2023

WE CAN DO THIS / JUNTOS SÍ PODEMOS: A CRITICAL DISCOURSE ANALYSIS OF GOVERNMENT-SPONSORED COVID-19 PUBLIC SERVICE ANNOUNCEMENTS (PSAS) IN ENGLISH AND SPANISH

Katalina R. Peterson
Scripps College

WE CAN DO THIS / JUNTOS SÍ PODEMOS:

A CRITICAL DISCOURSE ANALYSIS OF GOVERNMENT-SPONSORED COVID-19 PUBLIC SERVICE ANNOUNCEMENTS (PSAS) IN ENGLISH AND SPANISH

by

KATALINA R. PETERSON

SUBMITTED TO SCRIPPS COLLEGE IN PARTIAL FULFILLMENT OF THE DEGREE OF BACHELOR OF ARTS

PROFESSOR CARMEN FOUGHT, PITZER COLLEGE

PROFESSOR DAVID DIVITA, POMONA COLLEGE

PROFESSOR MARINA PÉREZ DE MENDIOLA, SCRIPPS COLLEGE

APRIL 28, 2023

Acknowledgements

I'd like to start by expressing my gratitude to my thesis readers for their guidance throughout my thesis experience. To Carmen, thank you for always motivating me to pursue things I find interesting, and for introducing me to the girls (Cricket and Frida) who kept me company on days when I worked on this project. To David, thank you for challenging me as a critical thinker, meeting with me during my mid-thesis crises, and being my honest voice of reason. To Marina, thank you for all of the love and grace you have shown me all these years, and for the excitement you felt for me when I couldn't feel it for myself. You all inspire me as a student and person as I move forward in my graduate studies.

Shout out to Galia and Robin for helping me flesh out the quantitative aspects of this project! Also to the rest of my Senior Sem peers who kept me caffeinated at Nambah!

Meg and Clare, you rallied on my behalf when I lost faith. Despite the late night zoomies, you accept me for who I am. You're the best friends I could ask for.

To the people who have been by my side my entire life: Thank you Grandma Judy, Grandma Rita, Papa Jim, and Uncle Mike for all of your love and support in college, but mostly in my life. To Kim, Mark, and Kaylana, you guys have always inspired me to live fully and I have grown by your example. Luke and Alayna, you make me want to make this a better world for you. There is no way to repay you all for what you've done for me, but I will keep trying.

Finally, I dedicate this paper and my college diploma to my parents. I've studied language for years now and still haven't found any words that describe how much you both are a part of everything I do and aspire to be. In these pages live the fruits of the determination, work ethic, curiosity, and care you planted in me.

To those I haven't mentioned here, thank you. I could write a thesis of acknowledgements to you. So without further ado…

Abstract

Communication inequities are known to negatively impact people from socioeconomically and linguistically disadvantaged backgrounds during public health crises (Gomez-Aguinaga et al., 2021). In the United States, Hispanics—especially those who speak Spanish—have been among the communities most disproportionately affected by COVID-19 (National Center for Health Statistics, 2022). The pandemic has underscored the importance of understanding the linguistic and discursive strategies implemented by institutions entrusted with disseminating public health information to reach diverse audiences, especially the most vulnerable. This paper analyzes the linguistic practices used to create public service announcements (PSA) produced in English and Spanish by health agencies at both the federal and local levels during the COVID-19 pandemic. Using mixed methods—including quantitative corpus analysis and qualitative discourse analysis—I examine how public health institutions produce media that engages with the perceived sociocultural contexts surrounding intended audiences. By identifying linguistic trends from a corpus of PSAs produced by government-organizations surrounding COVID-19 vaccine efforts, this discourse analysis shines light on powerful but subtle ideologies about communities of language speakers. Despite the assumption of direct translation between languages, my analysis shows there is indeed a difference between messaging in comparable government-sponsored COVID-19 PSAs in English and Spanish. Generally, PSAs in Spanish at the federal level employ familiarity, collectivism, and community-specific language more than comparable PSAs in English. This could suggest beliefs about Spanish-speakers in the United States as being more collectively-motivated than individually-motivated. This linguistic analysis has the potential to elucidate the kinds of language strategies that make public health PSAs most effective for different audiences, or to suggest how the current strategies employed by public health agencies could be better suited for culturally diverse audiences.

Keywords: COVID-19, public service announcements, Spanish, public health media, critical discourse analysis, keyword analysis, sentiment analysis, sociolinguistics, ideologies

TABLE OF CONTENTS

We Can Do This / Juntos Sí Podemos:

A Critical Discourse Analysis of Government-Sponsored Covid-19 Public Service Announcements (PSAs) in English and Spanish

Relevant Background

COVID-19

The COVID-19 virus has unfortunately become a ubiquitous part of life throughout the world. Its status within public consciousness makes it difficult to remember a time before the onset of the devastating virus. Its progression over a short amount of time left many suffering from the physical, emotional, and financial effects of its onslaught. As early as January 2020, the World Health Organization declared the novel coronavirus as a "public health emergency of international concern" (World Health Organization (WHO), 2020). By early March 2020, the Director-General of the WHO classified COVID-19 as a "pandemic" and encouraged countries to take "urgent and aggressive action" in controlling the spread of COVID-19 (WHO, 2020). Development of a vaccine to combat the spread of coronavirus began and by December 2020, the FDA had approved the first emergency use authorization of a vaccine (Pfizer-BioNTech) in the United States for people over 16 at risk of infection (U.S. Department of Health and Human Services (DHHS), 2022). In April 2021, the White House announced that COVID-19 vaccines would become available for everyone over the age of 16 (DHHS, 2022). In November 2021, the first booster access for adults was approved by the Health and Human Services Secretary (DHHS, 2022). Still, despite the rapid progress made on vaccine development, about 15% of American adults remained unvaccinated at the end of 2021 (Monte, 2021) while death rates in

the U.S. totaled over 450,000 in 2021 alone (National Center for Health Statistics (CDC), 2022). In the time since the first vaccines were approved for all adults, vaccination education efforts from healthcare organizations, community organizers, celebrity figures, and local governing bodies dominated mass media. The advent of this type of media was certainly not disconnected from coinciding political division and general sociocultural context in the United States. In fact, previous literature has contended that "managing pandemics is not *in vacuo*…[pandemics are] entangled in the cultural and social contexts that define them" (Musolff et al., 2022, p. 47). COVID-19 became a global event leading to a burgeoning collection of media that has shifted the landscape of public health media in the United States.

