A cultural shift in Western perceptions of Palestine

Terry Regier
Department of Linguistics
UC Berkeley
terry.regier@berkeley.edu

Muhammad Ali Khalidi
Department of Philosophy
CUNY Graduate Center
makhalidi@gc.cuny.edu

Abstract

We argue that a cultural shift in Western perceptions of Palestine began in the late 1990s to 2000s, leading to increased openness to Palestinian perspectives, including awareness of the Nakba. We present 3 computational analyses designed to test this idea against data from the 2020 Google Books English dataset. The results support the claim of a cultural shift, and help to characterize that shift.

Introduction

In the West today, there is some awareness of the Nakba — the displacement and dispossession through violence of Palestinian Arab society in 1948 (Zureiq, 1948). This awareness is evident from references to the Nakba in discussion of the current war in Gaza, and from the fact that a major computational linguistics conference is hosting a workshop on Nakba narratives. But not long ago, things were different. In her book *Perceptions of Palestine*, written from a U.S. perspective, Christison (1999, p. 2) remarked:

> The dispossession and dispersal of the Palestinians in 1948 has always been and to a great extent remains "an unrecognizable episode," as [Malcolm] Kerr put it, even for most informed Americans ... — unrecognizable in the sense not only that the dispossession has been forgotten but also that it is seldom recognized to be the ultimate cause of the conflict.

From today's perspective, these remarks seem dated, and it is striking that the author does not use the word *Nakba* in a direct reference to that historical event. In contrast, several books published in English since then have the word *Nakba* (Arabic for *catastrophe*) in their title (e.g. Sa'di and Abu-Lughod, 2007; Masalha, 2012; Abdo and Masalha, 2018; Allan, 2021). Thus it appears that the Nakba has become more prominent in the West, both as a recognized historical event and as a word, than it was not too long ago. Why is this?

We pursue the hypothesis that a cultural shift occurred in the late 1990s to 2000s, resulting in greater Western openness to Palestinian perspectives — including but not limited to awareness of the Nakba. On this hypothesis, the shift emerged in opposition to a dominant view that is indifferent or hostile to Palestinian perspectives, and that views Palestine and Palestinians from the outside, as a problem, and in terms of the so-called "conflict". The collision of these two views can be seen in current tensions in the West surrounding the Palestine question; our hypothesis concerns the origins of the more Palestine-sympathetic view.

Our study connects to prior computational work that has explored linguistic reflections of dehumanization (Mendelsohn et al., 2020; Caporusso et al., 2024). We emphasize in particular the feature of negative evaluation, which has been engaged in this prior work, and also the notions of psychological distancing and denial of subjectivity that have been identified as features of dehumanization (Opotow, 1990; Haslam, 2006) but have received less attention in computational work. A related body of work explores bias and propaganda in news media (e.g. Hamborg et al., 2019), and a line of this work has specifically addressed representations of the Palestine question in the news, in the context of the current war on Gaza (e.g. Zaghouani et al., 2024) and the 2014 Gaza war (Al-Sarraj and Lubbad, 2018). With respect to a cultural shift in Western perceptions of Palestine, Telhami (2018, 2020) and Serhan (2023) argued for a recent shift in the U.S., based on polling data, and Regier (2016) presented evidence that the term *the Nakba* entered the English language around 1998, and proposed that this might have signalled the beginning of a general shift. Regier and Khalidi (2024) provided an initial test of that proposal, based on historical

Proceedings of the Proceedings of the 1st International Workshop on Nakba Narratives as Language Resources, pages 9–17
January 20, 2025.

patterns of language use, suggesting support for it. However they did so briefly, in a few paragraphs in a non-peer-reviewed publication, with technical detail limited to 2 footnotes. Our study tests the same hypothesis in detail; the specific analyses we report here are novel.

In what follows, we first describe the data on which we rely, and then test the hypothesis of a general cultural shift in a series of 3 studies of increasingly fine temporal grain, asking: (**Study 1**) Was there a general shift?; (**Study 2**) If so, was the shift historically unusual?; (**Study 3**) Is such a shift supported by the first appearance of specific English words in a Palestinian context? We then conclude, and consider limitations of this work.

Data

We drew on data from the 2-gram subset of the Google Books English corpus (Michel et al., 2011), Version 20200217,[1] which we took to be reasonably representative of elite Western language use. We are aware of concerns surrounding this dataset, in particular relating to its change in composition over time, for example due to changes in the degree of representation of scientific work and fiction (e.g. Pechenick et al., 2015; Schmidt et al., 2021). In an attempt to address these concerns, we restricted attention to a subset of the overall 2-gram dataset: those bigrams in which the first term is a national adjective (explicitly marked as an adjective with the suffix *_ADJ*), and the second term is any noun (marked with the suffix *_NOUN*) composed of lower-case characters, which excludes proper names.[2] This yielded a dataset of bigrams such as *Korean students*, *Nigerian government*, *Israeli delegation*, *Palestinian territories*, etc., holding frequency counts for each bigram for each year. We also restricted the years under consideration to 1948 (the year of the Nakba) through 2019 (the last year covered by the dataset). We refer to this restricted subset as the *bigram dataset*. The restriction to (national adjective, noun) pairs is intended to reduce any influence of a change in composition of the larger dataset over time: while an influx of scientific publications or fiction could skew the overall Google Books dataset by increasing the frequency of certain terms, it seems less likely that it would substantially skew a subset of that dataset

[1] http://storage.googleapis.com/books/ngrams/books/20200217/eng/eng-2-ngrams_exports.html.

