrained on a larger set of sources than Google N-Grams, including Wikipedia and common crawls of the web (Bojanowski et al. 2017; Joulin et al. 2016). Prepared by Facebook's AI research team, FastText is among the most reliable and widely-used of embeddings.⁴

Selecting a salience cutoff requires researchers to balance the tradeoff between concept specificity and robustness. To our knowledge, there is no standardized best-practice for assessing this potentially-important tradeoff in the context of varying text types and theoretical goals. The general premise is that a higher cutoff level will select chunks that more closely relate to the focal group concepts in question, at the expense of smaller corpora. Recent research using CMDist to select "salient" texts have considered scores of "1" as "high," (Carbone and Mijis 2022). We use a salient cutoff of 1.65, representing roughly 5 percent of the sample on one side of a normal distribution, and employ various robustness checks to assess that decision. See online Appendix A for details regarding how we selected the appropriate cutoff.

We also examined the validity of the selected chunks. Since we know of no way to quantitatively establish the validity of chunks selected based on concept salience, we relied upon a variety of qualitative and interpretive checks. It is important to note that validity in this case does not mean that a salient chunk needs to be about Irishness, Mexican-ness and etc. Focal group concepts are in "the background" in some salient chunks. The value of our method is precisely in leveraging word embeddings' ability to access background meanings. Analyzing meaning through semantic contexts is the unique contribution of embedding approaches over human analysis or dictionary/keyword-based approaches. If salience were synonymous with explicit mentions, there would be no value added relative to keyword searches. If the association between the focal group concepts and the texts was always apparent to a reader, the only benefit of using word embeddings to select the text would be the ability to manage larger amounts of text.

To check validity, we sought to establish 1. that the pseudocorpora were plausibly connected to the focal groups, and 2. that they were somewhat unique to the focal groups. In terms of plausible connection, we confirmed that parts of *Etiquette* we previously identified through reading (e.g., the sections on quinceañeras) were also selected into the pseudocorpora. We also examined the salient texts. Using the word tokens in a selection of

salient chunks, we located the text excerpts and considered the connection to the focal groups. Some connections were implicit. Take, for example, the following implicit reference to *Catholic* (through the mention of rosaries) in a discussion of appropriate gifts from the 2004 edition (Post 2004, Chunk 3315, page 508):

perfume; or a scarf, belt, or other fashion accessory.

For quinceanera parties

For quinceanera—a girl's fifteenth-birthday celebration in Latin cultures—religious items such as rosaries or crosses are appropriate. Monetary gifts and personal items for teenage girls are also popular gifts. (See also Chapter 16, page 214: "Quinceanera.")

Explicit mentions referred to the focal group concepts directly. Take, for example the explicit reference to *Italian* immigrants and food from the 1975 edition (Post 1975, Chunk 254, page 33):

There are many foreign words that have become an accepted part of our language, and we should be familiar with their meaning and pronunciation. We cross paths so often with the many foreigners living in the United States that it is easy to increase one's knowledge of their languages. We have Italian grocers, German bakers, French and Spanish waiters—to name only a few—all of whom are delighted if you show an interest in their language, how it sounds, and what it means. In fact, if you ask the waiter how to pronounce "parmigiana", he very well may tell you not only how to say it, but how to cook it. Everyone loves to be an expert, and foreigners as well as Americans like to have you express interest in their language and their customs.

However, the connection to the focal group concept is not always clear to a human reader, as in the case of this salient chunk for *Jewish* from the 1965 edition (Chunk 1887, page 311):

On occasion, divorced parents may remain good friends and their daughter's time may be divided equally between them. If this is true, they may both wish to announce the engagement.

Mr. Gordon Smythe of Philadelphia, and Mrs. Howard Zabriskie of 12 East 72nd Street, New York City, announce the engagement of their daughter, Miss Carla Farr Smythe...

Since the word embeddings were generally effective in identifying clearly related chunks, we retained chunks like this one where the connection was not immediately apparent (see previous discussion of the unique value of word embeddings and online Appendix B for examples of salient chunks for all focal groups).⁵

As these excerpts show, salient chunks can be selected by multiple focal group concepts. The text about rosaries, and the text about foreigners living in the United States were salient for *Mexican* as well as *Catholic* and *Italian* – and for good reason since rosaries are discussed as a quinceañera gift, and Italian grocers are discussed alongside Spanish (a term often used in *Etiquette* to refer to Latin America) waiters. To some extent, this is an artifact of our corpus, which is relatively small and takes up a set number of topics relevant to everyday manners. Overlaps are also in tune with our theory, which posits that the intermingling of focal group concepts is to be expected with immigrant incorporation – for example, including quinceañera gifts on a list of presents for high moments one is likely to encounter.

However, in terms of validity, it is important to establish the differences in how different focal group concepts are implicitly and explicitly present in *Etiquette*. As a robustness check, we examined the uniqueness of each focal group concept's pseudocorpus. Table 6 presents the 60 most common tokens (words), ranked by frequency in each pseudocorpus within 25-year periods (1920–1945; 1945–70; 1970–1995; 1995–2020). To visualize these distinctions, we split the list into three tiers per period, the 20 most frequent words, and terms ranked 21–40 and 41–60 in frequency. The highlighted words in the table are unique to the focal group concept's respective pseudocorpus in that period and within that tier. For example, “wedding” is only present in the top 20 words of the *Catholic* pseudocorpus in period 1, so it is highlighted. By period 4, “wedding” ranks top 20 in all pseudocorpora, so it is not highlighted. These descriptive measures demonstrate the pseudocorpora are, to some degree, distinct, although some concepts are less uniquely represented.⁶ These analyses also demonstrate stability and change in the generalized contexts present in pseudocorpora. Although there is overlap in frequent words, there are clear differences in the actual frequency of those words. These focal-group-concept-specific differences in word frequency are reflected in the sentiment analyses and salience scores in CMDist.

Table 6. Top 60 Most Frequent Terms, by Group and Period, with Groups' Unique Terms in First, Second and Third 20 Highlighted.