Elevated Risk for Latino Communities

Throughout the COVID-19 pandemic in the United States, health disparities have come to the forefront, sparking conversations about access, identity, and elevated risk for death. It is well-documented through epidemiological data that Hispanic, non-Hispanic black, and Native American communities have been negatively impacted by COVID-19 at disproportionate rates compared to non-Hispanic White communities (National Center for Health Statistics (CDC) 'Mortality', 2022; National Center for Health Statistics (CDC) 'Health Disparities', 2022; and Mehring et al., 2021). One of the reasons why Latinos might experience higher exposure risk, disease prevalence, and mortality is due to the lack of public health information in Spanish.

The United States is the fourth largest Spanish-speaking country in the world, and by 2060, it's estimated to become the second largest (Instituto Cervantes, 2021). As of 2017, approximately 73% of 37 million Latinos in the U.S. spoke Spanish at home (Krogstad & Lopez, 2017). Given that producers of public health media have focused on "the potential for health

education of newer media technology such as the internet", one might expect that diverse media, such as media in languages other than English, are abundant and accessible (Wilkin & Ball-Rokeach, 2006). In fact, a 2021 publication found that only 15% of YouTube media is in Spanish, while 66% is in English (Instituto Cervantes, 2021). Nonetheless, "the extant literature on infectious disease outbreaks explores how communication inequalities during public health crises produce negative outcomes for individuals with lower socio-economic status, communication barriers, and lower knowledge levels" (Gomez-Aguinaga et al., 2021). Significantly, Knuesel et al. (2020) found that "when clinicians and patients communicate in the same language, there is demonstrably improved adherence to treatment plans, increased patient insight into health conditions and improved delivery of health education" (p. 109). Furthermore, the American COVID-19 Vaccine Poll showed that trust in Spanish-language news among Latinos not only resulted in more positive assessments of the COVID-19 vaccine, but higher vaccination rates as well (African American Research Collaborative, 2021). Given these high stakes, I seek to understand how public health agencies in the United States approach COVID-19 PSAs in English and Spanish to reach diverse audiences.

YouTube

Given the scale of the COVID-19 pandemic and the eruption of material associated with it, it would be nearly impossible to study all the kinds of media covering COVID-19 public health information. As the pandemic reached the United States, many institutions of learning and workplaces shifted to an online model, which made the internet a resource that 302.28 out of 331.9 million Americans had access to and relied on for everything from grocery shopping to public health information in 2021 (Statista Research Center, 2022; United States Census Bureau,

2022). For this reason, I decided to use YouTube to find suitable media for analysis. To establish YouTube as a viable platform for public health information, I researched additional information about accessibility and academic precedence. The results of this investigation show that over 2 billion users access YouTube monthly, making it the second most popular online social platform[1], with many more accessing the platform while not logged into a Google account (YouTube, 2022). Furthermore, with over 500 hours of content uploaded every minute, there would be no shortage of media to choose from on the site. Numerous academic publications in the last two decades have utilized YouTube for data, including publications that mainly studied public health videos for previous epidemics such as the Zika virus, Ebola virus, and Swine flu pandemic (Hernández-García & Giménez-Júlvez, 2020). These precedents confirmed that YouTube could be considered a viable platform for this type of academic analysis.

Government-Sponsored Public Service Announcements

Public Service Announcements (PSAs) are "a form of communication used by non-profit organizations, grassroots movements, and government/military divisions to campaign for social change or educate the public about resolving specific issues" (Purdue Global Academic Success and Writing Resource Center and Blog, 2023). They are usually produced in either video or audio mediums and circulated on radio, television, or online platforms. They often attempt to reach a target audience to raise awareness about a specific issue; for example, underage drinking, the effects of medication, or in this case, COVID-19.

[1] Second to Facebook.

After choosing YouTube as the source of public service announcements, the next task was narrowing PSAs by sponsor or authorship type. Despite the wide-ranging options made available by the internet, I decided that the media used in this research needed to be robust enough for analysis, have a verifiable date, have an attached sponsor or funder, and contain credible information that is likely to reach American audiences. For this reason, I chose verified accounts of government organizations for this analysis: the U.S. Department of Health and Human Services (DHHS) and the Los Angeles County Department of Public Health (LACDPH).

The reasons for choosing government media on a logical level were clear: the videos were neatly organized, all centered around COVID-19, had attached sponsors, utilized credible information and speakers, and were likely to be used in the daily messaging Americans receive about COVID-19 at work, school, and in medical spaces. Additionally, by choosing government organizations as the sources of PSA media, institutional authority regarding public health information could be analyzed more closely since all accounts chosen are leading departments of public health in the United States. Combating misinformation and avoiding hoaxes—which formed approximately 11% of all videos analyzed in a previous study analyzing COVID-19 YouTube videos in Spanish (Hernández-García & Giménez-Júlvez, 2020)—are central to the goals of government-sponsored public health messaging and is perhaps one of the reasons why there is a robust collection of COVID-related material on government-sponsored YouTube accounts. Furthermore, Abbas (2021) adds that misinformation "push[es] the WHO to stress the necessary need for fighting *infodemics* on one hand and trusting and respecting science on the other" (p. 1169). The role government-sponsored media plays in mitigating this so-called infodemic places it in a position ripe for further consideration in terms of how the messaging

implicitly makes assumptions about or impacts audiences at the language level. Further discussion about how specific media was chosen will be included in the methodology section to follow.

Federal vs. Local Level Justification

I decided to compare PSAs at the federal (DHHS) and local (LACDPH) levels because I was interested to see if any differences in messaging emerged due to the differences in expected roles at different levels of government. The 10th Amendment holds that any responsibility not constitutionally outlined for the federal government is reserved for state governments. The responsibility of "promoting health and well-being" is one that is constitutionally reserved for states under the 10th Amendment; however, "states and their local governments suffer from a significant lack of resources and interjurisdictional competition during major emergencies" (Cigler, 2021, p. 1). Some have suggested that the pandemic has challenged federalism's ability to implement a united response to COVID-19, "with its division of powers and reservation of significant authority to state governments resulting in fragmentation of authority, policy-making, and implementation" (Cigler, 2021, p. 1). Although federal level public health agencies tend to focus on infectious diseases while local level agencies focus on chronic diseases like diabetes, both levels ought to be responsible for relaying "honest, accurate and timely information to the public" (Cigler, 2021, pp. 6-7). How do the agencies differ (or not) in their approach to public health messaging about COVID-19? What about in multilingual contexts? What type of preparation do the agencies use to develop audience-specific media with the resources they are allotted?