[2] When listing words here, we omit the *_ADJ* and *_NOUN* suffixes for readability.

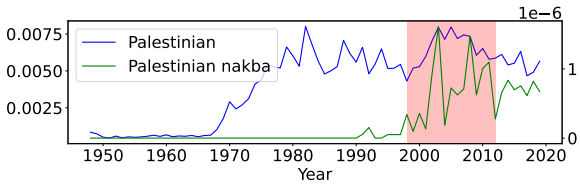

Figure 1: Relative frequency over time for *Palestinian* (left y-axis) and *Palestinian nakba* (right y-axis), in the subset of the Google Ngram dataset that we use: the "bigram dataset". The time period in pink is 1998-2012.

that is composed only of bigrams of this character. When we calculate the relative frequency of a target bigram, we normalize using only the frequencies of bigrams in this bigram dataset. Figure 1 shows such relative frequency traces over time for the term *Palestinian* (i.e. *Palestinian* * where * is any noun), and for *Palestinian nakba*. The time period highlighted in pink (1998-2012) shows an increase in usage for *Palestinian*, and a sharp increase from near zero for *Palestinian nakba*. On this basis, we take 1998-2012 to be a target period of apparently greater attention to Palestinians, and we test the hypothesis of a cultural change during that period.

Study 1: A general cultural shift?

If there was a general cultural shift, then there should be expressions other than *the Nakba* or *Palestinian nakba* that: (1) increased in frequency during the target period, (2) are still in use, and (3) convey openness to Palestinian perspectives. To identify those words that increased in frequency in a Palestinian context during the target period, and that remained high-frequency afterwards, we extracted from the bigram dataset those nouns n that appeared in the bigram "Palestinian n" disproportionately more often after the target period than before it. The target period is 1998-2012, so we defined two periods on either side of it: the distant past, which we call "then", defined as 1948-1997, and the recent past, which we call "now", defined as 2013-2019. We calculated, for each noun n that appears in a bigram of the form "Palestinian n", the log of the likelihood ratio:

$$G(n, \text{now}) = \log_2 \frac{p(\text{Palestinian } n|\text{now})}{p(\text{Palestinian } n|\text{then})} \quad (1)$$

where the two conditional probabilities are estimated as relative frequencies based on corpus counts in the bigram dataset, and with normalization also based on counts in the bigram dataset,

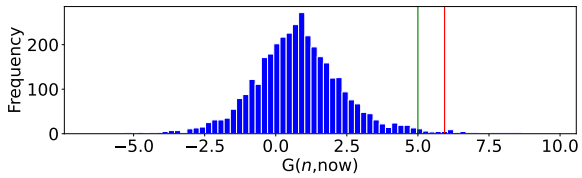

Figure 2: The histogram shows $G(n, \text{now})$ for all nouns n that are modified by the national adjective *Palestinian* in the bigram dataset. The red line marks this quantity for the noun *nakba*, and the green line marks $G(n, \text{now})$ = 5, which means that the bigram probability for "Palestinian n" is 2^5 = 32 times greater in the recent than in the distant past. Of the full set of 3698 nouns, 64, or 1.7%, are at or above this value.

with add-one smoothing applied to all corpus counts. $G(n, \text{now})$ measures the extent to which a given noun n was more frequent in a Palestinian context in the recent than in the distant past, and Tenenbaum and Griffiths (2001) argue that this quantity, the log of the likelihood ratio, captures the cognitive notion of being a good example of a category — in this case, the category of the recent rather than the distant past. Thus, nouns with high values for $G(n, \text{now})$ should be good examples (hence the letter G), or prototypical, of text from the recent rather than the distant past.[3]