rank	Catholic								Chinese								Irish				Italian			
	p1	p2	p3	p4	p1	p2	p3	p4	p1	p2	p3	p4	p1	p2	p3	p4	p1	p2	p3	p4	p1	p2	p3	p4
1	house	church	church	religious	people	american	formal	dinner	formal	john	woman	formal	dinner	people	john	woman	formal	people	formal	woman	formal	people	american	american
2	people	family	wedding	service	house	social	business	table	party	united	social	business	table	house	title	formal	address	house	woman	address	american	hand	formal	united
3	family	house	ceremony	ceremony	american	united	party	food	dinner	president	social	business	don	formal	wife	class	don	formal	social	united	american	service	president	
4	woman	father	family	catholic	class	country	social	page	address	house	house	party	page	woman	house	miss	business	franch	business	united	franch	house	address	
5	class	wedding	home	church	words	formal	house	held	cards	foreign	cards	served	party	country	jones	miss	names	class	names	business	franch	country	address	
6	mother	john	parents	home	country	day	house	held	party	hand	foreign	business	page	american	house	miss	business	class	names	business	franch	country	address	
7	social	mother	friends	held	hand	foreign	house	family	cards	day	foreign	served	party	american	jones	miss	business	class	names	business	franch	country	address	
8	girl	friends	service	family	day	foreign	house	held	party	hand	foreign	served	party	american	jones	miss	business	class	names	business	franch	country	address	
9	friends	parents	page	jewish	person	rank	united	business	business	person	day	business	party	american	house	miss	business	class	names	business	franch	country	address	
10	daughter	address	cards	parents	section	official	united	business	business	person	day	business	party	american	house	miss	business	class	names	business	franch	country	address	
11	smith	wife	class	address	social	letter	home	eating	guests	lady	mary	united	correct	miss	husband	church	women	woman	woman	woman	woman	woman	woman	
12	american	daughter	woman	formal	lady	english	wedding	guests	social	matter	correct	names	cards	mary	united	correct	names	mary	united	correct	names	mary	united	
13	husband	mary	address	wedding	world	friends	foreign	wedding	social	matter	correct	names	cards	mary	united	correct	names	mary	united	correct	names	mary	united	
14	mary	names	social	guests	french	people	woman	home	wedding	service	miss	husband	church	women	woman	woman	woman	woman	woman	woman	woman	woman	woman	
15	wedding	united	formal	children	dinner	words	conversations	meal	plate	section	gentleman	house	dress	women	woman	woman	woman	woman	woman	woman	woman	woman	woman	
16	day	ceremony	marriage	gifts	dinner	words	conversations	meal	plate	section	gentleman	house	dress	women	woman	woman	woman	woman	woman	woman	woman	woman	woman	
17	country	catholic	business	party	friends	title	friends	plate	plate	page	daughter	heading	religious	lady	lady	lady	lady	lady	lady	lady	lady	lady	lady	
18	formal	formal	guests	social	united	word	parties	american	american	united	friends	names	home	table	table	table	table	table	table	table	table	table	table	
19	wife	husband	section	business	white	class	family	class	class	correct	president	section	house	section	section	section	section	section	section	section	section	section	section	
20	york	letter	catholic	traditional	english	correct	class	service	service	person	social	family	traditional	formal	formal	formal	formal	formal	formal	formal	formal	formal	formal	
21	father	reception	father	services	foreign	ambassador	gifts	family	family	call	family	conversation	country	friends	friends	country	friends	friends	country	country	lady	lady	lady	
22	home	social	running	holy	service	chief	honor	white	white	words	lady	dear	dear	guests	guests	dear	guests	guests	dear	dear	secretary	secretary	secretary	
23	lady	american	heading	house	society	service	country	dress	dress	mother	mother	friends	parents	country	country	parents	country	country	parents	parents	chief	chief	chief	
24	title	class	religious	official	women	visiting	official	serving	serving	york	country	marriage	party	person	person	party	person	person	party	party	sir	sir	sir	
25	section	correct	held	page	matter	washington	page	traditional	business	called	called	title	title	business	business	title	business	business	church	church	john	john	john	
26	church	woman	mother	christian	title	woman	parents	women	women	smith	smith	party	style	home	smith	party	family	family	family	family	foreign	foreign	john	
27	page	friend	house	close	table	honor	church	style	style	friend	friend	sir	service	hand	hand	hand	hand	hand	hand	hand	hand	hand	hand	
28	called	marriage	close	person	america	secretary	section	cards	cards	form	father	guests	hand	correct	correct	correct	correct	correct	correct	correct	person	person	person	

(continued)

Table 6. Continued.

rank	Catholic				Chinese				Irish				Italian			
	p1	p2	p3	p4	p1	p2	p3	p4	p1	p2	p3	p4	p1	p2	p3	p4
29	join	title	dear	wear	form	world	table	conversation	public	sir	john	person	united	visiting		
30	bride	close		couple	york	vice	office	dishes	secretary	ambassador	official	white	home	wife		
31	call	home	couple	day	formal	written	public	fork	smith	rank	foreign	dinner	title	governor		
32	matter	jones	special	father	guests	invitation	running	gifts	table	class	service	guests	left	speech		
33	form	page	jewish	reception	public	page	children	religious	hand	addressed	ceremony	held	wedding	invitation		
34	taste	cards	children	united	wedding	french	heading	chapter	honor	call	country	office	party	jones		
35	hand	dear	honor	priest	page	person	sir	hand	rank	cards	women	form	called	office		
36	miss	children	miss	woman	beautiful	speak	american	set	wife	introduction	form	sir	family	washington		
37	society	smith	party	friends	family	home	ceremony	friends	addressed	dear	held	catholic	page	word		
38	takes	takes	relatives	officials	head	day	personal	parties	guests	secretary	closing	marriage	mother	cards		
39	united	invitation	sir	office	introduction	proper	couple	country	names	page	father	miss	rank	church		
40	possibly	service	married	roman	takes	representatives	person	eat	chapter	wedding	honor	married	sir	class		
41	president	day	closing	expected	taste	dinner	dear	host	manners	official	daughter	mother	takes	introduction		
42	dress	honor	takes	marriage	wear	family	held	house	society	introduced	dinner	class	ambassador	addressed		
43	person	miss	wear	women	woman	write	washington	restaurant	women	form	dinner	don	life	day		
44	dinner	roman	women	cards	correct	army	women	united	ambassador	person	forms	couple	beautiful	dear		
45	names	sir	invited	class	evening	cards	closing	ceremony	friends	church	married	president	dining	justice		
46	public	president	white	dear	rank	hand	gift	heading	john	taste	washington	addressed	manners	page		
47	service	special	invitation	life	home	life	guest	knife	letters	chief	city	day	public	representatives		
48	address	form	united	mother	called	cabinet	jewish	left	spoke	words	husband	foreign	america	senator		
49	children	country	christian	white	dress	close	names	times	address	york	request	respect	honor	service		
50	friend	envelope	daughter	president	party	office	invitation	include	running	vice	smith	washington	word	written		
51	left	guests	day	title	ambassador	private	white	section	office	honor	addressed	card	world	army		
52	parents	hand	front	churches	americans	countries	forms	children	french	honor	children	ceremony	bride	persons		
53	running	husband	follow	names	flowers	speaker	marriage	day	friend	representatives	mother	city	husband	conversation		
54	honor	addressed	minister	names	left	america	travel	respect	personal	spoke	secretary	called	means	speech		
55	letter	called	official	conversation	letters	guests	written	woman	card	visiting	white	children	meet	wedding		
56	introduction	relatives	reception	sir	manners	lady	countries	address	foreign	foreign	american	john	set	friends		
57	married	visiting	send	child	clothes	persons	eating	casual	introduction	hand	club	honor	americans	letters		
58	means	ambassador	dress	dress	gentleman	special	president	eat	mary	justice	jones	roman	evening	navy		
59	addressed	card	office	age	word	speech	rules	dining	mother	envelope	office	cards	father	people		
60	correct	person	roman	days	guest	york	tea	personal	cards	gentleman	couple	close	gentleman	york		

(continued)

Table 6. Continued.