Critical Discourse Analysis

The qualitative analysis tool chosen for this project is critical discourse analysis (CDA). Broadly, CDA is an interdisciplinary, often multimodal approach to analyzing language within its wider sociocultural context. Linguistic articles in the recent past have relied on CDA for their analyses of political statements made by national leaders during the COVID-19 pandemic (Abbas, 2021; Papamanoli & Kaniklidou, 2022), news media related to COVID-19 (Zorzi, 2022), public health organizations' webpages (Salama, 2022; Luzón, 2022), and even children's books about COVID-19 (Muelas-Gil, 2022). However, despite this rather large collection of articles using CDA for COVID-19 media, few used it to study YouTube videos, let alone those sponsored by government-organizations in two languages.

The purpose of discourse analysis has been defined by theoretical sociolinguist James Paul Gee as a tool to understand the primary functions of human language, which he suggests are "1) to scaffold the performance of social activities, and 2) to scaffold human affiliation within cultures, social groups, and institutions" (Gee, 1999, p. 1). I was particularly drawn to Gee's description because it highlighted the importance of critically analyzing language within its sociocultural context and effectively establishes language as a tool that facilitates social interaction among people from different backgrounds. For my analysis, it was important to understand Gee's description of what he calls *social languages*, which he says "represen[t] the values and interests of distinctive groups of people" (p. 29). Working under the frameworks of *social languages* and *discourse* allowed me to carry out an analysis that bore in mind the construction of messages that either implicitly or explicitly served to situate social identities and activities. Gee (1999) even goes so far as to suggest that these messages are constructed

reflexively, meaning the language used in different *social languages* "both creates and reflects contexts in which [it is] used" (p. 40). In all, critical discourse analysis was chosen because of its potential as a reliable tool to "uncover hidden ideologies and power relations in discourse" surrounding public health media disseminated by government health organizations.

Research Goals and Overview of Project

Among COVID-19 media that emerged early in the pandemic were public service announcements (PSAs) that spread COVID-19 information in creative ways to reach American audiences. One of my earliest memories from the vaccine rollout in the United States was hearing a PSA on the radio featuring a Native American speaker advocating for vaccine acceptance in Native American communities. I was struck by the PSA's attention to cultural detail and incorporation of perceived values such as connection to the land to reach a specific audience. Some groups have attempted to develop culturally-appropriate public health media in Spanish by relying on "transcreation": content "recreated in Spanish while preserving the meaning, style, tone, and intent of the English text" (Byrne et al., 2020, p. 4). Analyzing attempts at cultural translation like "transcreation" can uncover a host of underlying biases related to language use and the identities of speakers. How are subtleties in language used to impart sociocultural messages in PSAs? What do differences in cultural or linguistic translation reveal about implicit beliefs tied communities of language speakers? How do Spanish-language PSAs differ from English-language PSAs—if at all—in the face of COVID-19 health disparities disproportionately impacting Latino communities in the U.S.? What can public health agencies do to better communicate their audience approaches and effectively reach their target audiences?

In my study, I have taken government-sponsored YouTube media from two different sources: the United States Department of Health and Human Services (DHHS) and the Los Angeles County Department of Public Health (LACDPH). For each source, I matched media in English and Spanish based on them being comparable translations of each other. I then transcribed all media in English and Spanish before coding the videos for variables such as gender, theme, date of publication, and speaker type. A set of corpora made up of the transcriptions were run through both a keyword analysis and a sentiment analysis program to look for superficial differences. Then, each matched pair was analyzed side-by-side for qualitative discourse analysis. Based on both the quantitative and qualitative mixed methods, I give a final analysis of the media by the themes that emerged during the analysis. In the discussion sections of this paper, I explore the implications of my findings and comment on the key takeaways for an audience of academics, public health officials, and other interested parties.

QUANTITATIVE KEYWORD ANALYSIS

Keyword Analysis Method

A project such as this one would first benefit from the quantitative analysis of keywords in Spanish-language and English-language PSAs. Quantitative keyword analysis is a useful first step in this project because it is a simple method for determining any surface level lexical differences between the two bodies of texts. The results of this analysis will contextualize the later steps in my project. For this reason, I consulted two corpus professors to develop a plan for keyword analysis.

Keyword analysis is a tool that tells researchers "which words are unusually frequent (or infrequent) in the corpus in comparison with the words in a reference corpus" (Lihang et al., 2018). In other words, keyword analysis gives insight into which words are most highly representative of the target corpus when controlled for general frequency in the target language.

I began by taking the transcripts I had already created for my media and conjoined them into text files through the application Sublime (Skinner, 2023). For both the DHHS and the LACDPH, I made one text file each for all the transcriptions in Spanish and one each for all in English. This resulted in 4 text files for comparison. I then used a set of two reference corpora—one in English and one in Spanish—that would establish a baseline or "norm" for each language to be used in comparison with the target corpora I had created. For English, I utilized excerpts of the Freiburg-Brown Corpus of American English, otherwise known as Frown (Mair et al., 2009). For Spanish, I utilized excerpts from the Spanish Billion Words Corpus (Cardellino, 2016). Each reference corpus included about one million word tokens. All corpora were then uploaded to AntConc, a free application for corpus linguistics research developed in Python and Qt for data-driven language learning (Anthony, 2022).

Keyword Analysis Results

The top ten keywords most representative of the given samples[2] by agency (DHHS, LACDPH) and language (EN, ES) are as listed in Table 1 below.