The results of this analysis are shown in Figure 2. It can be seen that *nakba* (red line) is an extreme example of a word that showed increased relative frequency in a Palestinian context during the recent past compared with the distant past; thus, the bigram *Palestinian nakba* is identified as prototypical of the recent past, among bigrams starting with *Palestinian*. At the same time, *nakba* is not the only such extreme example. The green line in the figure marks a value that is lower than that for *nakba*, but high enough that only a small fraction of all nouns is at or above this value (see caption). The 64 words in this upper tail are listed in the top panel of Table 1, in order of decreasing $G(n, \text{now})$, i.e. decreasing prototypicality with respect to the recent past. A number of these newly prominent words capture a view of Palestinians from "up close," without much psychological distancing. This can be seen in words (other than *nakba*) that explicitly adopt Palestinian perspectives, as one would expect if there had been a general cultural shift, e.g. *sumud* (Arabic for steadfastness), *keffiyeh, intifadas*, words that highlight Palestinian subjectivity, e.g. *subjectivity, testimonies, narratives*, and words that capture ordinary human interests e.g. *football, restaurants, filmmakers*. Another theme represented here is that of the internet, e.g. *internet, website, hackers*; this is significant because it has been suggested (Regier, 2016; Regier and Khalidi, 2024) that the advent of the internet may have facilitated a cultural shift in Western perceptions of Palestine, by opening up access to information from a wide range of sources that previously would have been difficult to find. Overall, there seems to be a tendency for the newly prominent words to capture Palestinian perspectives from "up close", i.e. with minimal psychological distance, and thus in a way that emphasizes Palestinian humanity, i.e. the opposite of dehumanization.[4]

The bottom panel of the same table shows the 64 nouns n that appear in the highest frequency bigrams of the form "Palestinian n" for the same period, i.e. 2013-2019, listed in order of descending frequency. This is conceptually distinct from the words with greatest prototypicality for that period just seen in the top panel, and the contents of the high-frequency list contrast with the prototypical words. The high-frequency list contains many words that capture the dominant view of Palestinians "from the outside", i.e. with greater psychological distance – and in fact as a *problem*, which word is included in the list, along with words that, while not necessarily expressing psychological distance, do not express closeness, e.g. *state, people, conflict, issue, refugees*. There is no overlap between the two lists, although *intifadas* (note the plural, only applicable since 2000) appears in the first list, and *intifada* appears in the second. This lack of overlap indicates that the newly prominent words are not extremely prominent in absolute terms. There are some apparent exceptions to the general tendencies noted above — e.g. *intifada* among the high-frequency words, and *corruption* among the newly prominent words — but the general tendencies themselves seem at least qualitatively apparent.

We take these findings to support the view that there was a recent general cultural shift in Western

[3]We used this method rather than that of Monroe et al. (2008) because this method is simpler and has an independent connection to the cognitive notion of being a good example.

[4]As a control, we re-ran the same analysis but on nouns following the national adjective *American* (cf. Mendelsohn et al., 2020). The American list of the top 64 nouns contained many not found on the Palestinian list, e.g. *millennials, subprime, cybersecurity, cisgender, megachurches*. The lists for the two national adjectives contained the following nouns in common: *jihadist, jihadists, jihadi, website, websites, internet, indigeneity*; thus, these words highlight shared themes.

Newly prominent: hip queers intifadas hackers indigeneity gays masculinity jihadist internet rappers bid victimhood modernity spaces stakeholders accession cartoon rapper commemoration corruption jihadists website space rap patriation cleric football telecommunications custodianship filmmaking reform mobility testimonies restaurants shahid sumud nakba polling subjectivity textbook cinemas spouses practitioners coordinator websites keffiyeh imam presidency facilitators classrooms airspace narratives livelihoods campaigners filmmakers jihadi print standoff textbooks imams colonial cinema reforms statebuilding
High frequency: state people conflict refugees territories cause population women refugee society citizens issue resistance leadership territory land question nationalism children problem identity side groups rights leaders community struggle terrorists peace villages leader civilians economy suicide self statehood prisoners uprising communities factions politics residents organizations security history village movement youth woman delegation students areas case government workers militants terrorism lands minority intifada police cities context families

Table 1: **Top**: Nouns that were newly prominent in a Palestinian context as of 2013-19, and thus prototypical of that period rather than the distant past, listed by decreasing prototypicality. **Bottom**: Nouns that were high-frequency in a Palestinian context during the same period, 2013-19.

(specifically Anglophone) perceptions of Palestine, exemplified by newly prominent words that often convey openness to Palestinian perspectives, and psychological closeness to Palestinian experience, in contrast with high-frequency words from the same period, which do not show these features as clearly. This pattern suggests that the cultural shift was real but not especially high-profile at the time. Instead, it can be characterized as a quiet precursor to today's prominence for some of the same ideas, in the context of a continuing dominant view that is not particularly receptive to Palestinian perspectives.

Study 2: Was the shift historically unusual?

For a cultural shift to be noteworthy, it should be unusual — specifically it should be the case that the kinds of changes observed during that period are not observed regularly during earlier historical periods. To test whether this is the case, we ran variants of the analysis from Study 1 above, which picked out words that were newly prominent in a period of interest, but now for systematically varying target periods. Specifically, we conducted such analyses for several target decades: the 1960s, 1970s, 1980s, 1990s, 2000s, and 2010s. For each decade, the past or "then" period lasted from 1948 up until the last year before the target decade, and the "now" period covered the years of that decade.[5] A difference from the analysis in Study 1 is that here we compare an earlier time period with the decade immediately following it — whereas in the original analysis we compared time periods on either side of a specified target period. We adopt the form of analysis we do here because the original analysis spanned a large time range, and we wished to move to a temporally finer grain, to pinpoint more specifically when changes happened.