Italian				Jewish				Mexican				Muslim			
p3	p4	p1	p2	p3	p4	p1	p2	p3	p4	p1	p2	p3	p4	p1	p2
formal	formal	house	family	church	family	house	united	social	formal	house	country	formal	religious	house	country
wedding	dinner	people	parents	ceremony	parents	class	american	united	dinner	people	house	social	service	family	house
guests	page	class	ceremony	home	children	president	president	formal	don	class	family	formal	ceremony	social	family
dinner	don	social	religious	home	religion	united	formal	address	american	social	united	woman	home	woman	woman
class	guests	friends	service	jewish	service	rank	social	class	food	friends	social	address	wear	address	wear
home	service	page	church	service	friends	american	official	woman	table	society	president	church	guests	church	guests
page	dress	family	social	class	jewish	formal	rank	page	page	girl	american	class	jewish	home	party
church	table	table	children	friends	home	people	address	miss	served	section	friends	cards	party	cards	wedding
party	day	chapter	class	parents	ceremony	title	house	wedding	united	country	class	friends	family	friends	family
united	served	chapter	home	page	close	chief	title	running	women	hand	guests	guests	social	guests	social
running	american	country	life	wedding	wedding	official	letter	ambassador	business	home	formal	page	women	page	women
heading	business	world	father	social	couple	wife	ambassador	conversation	party	secretary	woman	parents	day	business	day
section	white	formal	day	catholic	child	lady	country	home	woman	world	secretary	business	services	house	services
served	wedding	section	country	guests	person	ambassador	john	home	guests	dinner	vice	house	held	house	held
address	food	life	people	religious	gifts	section	correct	official	home	dinner	chief	house	held	house	held
table	style	public	wedding	business	day	correct	chief	title	wedding	form	chief	service	parents	service	parents
house	traditional	york	page	cards	guests	day	secretary	dear	dress	person	women	ceremony	dress	ceremony	dress
service	class	dinner	world	formal	don	secretary	vice	foreign	plate	president	life	party	formal	party	formal
business	religious	person	running	held	held	washington	foreign	heading	service	guests	wedding	section	friends	section	friends
woman	women	society	special	section	marriage	service	honor	section	eating	service	wife	wedding	children	wedding	children
friends	social	american	families	special	catholic	form	service	sir	meal	woman	letter	heading	traditional	heading	traditional
ceremony	hand	home	girl	children	church	home	washington	house	white	page	rank	held	catholic	held	catholic
parents	home	service	husband	house	wear	social	wife	country	class	chapter	world	united	gifts	united	gifts
foreign	held	taste	service	running	party	army	justice	parents	host	family	address	conversation	church	conversation	church
honor	plate	gentleman	women	heading	page	country	sir	women	social	lady	home	running	close	running	close
social	casual	matter	american	party	relatives	honor	woman	church	family	table	official	sir	person	dinner	person
family	country	woman	child	address	families	justice	governor	marriage	fork	reason	parents	dinner	business	catholic	business
marriage	family	manners	live	close	traditional	letter	senator	names	serving	united	day	catholic	page	page	page
conversation	meal	mother	ceremony	woman	father	addressed	york	washington	hold	called	guest	dear	class	dear	class

(continued)

Table 6. Continued.

Italian				Jewish				Mexican				Muslim					
p3	p4	p1	p2	p3	p4	p1	p2	p3	p4	p1	p2	p3	p4	p1	p2	p3	p4
held american plate white cards children couple serving country father	host wine united ceremony chapter day guest heating running serving	wedding women club guests american intimate lady business tea	daughter guests woman wife friend heading names section hand person	day father relatives wear faith mother takes invited marriage	life services class age christian days social tradition birth couples	call called public vice chapter navy officers dinner friends guests	wedding army class jones cards cabinet representatives dear navy	closing addressed city family father office children forms friends john	traditional address country eat official wear cut hand heading section	formal american rank friend possibly public means personal york chief	page persons person service foreign honor invitation title ceremony	foreign honor minister official religious jewish special closing day	reception holy don life woman christian couple expected marriage address				
food form guest miss dear hand sir forms invitation john butter hot left mother person simply wine coffee jewish names prefer	person wear choice evening conversation respect section serve dishes event left times woman include including salad friends wine occasion set bow church	guest letters parents words means position takes rule book called form houses word include church difficult hand live occasion private stranger girl	reception relatives united called personal takes business catholic girls marriage leave matter write age difficult follow public taste care letter	christian city couple white conversation dear honor office official dinner foreign minister united chapter child conservative gifts parties telephone churches closing	reception respect women attend include parent conservative feelings follow gift mother special expected formal heading personal holy takes adult bride dress	introduction letters office page court foreign table cabinet governor senator sir address dear guest hand military naval person write husband words	office white called home introduction visiting dinner form friends minister page smith miss smith lady names speaker court family military city countries	husband life married service white american honor information personal smith chief daughter dr justice secretary york date james living nations	conversation foreign knife style ceremony dishes office day religious cards person honor serve friends guest nations president black casual evening	private taste life persons proper ambassador call matter position takes washing words wife party rules secretary york nations president black casual evening	representatives special club party running speaker takes people correct washing private cabinet catholic city justice york women america court houses dinner	father office miss washing women addressed children close couple rules invited mother parties person wear american city forms marriage chief child	respect tradition conservative days head united bride prayer attend gift white father follow heading special house priest faith takes age marriage chief				

Most frequent terms in the pseudocorpus for each edition for 4 25-year periods (p1: 1920-1945; p2: 1945-70; p3: 1970-1995; p4: 1995-2020). List is divided into three tiers. Terms that are unique to a group within that tier and period are highlighted.

That meaning is relational is a core principle of the word embedding method. We see this as a strength of embeddings. However, there are trade-offs in comparison with more direct methods. For example, in the case of the selection and analysis of only texts that explicitly mention the focal group concepts, we would be able to establish that all selected texts engaged the focal groups explicitly, and we could leave out texts that also include other focal groups to isolate that one group for the purposes of analysis. With the embedding method, even with robustness checks, we cannot completely rule out that some changes in sentiment and strange-to-normal associations could be unrelated to our focal group concepts.

We use the pseudocorpora to assess mainstream shifts in the texts that are salient to the focal group concepts. We first look at how the sentiment in these pseudocorpora changes over time. We then look at how these pseudocorpora have moved along a single dimension represented by the binary poles of “strange” and “normal”. We present results in two ways: combined and concept-specific. Combined results demonstrate the overall change across all the pseudocorpora. Concept-specific results demonstrate the overall change *within* each focal group concept’s pseudocorpus. It is important to note that there is no shared absolute middle position between strange and normal or negative and positive sentiment across focal-group-specific pseudocorpora. This means that we cannot compare CMDist scores across pseudocorpora, as the point of reference against which scores are standardized is unclarified. It is nonetheless possible to compare general trends within and across groups.

Sentiment

As evident in the differences between how sweet sixteen parties and quinceañeras are covered in *Etiquette*, it is not enough to establish the presence of focal group concepts in *Etiquette*. We expect increased positive sentiment to be an element of incorporation through mainstream shifts. Sentiment analysis uses pre-existing measures of the sentiment of words to establish the general emotive characteristics of the pseudocorpus (Prabowo and Thelwall 2009). Sentiment analysis on the pseudocorpus of salient chunks was conducted using the *SocialSent* package in R (Hamilton et al. 2016). *SocialSent* contains a collection of code and data for performing domain-specific sentiment analysis and includes historical sentiment lexicons. These lexicons include sentiment scores for the top 5000 words (excluding stop words) for all decades ranging from 1850–2000. Period-dependent sentiment lexicons are crucial since the sentiment of

words is influenced by both domain and sociohistorical context (Hamilton et al. 2016). Analyses that span many decades without considering changes in sentiment can be misled by modern-day understandings (Hamilton et al. 2016).

Sentiment analyses produce a sentiment score for each edition in the pseudocorpus. Since baseline sentiment varies across *Etiquette* editions, we present scores for the pseudoedition relative to the average sentiment of the pseudocorpora. Sentiment scores from the *SocialSent* package are obtained from an algorithm that propagates sentiment polarities from small seed sets that consider the context of use (for example, soft may have a negative connotation when speaking about intelligence but a positive connotation when speaking about bath towels). These propagated scores were then bootstrapped using $B = 50$ and 7 words per random subset to lessen the influence of corpus artifacts that might be present in a seed set (see Hamilton et al. 2016 for more information). Sentiment scores are finally represented as the average inferred sentiment across these bootstrap samples.

The sentiment analysis results are presented in Figures 2, 3, and 4. Results from simple regressions yielded significant coefficients across all 7 groups for the effect of year on sentiment scores ($p < .05$). Interestingly, the effect of year on *Muslim* and *Italian* sentiment is only significant when considered as a polynomial, suggesting, as depicted in Figure 4, that the century can be characterized first by decreasing, and then increasing positive sentiment ($p < .05$). We revisit this finding in the discussion.

Figure 2 shows trends in positive sentiment across all pseudoeditions. Though there is variation, sentiment is generally less positive from around the 1930s to the 1970s before becoming more positive. Positive sentiment increases at a greater rate after 2000.