The first major difference between the Spanish and English PSAs relates to the prevalence of collective-coded messaging. Three of the top ten keywords in DHHS Spanish are

[2] When compared to a reference corpus.

communal in nature (*podemos, juntos,* and *sabemos*), while two of the top ten keywords in DHHS English are communal in nature (*we* and *people*). In DHHS English, one of the top ten keywords was *your*, which is distinctly not communal in nature. It is possible that *juntos* and *podemos* are included in the keywords because the tagline for the Spanish language campaign by the DHHS was "Juntos sí podemos [Together we can]" whereas in English the tagline was "We can do this". However, this still reflects a difference even in the chosen taglines for each language. Therefore, although the languages appear to be relatively similar in the top keywords, subtle differences in the communal/individualistic nature of the other keywords indicate a difference in messaging for DHHS. For LACDPH media in English, there are no collective coded words and one distinctly non-collective coded word: *you*. In Spanish LACDPH media, there are no collective coded words in the top ten and no non-collective coded words.[3]

These results indicate—at least at the federal level—that more collective coded words are used in Spanish than in English PSAs. However, it is not the same at the local level, with no collective coded words found in the top ten keywords of both English and Spanish. This distinction warrants further exploration into the differences between the federal and local level PSAs, which can be found later in this paper.

[3] *Vacúnate* could possibly be considered a non-collective coded word because it is addressed to an individual, but I have chosen not to include it as such since comparable words in English such as *vaccinate* were not included.

Table 1: Top-ten keywords by agency and language based on keyness (likelihood)

DHHS EN	DHHS ES	LACDPH EN	LACDPH ES
covid	covid	vaccine	vacuna
vaccines	vacunas	covid	covid
vaccine	vacuna	vaccinated	angeles
vaccinated	vacunarse	county	vacunate
get	seguras	com	eloisa
we	**podemos**	eloisa	dosis
um	clinicos	vaccinate	doctora
people	**juntos**	gonzalez	visite
safe	soy	you	vacunas
your	**sabemos**	get	actualizada

QUANTITATIVE SENTIMENT ANALYSIS

Language and Affect

Sentiment or affect analysis is a tool used in linguistics to determine whether a lexical item or utterance indicates more positive inflection or more negative inflection. Linguistic affect can be understood as emotion words and phrases that describe positive or negative inflection, or create them through language. A recent study titled “Linguistic Affect: Positive and Negative

Emotion Words are Contagious, Predict Likability, and Moderate Positive and Negative Affect", the authors found that positive affect results in more likability of the speaker while negative affect results in unlikability (Knuppenburg & Fredericks, 2021). Positive affect has been defined as the "extent to which an individual is active, alert, enthusiastic, happy, [and] engages in pleasurable activities", while negative affect is characterized by "states of hostility, aversiveness, anger, contempt, guilt, fear, nervousness, [and] emotional distress" (Watson et al., 1988). Interestingly, negative affect has been tied to health outcomes, such as increased vulnerability to disease (Mendonça-de-Souza et al., 2007) and states of illness (Mayne, 1999). Meanwhile, positive affect has been shown to correlate with health and well-being (Mayne, 1999). Given the health implications of affect and role of trust in the acceptance of health media, I have chosen to study it in the context of the COVID-19 PSAs.

Sentiment Analysis Method

The tools I used to analyze affect in language were programs designed to assign a sentiment valency score to a given collection of texts. The Natural Language Toolkit (NLTK) (Bird et al., 2009) is a useful NLP tool that uses language corpora to complete various text processing tasks, including sentiment analysis. Various language data from across the internet were sourced to build a corpus of words that were then assigned sentiment scores by a variety of human coders. In this way, the sentiment scores can be said to be indicative of how the average person might interpret the affect of a given text or assign it a score if probed. The program's Sentiment Intensity Analyzer can assign a given sentence a score in each sentiment category: negative, neutral, and positive. For example, the sentence "I love kittens very much" might be assigned a higher positivity score than the negative sentence "I hate kittens" or the neutral

sentence “I have a kitten”. The NLTK model was used for the English-language text from the PSAs. A similar model pre-trained on tweets in Spanish[4] was used for the Spanish-language texts (Wolf et al., 2020). Text files containing transcriptions of all PSAs in English and Spanish included in this study were fed into the program, with aggregated sentiment scores printed for each text.

Based on findings from this study that suggest many Spanish-language materials employ communication strategies that emphasize family, togetherness, and collectivism more than the somewhat sterile English-language materials, I expected the PSAs in Spanish to reveal a more positive sentiment score than PSAs in English. Additionally, I hypothesized that widely-acknowledged distrust in healthcare and media among Spanish-speaking communities (López-Cevallos et al., 2014; Centers for Disease Control and Prevention, 2022) may have resulted in these public health agencies employing a more welcoming and positive tone in the Spanish-language media in an attempt to combat distrust.

Sentiment Analysis Results

The aggregated sentiment scores by text and sentiment type are shown in Table 2 below. These results show that, overall, the texts in English were assigned a significantly higher neutrality score than the texts in Spanish. The Spanish texts were assigned both higher negativity and positivity scores than English texts. This can be interpreted as Spanish being more “emotional” overall. Although all Spanish language data have a majority neutral score, they were assigned a higher negativity score than positivity score. Interesting to note is that the DHHS

[4] Built by a Hugging Face transformer.

(federal-level) Spanish data only shows a 0.01 point difference between negativity and positivity scores while the LACDPH (local-level) shows a 0.1 point difference. This means there is ten times as much difference between negativity and positivity scores (favoring negativity) between the federal and local levels. A breakdown of results appear as follows:

DHHS:

- EN 30.67% more neutral than ES
- ES 18.67% more negative than EN
- ES 12% more positive than EN

LACDPH:

- EN is 18.22% more neutral than ES
- ES is 13.92% more negative than EN
- ES is 4.3% more positive than EN

DHHS and LACDPH Comparison:

- DHHS EN is 0.26% more neutral than LACDPH EN: essentially identical
- LACDPH EN is 3.28% more negative than DHHS EN
- DHHS EN is 3.02% more positive than LACDPH EN
- LACDPH ES is 12.18% more neutral than DHHS ES
- DHHS ES is 1.47% more negative than LACDPH ES
- DHHS ES is 10.71% more positive than LACDPH ES