For each target decade, and for each *Palestinian*-modified noun n in that decade, we obtained the quantity $G(n, \text{now})$ as defined above, with the "now" period being the target decade and the "then" period being the period leading up to it, as just described. This is a measure of the extent to which the bigram "Palestinian *n*" is more frequent in the target decade, relative to the preceding period. Table 2 shows, for each decade, the top 50 nouns that were found to be newly prominent in a Palestinian context for that decade, in order of decreasing $G(n, \text{now})$.[6] Especially in earlier decades, there are several words relating to Judaism, e.g. *talmud, yeshiva, targum, gemara*. Many of the other words align with what one might expect given the historical record of the Palestinian national movement: *revolution, liberation, commando, guerillas, resistance* in the 1960s, when the PLO was first formed and began to assume prominence; *hijackings, hijackers, skyjackers* in the 1970s; *intifada, uprising, protestors* in the 1980s, at the time of the First Intifada; *spokeswoman, tracks, ministries, governance* in the 1990s, at the time the Madrid and Oslo processes. The 2000s are of particular interest because our target period (1998-2012) falls mostly within this decade. Here, the newly-prominent Palestinian words show a conflicting mix of perspectives, perhaps reflecting change in progress. On the one hand, we see the unfortunately familiar theme of violence, presumably because of

[5] For example, for the 1960s, "then" = 1948-1959 and "now" = 1960-1969; for the 1970s, "then" = 1948-1969 and "now" = 1970-1979, etc.

[6] All words listed in this table had $G(n, \text{now})$ values well above zero, i.e. all showed a substantial increase in frequency during the target decade, relative to the preceding period. These words include a few misspellings and partial words, presumably reflecting OCR errors: *lssue* (with a lower-case L, rather than issue), *confl* (presumably part of conflict).

1960-1969: revolution liberation commando guerillas resistance guerilla organisations laissez action freedom exodus fighter fedayin entity intellectuals guerrilla partisans commandos kerygma consciousness level schoolgirls commandoes guerrillas sign diaspora ampulla bourgeoisie patriots congress oak talmud yeshiva activist personality struggle organizations individual identity demands graduates targum saboteurs concept revolutionaries potential brigades fragment males grievances
1970-1979: objectives news camps splinter attempt mini targets bases rejectionists voice diplomacy dimension demonstrations acts hijackers component approach attempts lssue statelet desire quarter reconciliation autonomy selfrule coexistence militiamen leaderships recognition operation dialogue presence resolutions hijackings perceptions role peoplehood dissidents involvement bomb gemara summit participation reporter clashes concentration murder skyjackers spokesmen distinctiveness
1980-1989: dies cartoonist protester intifadah uprising businesses constants taxi folktale fida'iyyin boycott neighborhoods lssues communication reunification scarves anthem marketing collaborator rioting separatists infighting departure unionists intifada evacuation shells infrastructure passenger analyst intifadeh decisionmaking mujahidin rioters protesters nonviolence entrepreneur missile decisionmakers interdependence savings genocide boyfriend agenda assaults cars childhood step sample fringe
1990-1999: tourism track spokeswoman tracks subcontractors debts preventive adm ministries residency donor governance knife television islands bantustans selfgovernance hanging empowerment groundwater police grassroots jails websites management voter trucks worshipers businesspeople polls patriarchy capacities mufti nakba histories lntifada engagement paramilitaries pork airport negotiator demilitarization publics investment custody lesbians plates democratization stateless expenditure
2000-2009: hackers intifadas jihadists mortar moms reform rappers jihadist imams gays patriation modernity corruption reforms shahid queers hip website landholder masculinity airspace internet sniper textbook firebrand suicide textbooks cartoon spouses cinemas custodianship homicide schoolbooks victimhood pollster presidency apologists facilitators contiguity cleric coordinator bombers classrooms medics schoolboys nakba rapper reformers premier descendants
2010-2019: indigeneity accession bid hip internet kitchen queers presidents dessert football rap push application hagiography plaintiffs conict rapper websites rabbinic keffiyeh subjectivity campuses museum operative stakeholders spaces masculinity space collective victimhood schism commemoration restaurants food claimants mobility practitioners filmmaking st confl meals web cartoon campaigners telecommunications cuisine interviewees applications foods rappers

Table 2: Nouns that were newly prominent in a Palestinian context in several target decades, and thus prototypical of that decade rather than the preceding past, listed by decreasing prototypicality for each decade.

the events of the Second Intifada (*sniper, suicide, bombers*). On the other hand, there are a number of words that suggest an awareness of Palestinian perspectives and concerns; these include *nakba*, and also *contiguity*, and arguably *moms, medics, cinemas*. Also of note are e.g. *website, internet, hackers*, which, as noted above, likely reflect the advent of the internet, a possible mechanism for the proposed cultural shift. In contrast, the 2010s appear to be a period in which newly prominent words more consistently reflect Palestinian perspectives and experiences. Of note, 7 of the top 50 newly prominent words in this decade, or 14%, concern food: *kitchen, dessert, restaurants, food, meals, cuisine, foods*. There are very few food-related words in other decades (but see *pork* in the 1990s) — this development, along with other words from this decade, suggests a move away from viewing Palestinians as a political problem, and toward viewing them "up close" as normal human beings with normal interests such as food, and suggests more generally a new openness to and interest in Palestinian culture.