Figure 3 presents within-concept sentiment plots for texts most salient to the focal group concepts *Italian*, *Jewish*, *Catholic*, and *Irish*. Based on the previous literature, we expect this set of focal concepts to show a clear trajectory of integration through mainstream shifts (Brodin 1998; Ignatiev 2009; Roediger 2005; Takaki 1993). Note that concept-specific plots show trends in sentiment across editions for an individual focal group concept, making it possible to view how sentiment in texts salient to the focal group concept in question have changed irrespective of trends in the other focal group concepts. While the general pattern of increasing positive sentiment in more recent editions after declining or stagnating positive sentiment is evident for all focal group concepts, there is variation across focal group concepts. For example, the sentiment around *Italian* and *Jewish* crosses into positive territory later than *Catholic* and *Irish*.

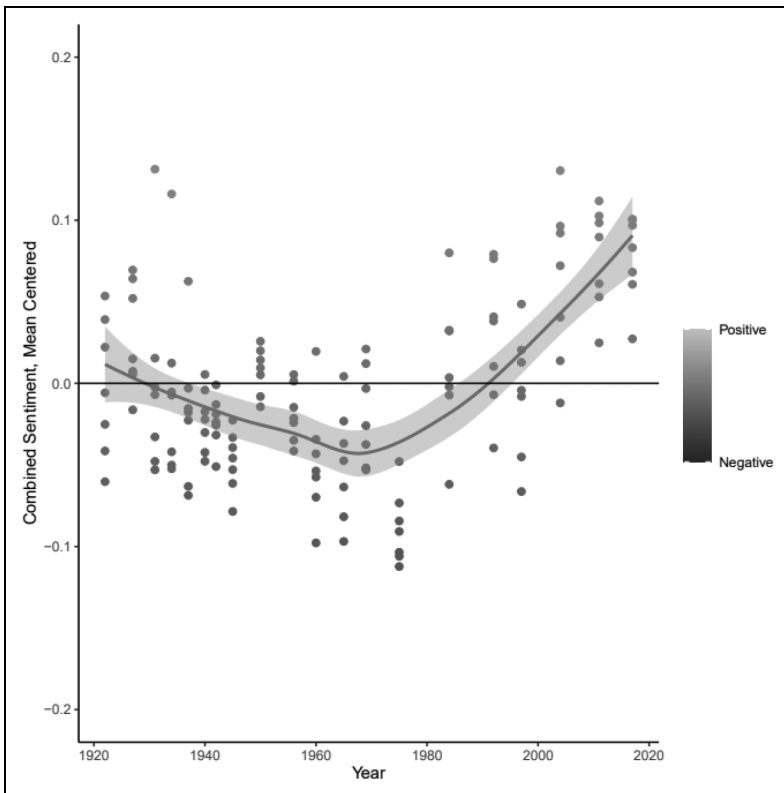


Figure 2. Sentiment: more or less positive relative to the edition for all focal groups, by edition year. Higher scores are more positive. Figure includes a LOWESS plot of median scores and a bar representing a 95 percent confidence interval.

Figure 4 plots the sentiment scores for texts most salient to the focal group concepts *Muslim*, *Chinese*, and *Mexican*. Based on the literature, we expect these focal concepts to show less evidence of integration through mainstream shifts (Lee 2003; Maghbouleh 2017; Massey, Durand, and Malone 2002; Takaki 1993). In this case, our observations were not consistent with our expectations. In the texts most salient to these focal group concepts there are also trends toward increased positivity after the 1970s. This is especially evident for texts salient to *Muslim* and, to a lesser extent, *Mexican*, focal group concepts. The shift into above-average positive sentiment does not fall outside the confidence interval for *Chinese*.

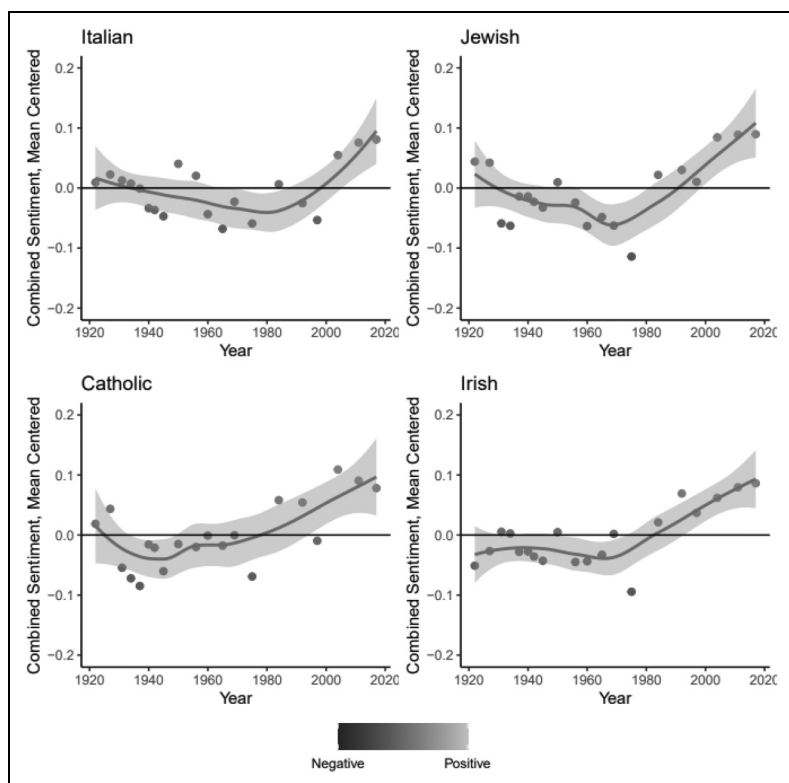


Figure 3. Sentiment: more or less positive relative to the edition for focal group set I, by edition year. Higher scores are more positive. Figure includes a LOWESS plot of median scores and a bar representing a 95 percent confidence interval.

Trends in sentiment are fairly consistent across the focal group concepts – both those that are often considered to be successfully incorporated into the mainstream and those that are not. This suggests a broader transformation in sentiment toward difference as an element of the mainstream over and above group-specific processes. However, there are limitations to consider in interpreting the results of the sentiment analyses. Although we have selected texts with high salience to the focal group concepts, we cannot be certain that these concepts drive the observed sentiment changes. As a robustness check, we conducted the same analyses for pseudocorpora built around various concepts that are neutral to the focal groups. These analyses

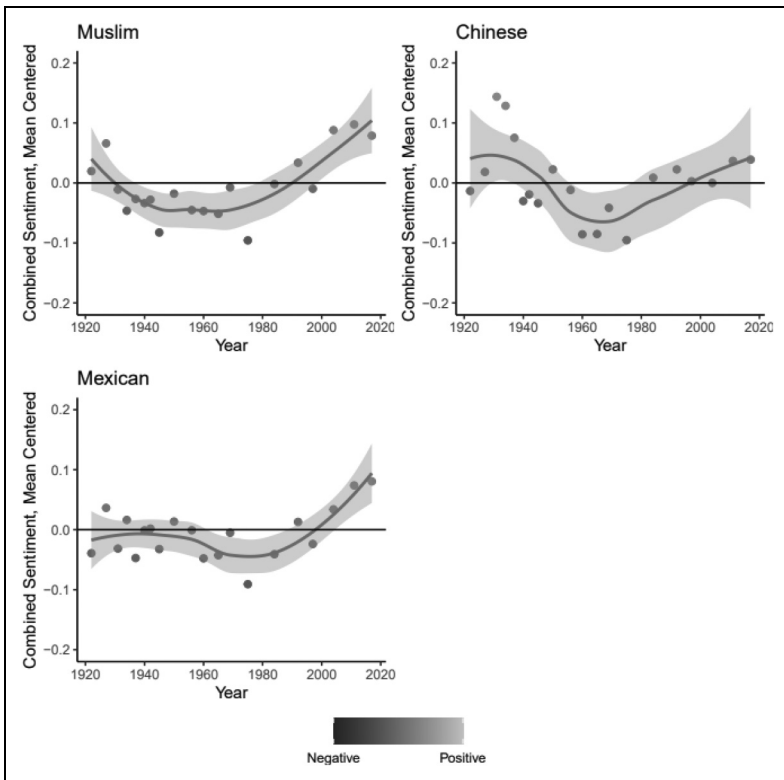


Figure 4. Sentiment: more or less positive relative to the edition for focal group set 3, by edition year. Higher scores are more positive. Figure includes a LOWESS plot of median scores and a bar representing a 95 percent confidence interval.

demonstrated various sentiment trajectories (see online Appendix D) and lend credibility to the conclusion that increased positive sentiment is not an artifact of the method.