Table 2: Aggregated Sentiment Scores for Individual Text Files

File Content	Agg. Negativity Score	Agg. Neutrality Score	Agg. Positivity Score
DHHS_EN_ALL	0.0547724868	0.8337619048	0.1114550265
DHHS_ES_ALL	0.2414310588	0.5271545672	0.2314143762
LACDPH_EN_ALL	0.0875617978	0.8311573034	0.0812977528
LACDPH_ES_ALL	0.2267451361	0.6489203737	0.1243344777

Sentiment Analysis Discussion

My hypothesis that positivity sentiment scores would be higher in Spanish texts than English texts is supported by this sentiment analysis. The general trend of English texts being more neutral than Spanish texts is also suggested by these findings.[5][6] The quantitative methods I have used—keyword analysis and sentiment analysis—shed light on overall trends but do not pinpoint examples in the text where language interacts with specific themes and ideologies. They suggest that Spanish language PSAs contain more collective and less neutral language while English language PSAs contain more clinical and neutral language, but to better understand the

[5] It is important to note that while these results tend to follow general trends found in keyword analysis and the qualitative work in the next section, there remain limitations to the degree of certainty that can be attached to the results. Since two different pre-trained models were used for the two languages, methodological differences in the algorithms of the models could produce exaggerated differences in results. This is paired with the novelty of sentiment analysis as a natural language processing tool. Despite the limitations and caveats, the results here *are* suggestive of a trend supported throughout this paper, and as such, these sentiment analysis findings are meant to contextualize and justify my use of qualitative discourse analysis.

greater context in which these limited findings occur, I decided it would be valuable to perform a qualitative critical discourse analysis of the texts. This discourse analysis is what I undertake in the following section.

QUALITATIVE DISCOURSE ANALYSIS

Media Selection

To analyze if and how English-language and Spanish-language COVID-19 PSAs differ linguistically, I compiled media from the YouTube channels of the U.S. Department of Health and Human Services (DHHS) and the Los Angeles County Department of Public Health (LACDPH). The DHHS YouTube channel contains curated playlists labeled "We Can Do This", "Juntos Sí Podemos USA", and "Trusted Voices/COVID-19 Vaccine Q&A". The LACDPH YouTube channel contains curated playlists labeled "COVID-19 Vaccine Q&A with Dr. Eloisa Gonzalez". It was from these playlists that I chose the media since they were the only playlists on the channel that contained relevant COVID-19 PSAs. Then, I narrowed the pool of available media by applying a publication date filter, which limited the options to videos that were published between April 1 and November 30 of 2021. This is because this timeframe straddles a critical point in the pandemic in the U.S.—it is the 8-month window in which the first set of vaccines against COVID-19 were approved for all U.S. adults before a booster was offered (DHHS, 2022).

Initially, videos were excluded if they were in a language other than English or Spanish, if they were duplicate videos in a single language, or they were not comparable translations in English and Spanish. This resulted in a rather small sample for the DHHS: 12 pairs of videos, or

24 videos total. From there, I selected some videos for comparison on the basis of shared theme, speaker, and length. This provided 7 more video pairs, or an additional 14 videos for analysis. In all, 19 videos from the DHHS were compiled in Spanish, ~84% of which were comparable to videos in English, and totaled 20.32 minutes of content. For DHHS videos in English, 22 videos were compiled, ~73% of which were comparable to videos in Spanish, and totaled 20.72 minutes. The LACDPH search resulted in 19 PSAs in English and 19 in Spanish, 100% of which were comparable translations.

Transcription and Coding

Following the selection of media, I organized it in a central spreadsheet with a video name that indicated the video number, language, and source; for example, the first Spanish video for the U.S. Department of Health and Human Services would be 1_ES_DHHS and the first English would be 1_EN_DHHS. This naming system also ensured that comparable videos could be easily matched and found. The YouTube links to these videos were then attached to the video titles for ease of revisitation. Along with each video code in the spreadsheet, I included the full video title, date of publication, and length in seconds, which are all basic metrics that can be found on YouTube.

Each video was then carefully transcribed and coded for several variables such as language, theme, speaker type, speaker gender, and month. Language was either coded as English (EN) or Spanish (ES - español). Theme was divided into several interrelated but distinct codes as they related most to the content of the given video: safety (SFTY), access (ACSS), science (SCI), inspiration (INSP), special populations (SPEC. POP.), clinic (CLIN), youth

(YOUTH), community (COMM), and misinformation (MISINFO). While most videos were only assigned one theme, several were assigned two, but no more than three. Next, videos were coded by speaker type: doctor/health professional (D), narrator (N), community organizer (CO), and testifier (T). Speaker gender was coded as either male (M) or female (F) based on conventional cues given by speaker voice and appearance. The last variable to be coded was month: April (1), May (2), June (3), July (4), August (5), September (6), October (7), and November (8). After all media was transcribed and coded, it underwent a rigorous verification phase in which transcriptions were meticulously checked for mistakes, all variables were double-coded for accuracy, and information such as publication date and length in seconds was verified. After this final verification, the data could then be used for analysis.

Qualitative Analysis Method

To develop my methodology for qualitative analysis, I first looked at what scholars have done in the past. In similar studies, authors have hand-coded videos by basic measures to prevent COVID-19 such as handwashing or social distancing, but that was not possible in my study since the media was specifically about encouraging people to get vaccines (Hernández-García & Giménez-Júlvez, 2020). Another study used lexicalization as a tool for discourse analysis, such as how differences in word choice could reveal underlying ideologies (Abbas, 2021). Some examples the author gave of lexicalization included, “Emphasize our positive things, de-emphasize our negative things, emphasize their negative things, de-emphasize their positive things” (p. 1170). The continuum of positive to negative sentiment suggested by the quantitative sentiment analysis made qualitative analysis a curious entrance point to understand the

contextualized differences. It is useful to think of affect as a permeating variable throughout the qualitative analysis.

After choosing and justifying all the qualitative analysis variables, I undertook the task of carrying out a critical discourse analysis. I followed Gee (1999)'s methodology for careful side-by-side analysis, which includes analyzing language at the line, macro-line, stanza, and macro-structure levels to pull themes from the media, resulting in structure that conveys meaning across languages (p. 122). I then analyzed all comparable video pairs line-by-line and identified examples related to my pre-established variables: *speaker introductions, virus agency, speaker identification with the audience, formality, affect, calls-to-action, certainty, pronouns and (non-)inclusive language,* and *family and community*. The findings are presented below in the results section.