We supplemented this qualitative analysis with a quantitative one. Specifically, we assessed the valence, positive or negative, of each noun (without the adjective *Palestinian*) via sentiment analysis.[7] We used this to classify each of the words in Table 2 above, and noted what proportion of the top 50 words per decade had positive sentiment. The results are shown in Figure 3; Fisher's exact test conducted on the counts underlying those proportions indicated a significant association ($p < 0.025$, 2-tailed) between positive sentiment and the 2010s, compared with the 1960s-2000s pooled together.

[7]We considered several sentiment analysis methods and eventually settled on DistilBERT base uncased finetuned SST-2, based on DistilBERT (Sanh et al., 2019) and made available by HuggingFace. The model accepts text as input (in our case a single noun) and returns a sentiment classification (positive or negative) together with a score showing the probability of the positive or negative classification. We found that all nouns in our data were classified as either very strongly positive or very strongly negative, such that the score added negligible information, and so we retained only the positive/negative classification and not the score. We compared the coverage and classifications of this model with those of two crowd-sourced sentiment lexicons, the NRC Word-Emotion Association Lexicon, or EmoLex (Mohammad and Turney, 2013), and the NRC VAD Lexicon (Mohammad, 2018), and found that: (1) DistilBERT covered approximately twice as many of the nouns in our data as did the crowd-sourced lexicons; (2) DistilBERT and EmoLex agreed in their negative/non-negative classifications for 86% of the nouns they both covered; and (3) the results we report here are qualitatively unchanged if EmoLex or VAD are used instead of DistilBERT, with attention restricted to the words covered by those lexicons.

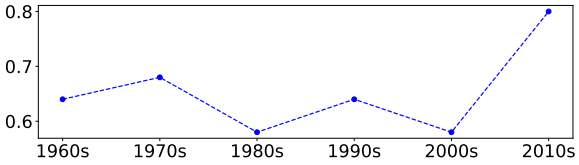

Figure 3: Proportion of the top 50 nouns that are newly prominent in a Palestinian context that have positive valence, per decade. The words analyzed here are those shown in Table 2.

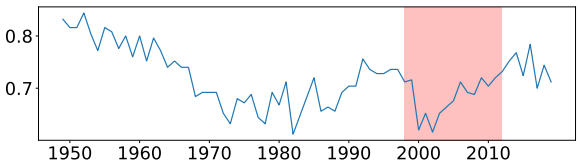

Figure 4: Proportion of the top 250 nouns that are newly prominent in a Palestinian context that have positive valence, by year. The time period shown in pink is the target period 1998-2012.

This confirms the impression that there is more positive sentiment among the newly prominent *Palestinian*-modified words of the 2010s, compared with those of earlier decades.[8]

Finally, we wished to move to an even finer temporal grain, so we re-ran the above analysis, but now targeting specific years rather than decades. For each year, we used Equation 1 to identify which words showed disproportionately greater relative frequency in that year, compared with all previous years since 1948 — that is, which words were newly prominent in a Palestinian context in that year. Space does not permit presenting the actual words retrieved, but Figure 4 shows the proportion of positive valence among the top 250 words per year; the graph for the top 50 is similar but noisier. As with the by-decades analysis above, the overall timeline seems broadly consistent with the historical record. We see a decline in positivity, from a Western standpoint, with the rise of the PLO in the mid-1960s through the 1970s, at a time when Palestinians were most saliently associated in the West with guerilla activity. There is a clear rise in positive sentiment in the 1990s, perhaps reflecting the brief period of Palestinian quasi-respectability in the West that was associated with the Madrid and Oslo processes, and the hopes that this could lead to a resolution of the Palestine question. This positivity continues into the beginning of the target period (1998-2012), suggesting that the cultural shift we are interested in may have ridden on the coattails of the generally more positive sentiment of the 1990s. This is followed by an abrupt decrease in positive sentiment at the beginning of the Second Intifada in 2000 — followed by a recovery by the mid-2000s. Overall, this by-year profile is consistent with the by-decade analysis above, and in particular with the mixed picture for the 2000s seen in that analysis.

We take these results — especially the by-decade results above — to suggest that the cultural shift is historically novel.

Study 3: When did specific words first appear in a Palestinian context?