Position Between “Strange” and “Normal”

A trend toward more positive sentiment in texts associated with the focal group concepts may suggest a decrease in the sense that the focal group is a “problem” for the mainstream. However, with mainstream shifts, we would expect that *Catholic*, *Chinese*, *Irish*, *Italian*, *Jewish*, *Mexican*, and

Muslim would not just have positive sentiment, they would also be seen as normal and ordinary – characteristics associated with the kind of Americans looking for advice on hosting their daughter’s quinceañera, for example. Our final analyses explore whether *Etiquette* shows these expansions of the boundary of the mainstream (see also Schachter 2016).

Our assumption is that incorporation into the mainstream would lead texts associated with the focal group concepts to be more “normal” over time. To test this assumption, we specify a semantic dimension anchored by a conceptual antonym pair: *strange* and *normal*. We then analyze the same salient pseudocorpora as in the sentiment analyses by establishing the position of each edition of the pseudocorpus within the semantic dimension. We continue with the convention of using italics when referring to *strange* and *normal* as concepts operationalized in CMDist. The premise of these analyses is that incorporation through mainstream shifts can be measured based on a focal group concept’s relative position, a proxy for meaningful association, between *strange* and *normal*. By assuming a continuum between this antonym pair (Fellbaum 1998), we can assess how the semantic association of texts that are salient to the focal group concepts move along that conceptual continuum over time (see also Kozlowski, Taddy, and Evans 2019).

Analyses are conducted using the CMDist antonym pair function (Taylor and Stoltz 2020). FastText embeddings are used to prepare a semantic continuum between *strange* and *normal*. CMDist calculates the focal groups’ pseudoeditions’ position between *strange* and *normal*, giving each pseudoedition a standardized score. These scores are used to show how the texts that are salient to each focal group concept changed in relation to being “normal” from the perspective of *Etiquette*’s depiction of the mainstream.

In specifying the strange to normal semantic dimension, we determine that mainstream inclusion should be seen as conceptually and definitionally independent of immigrant-ness (e.g., *American/Foreign*; *Immigrant/Citizen*). It is possible, and even likely, that a focal group concept becomes more or less associated with immigrant-ness over time, for example, as part of the process of immigrant incorporation or as a result of new waves of immigration from the same location. In theory, the relationship between these shifts in the immigrant-ness of the focal group concepts are independent of whether or not a focal group concept becomes more or less *normal*. Likewise, a focal group concept could maintain a similar level of association with immigrant-ness and become more *normal* through mainstream shifts that normalize immigration. We conducted multiple sensitivity analyses to buttress this claim, which can be found in online Appendix E. Analyses include reproduction of the results using alternative embeddings and alternative concept

specifications. Other specifications may be relevant with other data (e.g., news articles, see Stoltz and Taylor 2021).

Simple regressions with the CMDist *strange-to-normal* scores yielded significant coefficients across all seven groups ($p < .05$). Figure 5 presents results from an analysis that included all pseudocorpora. Results depict clear movement towards *normal*. The scores are standardized and mean-centered at 0 across all focal group concepts, reflecting overall trends.

Figure 6 plots the *strange-to-normal* CMDist scores for the first set of focal group concepts: *Italian, Jewish, Catholic, and Irish*. These scores are

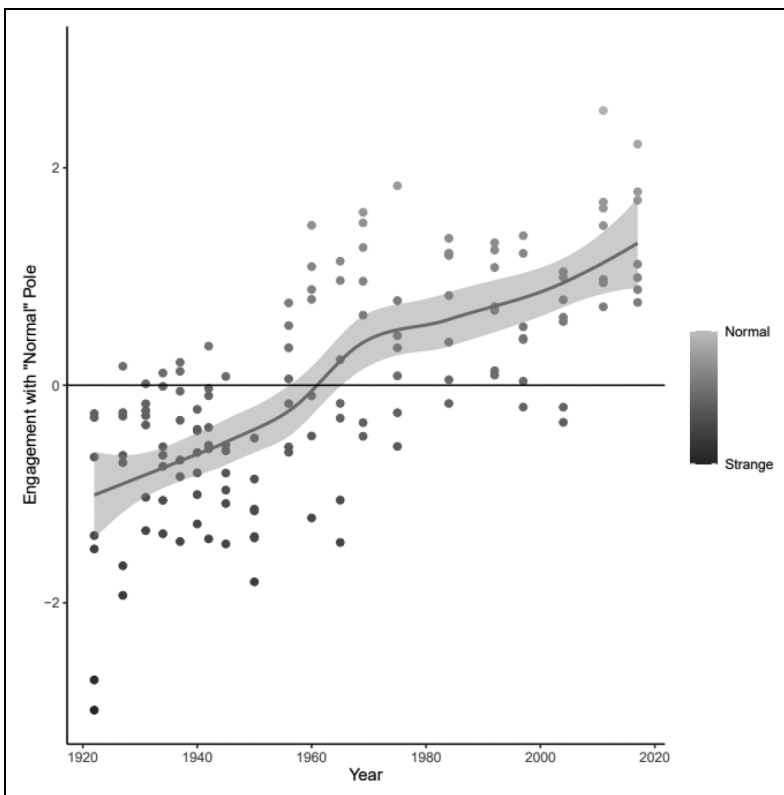


Figure 5. Meaning: normal versus strange over time for all focal group pseudocorpora, by edition year. Lower scores are closer to strange and higher scores are closer to normal. Figure includes a LOWESS plot of median scores and a bar representing a 95 percent confidence interval.

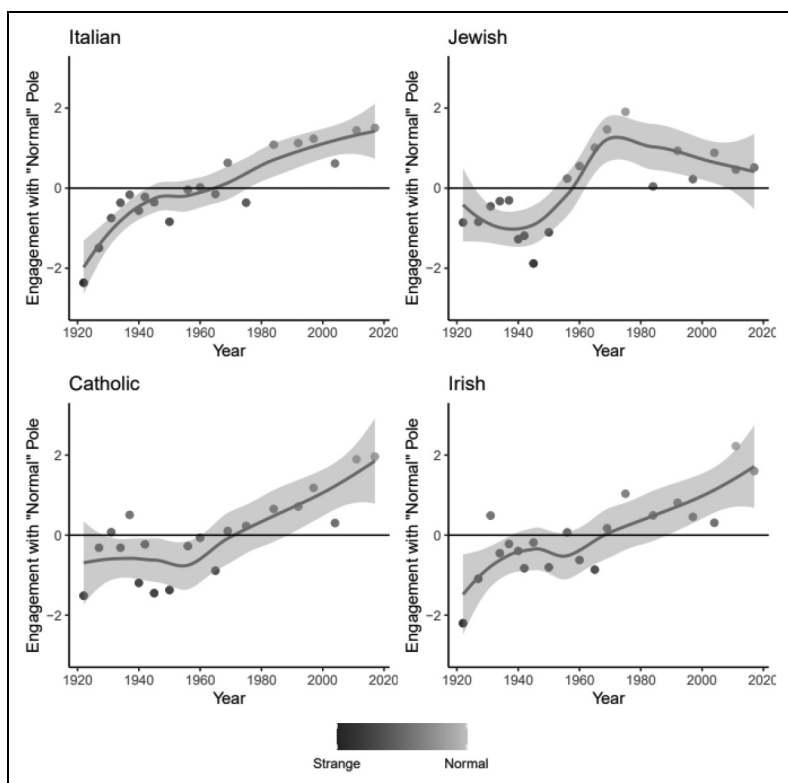


Figure 6. Meaning: normal versus strange over time for set I focal group pseudocorpora, by edition year. Lower scores are closer to strange and higher scores are closer to normal. Figure includes a LOWESS plot of median scores and a bar representing a 95 percent confidence interval.