Qualitative Critical Discourse Analysis Results

Speaker Introductions

One general theme drawn from the transcriptions was whether or not speakers introduced themselves. One important aspect of building ethos with an audience is introducing the speaker as a credible source or messenger. When speakers did introduce themselves, I noted differences in how they did so. For example, in several DHHS video pairs, the speaker introduced themselves in English, but not in Spanish. There is one example in the reverse, in which the speaker introduces herself and her credentials in Spanish, but not in English. Interestingly, in one of the video pairs, the speaker introduces herself by name in the Spanish version but introduces herself with the title "Doctor" in English. This difference suggests the speaker might have made

this choice to establish more personal rapport with the Spanish-speaking audience or a more professional rapport with the English-speaking audience. Could she have been attending to a perceived difference in ethos preference among speakers? That is, a trustworthy and familiar figure to Spanish-speaking audiences and a professional and credentialed source for English-speaking audiences? The LACDPH set of PSAs features the speaker introducing herself in every video, switching between four different versions of introductions: general, here to provide info for before- and after-vaccines, here to answer top questions, and here to address myths. None of these introductions differ lexically; in fact, they are more closely direct translations than the DHHS video pairs.

Virus Animacy

One interesting variable Papamanoli & Kaniklidou (2022) studied in political speeches about COVID-19 was how the virus is characterized, for example, as a passive threat or an active adversary (p. 51), which is how I established *virus animacy* as a variable. The videos analyzed here are not exempt from alternately describing the COVID-19 virus and vaccines as both animate agents and inanimate threats. One of the most common translation differences in the DHHS sample regarding the animacy spectrum was the labeling of the vaccines themselves. In English, vaccines were most frequently referred to as “COVID-19 vaccines”, while in Spanish this was translated to “vacunas contra el COVID-19 [lit. vaccines against COVID-19]”. While this may be a standard translation from English into Spanish, it does reveal a difference in animacy, with the latter suggesting that the virus is something that can be actively fought against in a way that the English equivalent does not capture. Some English videos did include the phrase “fight COVID”, which is similar in animacy to the Spanish “contra”, but is not nearly as

frequent in the sample. This active/passive distinction is echoed by the LACDPH set in which phrases like “la propagación del virus [the spread of the virus]” and “the virus from being spread” display more virus animacy in the Spanish version.

Two videos in particular from the DHHS set—one in Spanish and one in English—make extensive use of metaphors that lend the virus an animate quality. For instance, the Spanish video used figurative language such as “el virus sigue feliz entre nosotros [the virus happily stays among us]”, “no se para [it won’t stop]”, “el virus queda como acorralado [the virus remains contained/cornered]”, “no encuentra dónde ir [it cannot find where to go]”, and “no sabe ya a quien infectar [it no longer knows who to infect]”. A video in English made extensive use of animacy language as well, such as “infection can’t jump from person to person”, “doesn’t have anywhere to go”, and “not enough hosts to survive”. Therefore, although animacy values are given to the virus in both English and Spanish as a rhetorical tool, there seems to be a tendency for virus animacy to be more pronounced in Spanish versions of PSAs.

Speaker Identification with the Audience

One element of any successful persuasive campaign is having the speaker relate to the audience in a way that builds trust in the speaker, and therefore the message. Gee (1999) suggests a possible option for discourse analysis is the “repeated references to self as claimer”, on which I based the variable of *speaker identification with the audience* (pp. 92-94). In English-language DHHS videos, speakers do make attempts to relate to a broad audience; for instance, one speaker shares how she continues to be protected from smallpox despite not having the vaccine as an analogy for herd immunity, and another describes herself as a relatable “mom of two”. However, the majority of obvious identification with the audience in English videos came

from PSAs in which the target audience was decidedly Latino or another minority community. For example, in one comparable translated video in English, the speaker says “many Latinas and Latinos like us”, while in another the speaker proudly describes herself “as a woman, as a woman of color, and as the first physician in [her] family”. While we can acknowledge the ways in which speakers connected with their English-speaking audiences, it is notable that this effect was especially pronounced in videos directed towards minority audiences.

In Spanish, establishing rapport with the audience seemed to be much more common. Examples range from “para quienes ya no estamos tan jóvenes [for those of us who are no longer that young]”, “como mujer latina [as a Latina woman]”, and even going so far as to claim “somos la voz de nuestros latinos aquí en Gainesville [we are the voice of our Latinos here in Gainesville]”. This rhetorical tool is used to build trust with and express vulnerability to the audience during a public health crisis, and to an extent, with an audience that is disproportionately impacted by the effects of the virus. One particularly striking example of difference in speaker identification with the audience is in a pair of comparable translated videos about the vaccine’s risk to pregnant women and fertility in general. The speaker is the same in each video. In the Spanish video, the speaker at one point identifies with the assumed audience of pregnant people by saying “yo estando embarazada [being pregnant myself]” as a way to relate to listeners’ anxieties. However, in the English translation of this video with the same exact speaker, there is never a mention of the speaker’s own pregnancy, and instead the video comes across as more clinical and emotionally distant than the version in Spanish. In the LACDPH set, the speaker never employs this type of language, even when the speaker herself is Latina.

Formality

Here, formality is being used to mean complexity of language, jargon, and how conversational the speaker is with the audience. Admittedly, grammatical formality is simpler to pinpoint in Spanish, given the existence of formal ("usted") and informal ("tú") second-person forms that do not exist in English. Almost all videos presented in Spanish used the formal "usted" form to address the audience, with only a few opting for the informal "tú" in the case of those geared toward a youthful audience. A couple of interesting lexical differences arose from the transcripts that can be loosely related to formality. For example, in one pair of comparable translated transcripts, the speaker describes the potential side effects of the vaccine. In English, this side effect is described as "pain at the injection site", while in Spanish it is referred to as "dolor en el brazo [pain in the arm]". This difference—albeit slight—does reflect a difference in the complexity of language used, with the Spanish version appearing more simplified. The fact that in other videos in Spanish the more direct translation "dolor en el lugar de la inyección [pain at the place of injection]" makes me consider how perhaps subconscious lexical decisions the speaker makes in the video can reveal something about the level of complexity he deems appropriate for certain audiences.