The above analyses concern increases in frequency, of the sort that we have seen for *Palestinian nakba*. But *Palestinian nakba* did not just increase in frequency — it increased from essentially zero (recall Figure 1). Are there other *Palestinian*-modified words that also increased from essentially zero during the same target period, and if so, do these words also suggest an openness to Palestinian perspectives? And to maintain a fine temporal grain: in which specific years did such words appear?

To answer these questions, we determined, for each *Palestinian*-modified noun in the lexicon (i.e. each noun preceded by the national adjective *Palestinian*, e.g. *Palestinian restaurants*), when that adjective-noun pair entered the lexicon. We did this by examining the raw frequency profile for that adjective-noun pair over time, and identifying the first year for which the frequency of that adjective-noun pair reached 10% of its maximum.[9] Figure 5 illustrates this for the noun *subjectivity* (i.e. the bigram *Palestinian subjectivity*). We were interested specifically in which bigrams entered during or after the period of interest, 1998-2012, although of course many bigrams entered before

[8]The DistilBERT-based sentiment analysis model returns the values positive and negative, but not neutral, and many words that intuitively appear to have neutral valence, e.g. *materials, comment*, are classified as positive by this model. For this reason, it may be helpful to think of positive classification from this model as meaning either positive or neutral, or equivalently non-negative.

[9]Regier and Khalidi (2024) used a different method, based on that of Xu et al. (2016, Supporting Information), which is in turn based on that of Kass et al. (2014, sections 14.2.1 and 14.2.2). We found that our simpler method yielded more satisfactory results. For example, the Regier and Khalidi (2024) method marks *Palestinian intifadeh* as entering the language in 1983 and *Palestinian intifada* in 1991, despite the similarity in the frequency profiles for the two bigrams. In contrast, the method we use here has both bigrams entering the language in the same plausible year, 1988.

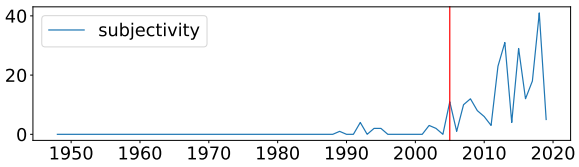

Figure 5: Frequency trace for the bigram *Palestinian subjectivity*, with its year of entry into the lexicon, 2005, marked by a vertical red line.

that time. For this reason, we focused on those *Palestinian*-modified nouns that entered the language between 1998 and the last year of the dataset, which is 2019. These are shown in Table 3, with the adjective *Palestinian* left implicit. It can be seen that there were *Palestinian*-modified nouns other than *nakba* that entered the lexicon during this period and that suggest openness to Palestinian perspectives. An early example is *sumud*, and the food theme appears again as well: *restaurants, dishes, food, kitchen*, along with other words that suggest a psychological closeness to Palestinians: *memory, testimonies, moms, filmmaking, subjectivity*. We saw earlier that many of these words increased in frequency in a *Palestinian*-modified context during the target period; now we see that the corresponding bigrams in fact entered the lexicon at that time. However this is interrupted by a sudden theme of violence around the year 2000, at the beginning of the Second Intifada: e.g. *mortar, homicide, clashes, sniper*. This is consistent with the mixed picture of the 2000s that we saw in the by-decade analysis above, and the by-year sentiment analysis, but now with specific words attached to specific years. By the mid-2000s, many of the newly entering words again suggest a psychological closeness to Palestinian experiences, e.g. *subjectivity* (2005), *food* (2007), *indigeneity* (2010) — but see also *jihadists, jihadi, jihadist*.[10] Overall, we see a shift from initial openness to Palestinian perspectives, to a theme of violence during early 2000s, then back to more openness later on — giving some temporal detail to the idea proposed earlier of the 2000s as a time of contestation and possible change.

We take these results, concerning when *Palestinian*-modified nouns first entered the English lexicon, to support the idea that a general cultural shift occurred during the period of interest, marked by greater openness to Palestinian perspectives. These results also help to characterize that shift at a finer temporal grain.

[10]Recall that the *jihadist* and *indigeneity* themes were shared with the American prototypicality analysis.

1998: restaurants practitioners sumud imam livelihoods debts nakba photographers
1999: edition gunfire bomber soccer plates print dishes noncompliance survey testimonies compilations lawmaker localities
2000: facilitators mortar schoolbooks homicide airspace classrooms clashes telecommunications sniper con gays descendants victimhood hackers snipers medics
2001: publics narratives football stakeholders teens intifadas apologists suicide auxiliaries textbooks corruption imams textbook toddler ceasefire
2002: license symbol reform masculinity rapper presidency reforms website gatherings filmmaker landholder polling premier memories web moms patriation custodianship
2003: rabbinic coordinator spouses mobility spaces space tanna compound memory cinema
2004: cartoon internet jihadists noncitizens jihadi applications jihadist rappers
2005: modernity cinemas queers filmmaking shahid cohort subjectivity interviewee rap
2006: member confl commemoration
2007: nonviolence movies food series
2008: heathenism
2009: hip
2010: indigeneity
2011: bid websites
2012: accession st kitchen
2014: plaintiffs

Table 3: Nouns *n* for which the bigram "Palestinian *n*" entered the lexicon, 1998-2019. Each noun is shown listed by its year of entry. The years 2013 and 2015-2019 are missing from the table because no *Palestinian*-noun bigrams entered the lexicon in those years.