mean-centered within the pseudocorpus for the individual focal group concept, with a score of 0 representing the average for the pseudoeditions salient to that focal group. As such, scores should not be compared between focal group concepts, but trajectories can be compared. A positive trend doesn't suggest incorporation so much as relative incorporation. Similarly, no positive trend doesn't suggest a lack of incorporation so much as a lack of change in incorporation. Figure 6 shows texts salient to *Catholic*, *Irish*, and *Italian* trend towards *normal*. Texts salient to *Jewish*, are less *normal*, beginning in the 1970s. This is the same period in which

there was more positive sentiment associated with the texts for *Jewish* and many other focal group concepts. This finding is taken up in the discussion.

Figure 7 plots the scores for *Muslim*, *Chinese*, and *Mexican*. This figure also demonstrates varied patterns. *Mexican* and *Muslim* have flatter trajectories toward *normal* relative to *Chinese* and most other focal groups. *Chinese* has a fairly clear trajectory toward *normal*. The confidence interval around *Muslim* is quite wide and remains close enough to 0, suggesting that *Muslim* may not have achieved a significant move towards *normal*.

The results from these analyses of the position of *Catholic*, *Chinese*, *Irish*, *Italian*, *Jewish*, *Mexican*, and *Muslim* between *strange* and *normal* show a

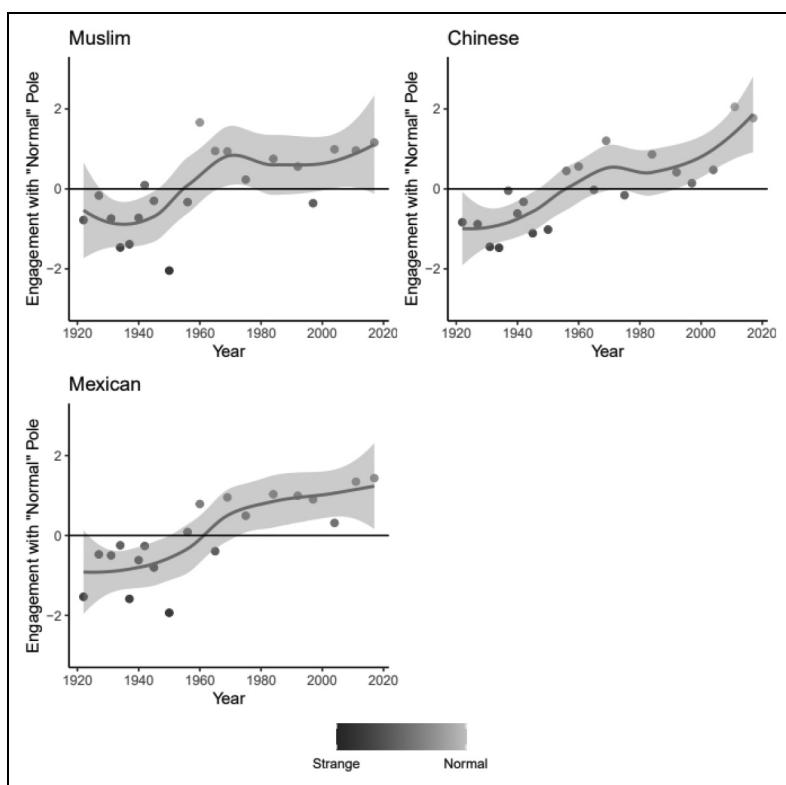


Figure 7. Meaning: normal versus strange over time for set 2 focal group pseudocorpora, by edition year. Lower scores are closer to strange and higher scores are closer to normal. Figure includes a LOWESS plot of median scores and a bar representing a 95 percent confidence interval.

general trend toward immigrant incorporation through shifts in the mainstream. However, particular focal group concepts follow different trajectories. Despite trends toward more positive sentiment, *Jewish*, *Mexican*, and *Muslim* are particularly noteworthy in their divergence from a solid upward trajectory from *strange* to *normal*. Although they remain closer to normal than in pre-1960, the texts associated with Jewish group concepts were less *normal* after the 1960s. Although there is not a clear decline for *Muslim*, there is a noticeable leveling of the trajectory. There is no similar situation observed for *Catholics*, which suggests this does not reflect mainstream shifts around religion. These findings are potential evidence of mainstream disaffiliation for *Jewish* and *Muslim*.

Discussion and Conclusions

The “remaking of the American mainstream” (Alba and Nee 2003) is crucial to the trajectory of immigrant incorporation. Past research on individuals’ attitudes, preferences, and practices has determined that established, non-immigrant, Americans become comfortable and familiar with the languages, foods, and cultural and religious practices that immigrant groups introduce. In some cases, established individuals develop a more expansive and less ethnocentric notion of what it means to be American. In this research, we sought to observe these mainstream shifts. Applying sociological computational text analysis methods to a unique text corpus, we examined mainstream shifts that show immigrant incorporation at the level of society’s social and symbolic organization. We examined mainstream shifts related to immigrants by analyzing texts focused on practical and mundane elements of everyday life. Our data consisted of Emily Post’s *Etiquette* – a book guiding Americans through high moments and everyday encounters for nearly a century. Taking *Etiquette* as a depiction of the mainstream, we examine the text for evidence of the incorporation of different groups.

Based upon exploratory analyses, we determined that our texts were best suited to observing mainstream shifts in relationship to long-standing groups who were and may still be seen as immigrants and outsiders in the United States: Catholic, Chinese, Irish, Italian, Jewish, Mexican, and Muslim. We theorized that this incorporation should be evident through positive sentiments attached to the people, practices, identifications, and objects associated with these groups. We theorized that incorporation would also include a background sense that these people, practices, identifications, and objects were not strange but normal elements of mainstream social life. To examine this, we used word embeddings to select text excerpts that were most salient to

these groups. We then analyzed the sentiment and the meaning of these texts in terms of position between being strange and normal.

We uncover mainstream shifts indicating some incorporation of all of these groups. We observed more positive sentiment and a movement from more *strange* to more *normal* on the aggregate, but there is variation within and across groups. Simple regressions showed these shifts were statistically significant ($p < .05$). However, as depicted in the figures, it often appears that the trajectory of incorporation over time is not linear. Instead, the analyses often demonstrate curves with increasing, decreasing, and stable incorporation periods. Furthermore, some focal groups acquire consistent positivity and a view of normalness over time, but others have less consistent trajectories of incorporation. Positive sentiment and normal-ness operated independently. In this article, we have not interpreted these trends in relation to the facts of migration history for each group, and in relation to periods of pro- and anti-immigrant attitudes and politics in the United States, but such analysis is a next logical step in the research. For example, in the case of *Chinese*, there was lagging positive sentiment but still a clear movement toward *normal*. Meanwhile, increased positive sentiment for *Jewish* and *Muslim* group concepts was coupled with declining or stagnating *normal*-ness, suggesting the possibility of rising disaffiliation instead of incorporation. Future research could follow up on these findings and examine the possible relationship between mainstream disaffiliation and related research that has observed Muslim American experiences being reclassified outside of whiteness after the 9/11 terrorist attacks (Maghbouleh 2017) and the rise of anti-Semitism in the contemporary United States (Moshin 2018).