Another interesting lexical difference comes from a pair of comparable DHHS videos about receiving the vaccine in-clinic. In English, the speaker says people receiving the vaccine will be required to stay 15 minutes "so you can be observed". In Spanish, the same speaker says this is "para asegurarnos que se encuentre bien [for us to make sure you are doing well/ok]". Again, while this semantic difference may be minor, it does carry a connotation difference in

which the English version can be perceived as more clinical or formal, while the Spanish version is more familiar and informal.

In the LACDPH, the style is conversational, but the tone is undoubtedly formal. Examples of differences in language complexity are difficult to find, but those that were annotated showed opposite findings to the DHHS set. In Spanish, lines like "Los viajes de cualquier tipo conllevan riesgos adicionales [Trips of any kind carry additional risks]" appear more formal than the comparable English version "Travel of any kind is risky". Another example is the phrase "desarrollar inmunidad [develop immunity]" compared to "build up immunity". This difference between DHHS and LACDPH is curious because there seems to be discrepancy in the formality of each, with the federal-level preferring more formality in English and the local-level preferring slightly more formality in Spanish. However, the examples from the LACDPH are less convincing than the examples from the DHHS since they could be interpreted as faithful translation or "transcreation" attempts from English to Spanish for phrases that would not sound as familiar to the Spanish-speaking ear if translated directly. The DHHS examples display more concrete differences in word choice that lend themselves to the understanding that PSAs in Spanish carry a more informal tone than those in English.

Calls-to-Action

Nearly every single video in the sample had a call to action at the end that related to getting vaccinated, accessing reliable facts, or inquiring for help from a healthcare professional. However, the manner in which the calls to action are framed in English and Spanish was different at times and could reflect differences in the type of messaging directed towards the respective audiences. For example, in a pair of comparable DHHS videos, the Spanish call to

action is “es su turno de vacunarse [it’s your turn to get vaccinated]”, which in English is instead “it’s time to get vaccinated”. The Spanish version directly addresses the viewer and is more personal than the rather objective English version. In another pair of comparable videos, the call to action in Spanish asks viewers to “asumir tu rol más importante [assume your most important role]” whereas in English it is more informal by directing speakers to “play your most important role”. Although the difference is slight, the connotation of responsibility included in the Spanish version is perceptible and could be drawing on a theme considered important to Spanish-speaking audiences.

In fact, another set of comparable videos draws on responsibility in the Spanish version as well. In the call to action in English, the speaker says “All Latinos should understand the importance of getting a vaccine and how they are able to access it”, while in Spanish the call is that all Latinos should understand that “no solamente están protegiéndose a ellos sino su familia también…tomemos eso como responsabilidad [they are not only protecting themselves, but their families as well…we take this as a responsibility]”. This difference seems to indicate who the target audiences are and perhaps appeals to individualistic thinking in English-speaking audiences and collective thinking in Spanish-speaking audiences. It is also interesting to note that “responsibilidad” in Spanish can be understood to mean both of the distinct English concepts “responsibility” and “accountability”. A final example relates to the mention of loved ones in calls to action. In the Spanish half of a comparable set of videos, the call is “Vacúnense por su salud, por el amor a los suyos, por estar juntos otra vez [Get vaccinated for your health, out of love for your loved ones, and to be together once again]”. In the English half, the call is to get vaccinated “So we can embrace what matters most.” This call is vague and while perhaps

alluding to loved ones and togetherness, makes no explicit mention of either despite being a comparable translation.

Calls-to-action are far less numerous in the local level LACDPH set, with the same slogan call-to-action stated at the end of every video: “Visite vacúnate Los Ángeles punto com para obtener la información más actualizada [Visit vacunatelosangeles.com to get/obtain the most current information” and “Please visit vaccinatelacounty.com for the latest information”. There is a stylistic difference in how this similar call-to-action is framed; in Spanish, there is notably a lack of the polite and respect-indicating word “please”, instead opting for just the imperative mood to give a command.

Certainty

Certainty is, for example, how authoritative a speaker aims to be when sharing public health information. Certainty is also related to another tool Abbas (2021) used in his critical discourse analysis of political statements during COVID-19: evidentiality. He says, “Evidentiality is concerned with the way a speaker or writer qualifies statements by referring to the information source and the degree of reliability of the information source on which the statement is based” (p. 1172).

In the DHHS Spanish set, absolute certainty phrasing was popular; namely, the phrase “todas las vacunas [all of the vaccines]”. English counterparts did not include the “all” qualifier and opted instead for simply “the vaccines”, which is lower on the spectrum of certainty than “all of the vaccines” An example of a difference in certainty between two comparable, translated videos is the certainty encoded in the line “muchas personas no han tenido ningún efecto [many people haven’t had any effect]” vs. “many people have reported no side effects at all”. While at

first glance this difference in language may appear unimportant, in Spanish the certainty is about people not having any side effects, while in English the certainty is about people not *reporting* having any side effects. This signals a difference in certainty about the possibility of side effects following vaccination in English and Spanish. At the same time, significant amounts of hedging are present in both English and Spanish. For example, in English many statements include hedging phrases like "very unlikely", "mild or moderate", "only last a couple of days", "very brief and short-lasting", and "very very rare". In Spanish similar hedging appears in phrases like "un poquito de dolor [a little bit of pain]", "un poquito de malestar en los huesos [a little bit of bone discomfort]", "un poquito más exagerado [a little bit more exaggerated]". In Spanish, there is also the use of hyperbole to convey certainty about the safety of the vaccines: "muy bien, muy bien [very good, very good]", and "prefiero cien mil veces [I prefer it a hundred thousand times]". Therefore, qualitative evidence suggests that messaging in both English and Spanish at the federal level relies on markers of certainty to motivate audiences.