Conclusions

We have shown (1) that the period 1998-2012 witnessed a shift in English language use, marked by greater openness to Palestinian perspectives; (2) that that shift was historically unusual; and (3) that the date of entry of specific *Palestinian*-modified words into English is consistent with that cultural shift. Our findings are consistent with the U.S. polling results of Telhami (2018, 2020) and at the same time elaborate the picture in several respects. The polling results highlighted the 2010s as a period of change; our results confirm that but also underscore the importance of the precursors to that moment: the 1990s as a period of generally positive sentiment toward Palestinians, and the 2000s as a time of apparently conflicting portrayals, and change, eventually resulting in the consolidation, in the 2010s, of a view that is more open to Palestinian

perspectives. Our results also help to characterize that recently-emerging view, highlighting in particular the element of psychological closeness to, and implicitly the humanity of, Palestinians.

Some aspects of the cultural shift we have discussed are specific to Palestinians (e.g. *nakba*), while other aspects appear to reflect a broader recent cultural emphasis on previously marginalized perspectives more generally, not just with respect to Palestine (e.g. *indigeneity*). Finally, some other aspects are even more general (e.g. *internet*).

None of this should be allowed to obscure the fact that to a large extent the Anglophone world and the West more generally remain a hostile informational environment with respect to the Palestine question, especially in the context of the ongoing war on Gaza. The Palestine-sympathetic element of language use that we have identified, although real and increasingly influential, remains a countercultural element, and is opposed by standard governmental and elite opinion in much of the West. Still, we think it is significant that the support for Palestine that can be seen in the West today has a relatively long history, dating back to the late 1990s — because that long history suggests that this support has a degree of momentum and is therefore unlikely to be easily halted or reversed.

We have suggested that this cultural shift may help to explain the increase in prominence of the Nakba in the West — but what explains the occurrence of the cultural shift itself? The evidence we have seen does not allow us to pinpoint a specific cause with any certainty, but does allow informed speculation. We have seen that the 1990s generally were a period of more positive sentiment toward Palestinians, likely because of the Madrid and Oslo processes and the hope that these might lead to a resolution of the Palestine question. The 1990s also saw the advent of the internet, as reflected in some of our analyses, and this new medium allowed wide access to a range of ideas and perspectives that previously had not been as easily accessible. This context may help to explain why mention of the Nakba in English began in the late 1990s, an entire half century after the event itself, and not before. This initial period of apparent openness to Palestinian perspectives was then abruptly challenged by the events of the Second Intifada (2000), also reflected in our analyses, and by the attacks of September 11 2001. Mansour (2002) has argued that the September 11 attacks led some in the West to "lump together indiscriminately" (p. 13) those attacks and the Second Intifada, resulting in a worldview in which the U.S. and Israel were confronting a single opponent. At the same time, those attacks also motivated a number of U.S. scholars to examine the Israel/Palestine question closely, in some cases leading to a critique of U.S. policy toward Israel (e.g. Mearsheimer and Walt, 2007). It appears that there eventually grew out of that contested period a set of linguistic usages that were substantially more open to Palestinian perspectives than those that had preceded them — an initially quiet but tangible precursor to today's still-contested environment.

Limitations

The subset of the Google Books dataset that we use may exhibit some change in composition over time (Pechenick et al., 2015; Schmidt et al., 2021), despite our best efforts to mitigate this issue. Moreover, the dataset only covers years through 2019, and thus misses the period of cultural change connected with the ongoing war in Gaza. The restriction to (national adjective, noun) bigrams misses cultural trends that may be reflected elsewhere in language. Finally, the data are drawn from books and are thus skewed toward the language of the cultural elite. These limitations could be addressed by attempting to replicate using a different data source that avoids these issues. The overall argument could be strengthened by more thoroughly probing the robustness of our results to changes in specific methods used for e.g. sentiment analysis, prototypicality for a given time period, and the like.

Acknowledgments

We thank Dan Jurafsky, Charles Kemp, and 3 anonymous reviewers for helpful comments on an earlier draft of this paper. Any errors are our own.

References

Nahla Abdo and Nur Masalha. 2018. *An oral history of the Palestinian Nakba*. Zed Books, London.

Wael F. Al-Sarraj and Heba M. Lubbad. 2018. Bias detection of Palestinian/Israeli conflict in Western media: A sentiment analysis experimental study. In *2018 International Conference on Promising Electronic Technologies (ICPET)*, pages 98–103.