Beyond immigrant integration, the techniques presented here could be adapted to a variety of questions related to social sentiment and symbolic inclusion. In particular, we emphasize the value of using a theoretically-selected text corpus instead of relying on a sampling logic in corpus selection. We also demonstrate the potential of computational sociological methods by moving beyond relying on explicit occurrences and focusing on implicit and explicit meanings through word embeddings and sentiment analyses, which are much more interpretive computational tools.

Finally, we want to underscore the continued centrality of qualitative epistemology in the practice of computational social science (Nelson 2021). Both human and machine reading and text analysis were crucial to this project - a mutual give and take combining the interpretive logics of the social scientists and machine neural networks of word embeddings in a computational grounded theoretic approach (Nelson 2020). Human analysis of the text

served two functions in our analyses: to confirm the validity of our embedding techniques and provide insight into how the mainstream shifts associated with immigrant integration actually played out in the texts. Computational approaches extend the interpretive toolkit, but human interpretation is crucial for correcting, confirming, and interpreting the results of the computational text analysis.

Acknowledgments

The authors wish to thank Lauretta Czarnezki, Alexander Karl, Asia Letlow, Pedro Pinales, Olivia Reid, and Stuart Riepl for their research assistance.

Author's Note

Voyer and Kline are equal first authors. Danton is second author. Volkova is third author. A replication package, including scripts for the analyses and appendices and instructions for accessing the Emily Post Corpus is available in the project repository, https://github.com/avoyer/Immigrants_and_Mainstream.



Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This article is based on research supported by the United States National Science Foundation Award No. 1851304, The University of Connecticut Office of the Vice President for Research, The University of Connecticut College of Liberal Arts and Sciences, and the University of Connecticut Department of Sociology.

ORCID iDs

Zachary D. Kline  <https://orcid.org/0000-0003-4942-929X>
Andrea Voyer  <https://orcid.org/0000-0003-0338-3273>

Supplemental Material

Supplemental material for this article is available online.

Notes

1. In autumn 2022, after this research was completed, a new edition was released.
2. Exploratory sentiment analysis was conducted using the language analysis software, Linguistic Inquiry Word Count (LIWC 2015; Pennebaker et al. 2015).

- LIWC is a well-established tool that considers the psychological properties of words based on their typical contexts of use. The exploratory sentiment analysis did not use a period-specific sentiment lexicon.
3. Positive sentiment for the text around quinceañeras in 1997 was 3.78. In 2011, it was 5.08, suggesting increased positive sentiment. The positive sentiment for texts around sweet sixteen in 1997 was 6.50.
 4. CMDist tracks how varying documents engage with a *concept* that is identified using the semantic meanings in word embeddings as a map, rather than tracking the changing meaning of a *word* over time in different word embeddings (i.e., Words Movers Distance; e.g., Hamilton et al. (2016). To verify that the results are not sensitive to the embeddings used, we recreated all of the main analyses of the paper using 5 different word embeddings across all analyses. These results, as well as a more detailed explanation of the underlying logic of selecting an embedding, is discussed in online Appendix D.
 5. In cases where none of the chunks are plausibly related to the concept and the researchers have not observed the presence of that concept in the text through reading, there is likely not sufficient validity for the analysis of that concept to move forward. If the researchers have observed the presence of the concept through reading, the concept specification should be adjusted and the pretrained embedding should be switched to see if better results are obtained before abandoning the concept.
 6. In cases where there are no unique terms in a pseudocorpus but reading confirmed the presence of the concept in the text and the selected chunks were deemed plausible in their relationship to the concept, it should be considered whether the lack of uniqueness is indeed the result of a lack of unique chunks, if this overlap is driven by similarity to just one other focal concept, and whether concepts that are too closely related in the embeddings could be respecified to limit overlap. If it is not possible to locate the concept in the text, select chunks that are plausibly related, or develop a pseudocorpus that is unique to the concept, the analysis of that concept should not move forward.

References

- Abrutyn, Seth and Michael J. Carter. 2015. "The Decline in Shared Collective Conscience as Found in the Shifting Norms and Values of Etiquette Manuals." *Journal for the Theory of Social Behaviour* 45(3):352-76.
- Alba, Richard D. 2005. "Bright vs. Blurred Boundaries: Second-Generation Assimilation and Exclusion in France, Germany, and the United States." *Ethnic and Racial Studies* 28:20-49.
- Alba, Richard D. and Victor Nee. 2003. *Remaking the American Mainstream: Assimilation and Contemporary Immigration*. Cambridge, MA: Harvard University Press.

- Arditi, Jorge. 1999. "Etiquette Books, Discourse and the Deployment of an Order of Things." *Theory, Culture & Society* 16(4):25-48.
- Bojanowski, Piotr, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. "Enriching Word Vectors with Subword Information." *Transactions of the Association for Computational Linguistics* 5:135-46.
- Bourdieu, Pierre. 1984. *Distinction: A Social Critique of the Judgement of Taste*. Cambridge, MA: Harvard University Press.
- Brodtkin, Karen. 1998. *How Jews Became White Folks and What That Says About Race in America*. New Brunswick, N.J.: Rutgers University Press.
- Brown, Wallace. 1970. "The View at Two Hundred Years: The Loyalists of the American Revolution." *Proceedings of the American Antiquarian Society* 80(1):25-47.
- Carbone, Luca and Jonathan Mijs. 2022. "Sounds Like Meritocracy to My Ears: Exploring the Link Between Inequality in Popular Music and Personal Culture." *Information, Communication, and Society* 25(5):707-25.
- Claridge, Laura. 2009. *Emily Post: Daughter of the Gilded Age, Mistress of American Manners*. New York, NY: Random House.
- Drouhot, Lucas G. and Victor Nee. 2019. "Assimilation and the Second Generation in Europe and America: Blending and Segregating Social Dynamics Between Immigrants and Natives." *Annual Review of Sociology* 45:177-99.
- Dumenil, Lynn. 1991. "The Tribal Twenties: 'Assimilated' Catholics' Response to Anti-Catholicism in the 1920s." *Journal of American Ethnic History* 11(1):21-49.
- Elias, Norbert. 2000. *The Civilizing Process: Sociogenetic and Psychogenetic Investigations*. Oxford, England: Blackwell Publishers.
- Ellis, Nick C. 2019. "Essentials of a Theory of Language Cognition." *The Modern Language Journal* 103:39-60.
- Fellbaum, Christiane. 1998 "A Semantic Network of English: The Mother of all WordNets." Pp. 137-48 in *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Dordrecht: Springer.
- Firth, John Rupert. 1957 "A Synopsis of Linguistic Theory, 1930-1955." Pp. 168-205 in *Studies in Linguistic Analysis*, edited by J. R. Firth. Oxford, England: Blackwell.
- Garvin, Paul L. 1962. "Computer Participation in Linguistic Research." *Language* 38(4):385-9.
- Giuliani, Luigi, Herman Brinkman, Geert Lernout, and Marita Mathijssen, edited by 2006. *Texts in Multiple Versions: Histories of Editions (Vol. 5)*. Amsterdam, Netherlands: Rodopi.
- Goffman, Erving. 1963. *Behavior in Public Places: Notes on the Social Organization of Gatherings*. New York, NY: Free Press of Glencoe.
- Hamilton, William L., Kevin Clark, Jure Leskovec, and Dan Jurafsky. 2016. "Inducing Domain-Specific Sentiment Lexicons from Unlabeled Corpora."

- arxiv Cornell University. Retrieved July 15, 2022 (<https://arxiv.org/abs/1606.02820>).
- Ignatiev, Noel. 2009. *How the Irish Became White*. New York: Routledge.
- Jacobs, Laura. 2001. "Emily Post's Social Revolution." *Vanity Fair*, December 2001. Retrieved August 19, 2020 (<https://archive.vanityfair.com/article/2001/12/emily-posts-social-revolution>).
- Jiménez, Tomás. 2017. *The Other side of Assimilation: How Immigrants are Changing American Life*. Berkeley: University of California Press.
- Jiménez, Tomás R. and Adam L. Horowitz. 2013. "When White is Just Alright: How Immigrants Redefine Achievement and Reconfigure the Ethnoracial Hierarchy." *American Sociological Review* 78(5):849-71.
- Joulin, Armand, Edouard Grave, Piotr Bojanowski, Matthijs Douze, Herve Jégou, and Tomas Mikolov. 2016. "FastText.zip: Compressing Text Classification Models." arxiv Cornell University. Retrieved August 19, 2020 (<https://arxiv.org/abs/1612.03651>).
- Kozlowski, Austin C., Matt Taddy, and James A. Evans. 2019. "The Geometry of Culture: Analyzing the Meanings of Class Through Word Embeddings." *American Sociological Review* 84(5):905-49.
- Kroneberg, Clemens and Andreas Wimmer. 2012. "Struggling Over the Boundaries of Belonging: A Formal Model of Nation Building, Ethnic Closure, and Populism." *American Journal of Sociology* 118:176-230.
- Lee, Erika. 2003. *At America's Gates: Chinese Immigration During the Exclusion Era, 1882–1943*. Chapel Hill: University of North Carolina Press.
- Lees-Maffei, Grace. 2012 "Small Change? Emily Post's Etiquette (1922–2011)." Pp. 217-48 in *Must Read: Rediscovering American Bestsellers*, edited by S. Churchwell and T. R. Smith. New York, NY: Continuum International Publishing Group.
- Lenci, Alessandro. 2018. "Distributional Models of Word Meaning." *Annual Review of Linguistics* 4:151-71.
- Maghbouleh, Neda. 2017. *The Limits of Whiteness: Iranian Americans and the Everyday Politics of Race*. Palo Alto: Stanford University Press.
- Massey, Douglas S., Jorge Durand, and Nolan J. Malone. 2002. *Beyond Smoke and Mirrors: Mexican Immigration in an Era of Economic Integration*. New York: Russell Sage Foundation.
- Massey, Douglas S. and Magaly Sánchez. 2010. *Brokered Boundaries: Creating Immigrant Identity in Anti-Immigrant Times*. New York: Russell Sage Foundation.
- McHugh, Jess. 2021. *Americanon: An Unexpected U.S. History in 13 Bestselling Books*. New York: Dutton.
- Mennell, Stephen. 2007. *The American Civilizing Process*. Cambridge, England: Polity.
- Moshin, Jamie. 2018. "Hello Darkness: Antisemitism and Rhetorical Silence in the 'Trump Era.'" *Journal of Contemporary Rhetoric* 8(1/2):26-43.

- Nelson, Laura K. 2020. "Computational Grounded Theory: A Methodological Framework." *Sociological Methods & Research* 49(1):3-42.
- Nelson, Laura K. 2021. "Leveraging the Alignment Between Machine Learning and Intersectionality: Using Word Embeddings to Measure Intersectional Experiences of the Nineteenth Century U.S. South." *Poetics* 88(21). <https://doi.org/10.1016/j.poetic.2021.101539>.
- Pennebaker, James, Roger Booth, Ryan Boyd, and Martha Francis. 2015. *Linguistic Inquiry and Word Count: LIWC2015*.
- Prabowo, Rudy and Mike Thelwall. 2009. "Sentiment Analysis: A Combined Approach." *Journal of Informetrics* 3(2):143-57.
- Portes, Alejandro and Robert L. Bach. 1985. *Latin Journey: Cuban and Mexican Immigrants in the United States*. Berkeley: University of California Press.
- Portes, Alejandro and Ruben G. Rumbaut. 2001. *Legacies: The Story of the Immigrant Second Generation*. Berkeley: University of California Press.
- Portes, Alejandro and Min Zhou. 1993. "The New Second Generation: Segmented Assimilation and its Variants." *The ANNALS of the American Academy of Political and Social Science* 530:22.
- Post, Elizabeth L. 1975. *The New Emily Post's Etiquette*. New York, NY: Funk & Wagnalls Company.
- Post, Emily. 1937. *Etiquette: The Blue Book of Social Usage*. 5th ed. New York, NY: Funk & Wagnalls Company.
- Post, Peggy. 1997. *Emily Post's Etiquette*. 16th ed. New York: HarperCollins Publishers.
- Post, Peggy. 2004. *Emily Post's Etiquette*. 17th ed. New York: HarperCollins Publishers.
- Post, Peggy, Anna Post, Lizzie Post, and Daniel Post Senning. 2011. *Emily Post's Etiquette: Manners for a New World*. New York: HarperCollins Publishers.
- Roediger, David R. 2005. *Working Toward Whiteness: How America's Immigrants Became White: The Strange Journey from Ellis Island to the Suburbs*. New York: Basic Books.
- Schachter, Ariela. 2016. "From 'Different' to 'Similar' an Experimental Approach to Understanding Assimilation." *American Sociological Review* 81(5):981-1013.
- Stoltz, Dustin S. and Marshall A. Taylor. 2019. "Concept Mover's Distance: Measuring Concept Engagement via Word Embeddings in Text." *Journal of Computational Social Science* 2:293-313.
- Stoltz, Dustin S. and Marshall A. Taylor. 2021. "Cultural Cartography with Word Embeddings." *Poetics* 88(21):231-242. <https://doi.org/10.1016/j.poetic.2021.101567>.
- Takaki, Ronald T. 1993. *A Different Mirror: A History of Multicultural America*. Boston: Little, Brown & Co.

- Taylor, Marshall A. and Dustin S. Stoltz. 2020. "Integrating Semantic Directions with Concept Mover's Distance to Measure Binary Concept Engagement." *Journal of Computational Social Science* 4:231-42.
- Timm, Jason. 2019. "Historical Word Embeddings and Lexical Semantic Change." *Github*. Retrieved April 4th, 2022 (<https://github.com/jaytimm/google-ngrams-and-r>).
- Voyer, Andrea. 2013a. "Notes on a Cultural Sociology of Immigrant Incorporation." *American Journal of Cultural Sociology* 1(1):26-41.
- Voyer, Andrea M. 2013b. *Strangers and Neighbors: Multiculturalism, Conflict, and Community in America*. New York: Cambridge University Press.
- Voyer, Andrea, Zachary D. Kline, and Madison Danton. 2022. "Symbols of class: A computational analysis of class distinction-making through etiquette, 1922–2017." *Poetics* <https://doi.org/10.1016/j.poetic.2022.101734>.
- Wimmer, A. 2008. "The Making and Unmaking of Ethnic Boundaries: A Multilevel Process Theory." *American Journal of Sociology* 113:970-1022.
- Wimmer, Andreas. 2009. "Herder's Heritage and the Boundary-Making Approach in Immigrant Societies." *Sociological Theory* 27:244-70.
- Wouters, Cas. 2007. *Informalization: Manners and Emotions Since 1890*. Los Angeles, CA: Sage Publications.

Author Biography

Andrea Voyer is an Associate Professor (docent and senior lecturer) in the Department of Sociology at Stockholm University. Her current research uses etiquette books as data in computational sociological and qualitative analyses of social inclusion and exclusion on the basis of immigration, race and class.

Zachary D. Kline is a PhD Candidate at the University of Connecticut and Teacher-Scholar Fellow at The College of New Jersey. Zachary is a stratification scholar whose substantive interests intersect with public policy, economic sociology, and urban sociology. Methodologically, he employs quantitative and computational techniques. His work has appeared in *Social Forces*, *The Sociological Quarterly*, *Poetics* and *Socius: Sociological Research for a Dynamic World*.

Madison Danton is a doctoral student in the Department of Sociology at the University of Connecticut.

Tatiana Volkova is a graduate student in the Department of Computer and System Sciences at Stockholm University.