Diana Allan. 2021. *Voices of the Nakba: A Living History of Palestine*. Pluto Press, London.

Jaya Caporusso, Damar Hoogland, Mojca Brglez, Boshko Koloski, Matthew Purver, and Senja Pollak. 2024. A computational analysis of the dehumanisation of migrants from Syria and Ukraine in Slovene news media. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 199–210.

Kathleen Christison. 1999. *Perceptions of Palestine: Their influence on U.S. Middle East policy*. University of California Press, Berkeley and Los Angeles, CA.

Felix Hamborg, Anastasia Zhukova, and Bela Gipp. 2019. Automated identification of media bias by word choice and labeling in news articles. In *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, pages 196–205.

Nick Haslam. 2006. Dehumanization: An integrative review. *Personality and Social Psychology Review*, 10(3):252–264.

Robert E. Kass, Uri T. Eden, and Emery N. Brown. 2014. *Analysis of Neural Data*. Springer, New York.

Camille Mansour. 2002. The impact of 11 September on the Israeli-Palestinian conflict. *Journal of Palestine Studies*, 31(2):5–18.

Nur Masalha. 2012. *The Palestine Nakba: Decolonising History, Narrating the Subaltern, Reclaiming Memory*. Zed Books, London.

John J. Mearsheimer and Stephen M. Walt. 2007. *The Israel Lobby and U.S. Foreign Policy*. Farrar, Straus and Giroux, New York.

Julia Mendelsohn, Yulia Tsvetkov, and Dan Jurafsky. 2020. A framework for the computational linguistic analysis of dehumanization. *Frontiers in Artificial Intelligence*, 3.

Jean-Baptiste Michel, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, and Erez Lieberman Aiden. 2011. Quantitative analysis of culture using millions of digitized books. *Science*, 331(6014):176–182.

Saif Mohammad. 2018. Obtaining reliable human ratings of valence, arousal, and dominance for 20,000 English words. In *Proceedings of The Annual Conference of the Association for Computational Linguistics (ACL)*, Melbourne, Australia.

Saif Mohammad and Peter Turney. 2013. Crowdsourcing a word-emotion association lexicon. *Computational Intelligence*, 29:436–465.

Burt L. Monroe, Michael P. Colaresi, and Kevin M. Quinn. 2008. Fightin' words: Lexical feature selection and evaluation for identifying the content of political conflict. *Political Analysis*, 16:372–403.

Susan Opotow. 1990. Moral exclusion and injustice: An introduction. *Journal of Social Issues*, 46(1):1–20.

Eitan A. Pechenick, Christopher M. Danforth, and Peter S. Dodds. 2015. Characterizing the Google Books corpus: Strong limits to inferences of socio-cultural and linguistic evolution. *PLoS ONE*, 10:e0137041.

Terry Regier. 2016. Perceptions of Palestine: The view from large linguistic datasets. *Journal of Palestine Studies*, XLV:41–54.

Terry Regier and Muhammad Ali Khalidi. 2024. Palestine and the Western "street". *Security in Context*. Policy Paper 24-06.

Ahmad H. Sa'di and Lila Abu-Lughod. 2007. *Nakba: Palestine, 1948, and the Claims of Memory*. Columbia University Press, New York.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. In *5th Workshop on Energy Efficient Machine Learning and Cognitive Computing - NeurIPS 2019*.

Benjamin Schmidt, Steven T. Piantadosi, and Kyle Mahowald. 2021. Uncontrolled corpus composition drives an apparent surge in cognitive distortions. *Proceedings of the National Academy of Sciences*, 118(45):e2115010118.

Yasmeen Serhan. 2023. The American public's views on Israel are undergoing a profound shift. Washington hasn't caught up. *Time*. July 19, 2023.

Shibley Telhami. 2018. Americans are increasingly critical of Israel. *Foreign Policy*. December 11, 2018.

Shibley Telhami. 2020. Changing American public attitudes on Israel/Palestine: Does it matter for politics? In *Israel/Palestine: Exploring A One State Reality — POMEPS Studies 41*, pages 76–82.

Joshua Tenenbaum and Thomas Griffiths. 2001. The rational basis of representativeness. In *Proceedings of the 23th Annual Conference of the Cognitive Science Society*, pages 1036–1041.

Yang Xu, Terry Regier, and Barbara C. Malt. 2016. Historical semantic chaining and efficient communication: The case of container names. *Cognitive Science*, 40(8):2081–2094.

Wajdi Zaghouani, Mustafa Jarrar, Nizar Habash, Houda Bouamor, Imed Zitouni, Mona Diab, Samhaa El-Beltagy, and Muhammed AbuOdeh. 2024. The FIGNEWS shared task on news media narratives. In *Proceedings of The Second Arabic Natural Language Processing Conference*, pages 530–547. Association for Computational Linguistics.

Constantin Zureiq. 1948. *Ma'na al-nakba [The meaning of the catastrophe]*. Dar al-'ilm lil-malayin, Beirut.