The LACDPH set heavily emphasized the role of the CDC, scientists, and vaccine developers in providing accurate information and containing the virus. For example, the speaker would state things like "Los centros para el control de enfermedades conocido como el CDC en inglés y el condado de Los Ángeles continúan instando a todos que… [The Center for Disease Control, known as the CDC in English, and Los Angeles county continue insisting that everyone….]"[7] as well as "Los médicos y los funcionarios de salud pública están tratando de brindar a las personas buena información [Doctors and public health officials are trying to bring

[7] All quotes from this paragraph are only being presented in Spanish since, as a direct translation, adding the English versions would be redundant for the reader. However, it should be made clear that this language is found in *both* the English and Spanish versions.

people good information]". This set was also more reliant on statistical information such as "una de cada veinte personas que reciben la vacuna no estarán protegidas [one in every twenty people that receive the vaccine won't be protected]" and "La vacuna tiene una eficacia del noventa y cinco para prevenir la enfermedad COVID-19 [The vaccine has an efficacy of 95% in preventing the COVID-19 disease]". What is perhaps surprising is that despite this reliance on hard science and deference to national level agencies, the LACDPH set is full of uncertain statements. A few examples of these types of statements include: "Todavía no se sabe cómo las variantes, o nuevas cepas de COVID-19, afectarán a la eficacia de las vacunas COVID-19 actuales [We still do not know how the variants, or new versions of COVID-19, will affect the efficiency of the current COVID-19 vaccines]", "Esto es porque no sabemos qué tan bien evita la vacuna la propagación del virus [This is because we do not know how well the vaccine prevents the spread of the virus]", and "Todavía no sabemos cuánto tiempo estará protegido después de haber tenido COVID-19 [We still do not know how much time you will be protected after getting COVID-19]". While it is understandable that the LACDPH would want to give the most honest information possible, this uncertainty likely does not calm the myths and qualms the agency has set out to do with this video series. The local level would likely benefit from more certain language use.

Pronouns and (Non-)Inclusive Language

Pronoun use in public campaigns is a particularly interesting variable to analyze since variation in "you" versus "we/us" language can result in differences in interpretation regarding belonging and audience. For instance, inclusive language is missing in some of the Spanish versions of comparable DHHS sets. Take for example the difference between "in our country's

history" and "en la historia del país [in the history of the country]". This small difference in pronoun usage can have implications for the inclusion of Spanish-speakers. It brings to mind whether audiences would even be aware of this difference if they did not have access to English or Spanish. This could be a bias that implicitly affects listeners. In other cases, mention of groups of people is used in English, but not in Spanish. For instance, "people in the United States" and "our kids", which are missing phrases in their Spanish counterparts. However, these differences should not distract from the fact that highly inclusive language is included in many more examples in Spanish, but not in English. Some examples include, "nuestros pacientes [our patients]", "lo que estamos sintiendo es nada más que nuestro cuerpo…[what we're feeling is nothing more than our body…]", " "no vamos a lograr [we are not going to achieve]", "si nos vacunamos [if we get vaccinated]", "sigue feliz entre nosotros [it continues happily among us]", "hemos [we have]", and "comenzamos [we begin to]". In some cases, Spanish versions of translated video pairs use personal possessive pronouns where English does not, including "mis pacientes [my patients]" versus "a lot of people". Another example is "su vacuna [your vaccine]" versus "a vaccine". It is also interesting to note that the phrases "our Latinos" and "Latinas and Latinos like us" are used in both English and Spanish, but only when the video is directed towards a Latino audience. This analysis points to some mixed results, but overall it could be argued that more personal and collective language is used in Spanish than in English. This supports the findings from the quantitative keyword analysis and can be further explored through quantitative means.

The LACDPH set was much more even in terms of (non)-collective language use. All PSAs were in second-person, and the speaker never identified with the audience, instead

reserving use of the pronoun "we" to refer to the LACDPH or public health officials in general. Only in one case is the collective form used, in the line "estamos trabajando para aumentar la inmunidad en nuestra comunidad [we are working to increase immunity in our community]" in which "our community" is used in both English and Spanish.

Family and Community

The quantitative keyword analysis above suggested there might be a difference in how collective or personal language is used in these sets of data. References to family and community were of particular interest in this project because they could reveal trends between the PSAs in terms of focus on collectivism versus individualism. In both the English and Spanish DHHS data, there is reference to "loved ones" and "seres queridos [loved ones]". In rare cases, community is mentioned in English but not its Spanish counterpart: "not only yourself, but your loved ones and the community" versus "a usted y a sus seres queridos [yourself and your loved ones]". However, language about community is used in other Spanish videos, for example "a usted, a su familia y a su comunidad sanos y salvos del COVID-19 [you, your family, and your community healthy and safe from COVID-19]". In Spanish videos, there is a strong emphasis on the contribution of the individual to the community, as exemplified by the following lines: "mientras más gente en la comunidad se vacune [while more people in the community get vaccinated]", "porque todo el mundo está vacunado [because everyone is vaccinated", "uno no se vacuna solamente para uno mismo, uno se vacuna por la gente a su alrededor, para su familia, para sus amistades [one does not get vaccinated only for themselves, one gets vaccinated for the people around them, for their family, for their friendships]", "no eres nada mas tú [it's not just you]", "puedes infectar esa otra persona [you can infect that other person]", "si lo das a tu abuela

o tu abuelo [if you give it to your grandma or your grandpa]". While in English there is reference to loved ones and grandparents (i.e. "yourself and your loved ones"), the focus in Spanish tends to revolve around the concept of family (i.e. "para que usted y su familia [so that you and your family]"). Therefore, while community is mentioned in both English and Spanish, the rate at which it appears could be better understood through quantitative analysis. Family seems to be a theme more unique to DHHS PSAs in Spanish, perhaps as a motivating factor in getting Spanish-speaking Latino communities vaccinated since 'family' is a concept commonly believed to be a value in Latino communities.

Surprisingly, the LACDPH videos did not mention family, loved ones, or friends. Only once was the word "community" used in both English and Spanish without defining who exactly made up the community. Instead, the producers seemed to opt for the more general terms like "people" when talking about the effect not being vaccinated could have on others: "evitará que usted transmita el virus que causa COVID-19 a otras personas [it will prevent you transmitting the virus that causes COVID-19 to other people]" or "para no infectar a los vacunadores [so you do not infect the vaccinators". I did not expect to see such a lack of reference to family and loved ones in any of the PSAs since it is often used as a motivating factor for many individuals.

DISCUSSION OF FEDERAL AND LOCAL LEVEL DIVIDE

As a supplement to this project, I initiated correspondence with public relations professionals at the U.S. Department of Health and Human Services (DHHS) and the Los Angeles County Department of Public Health (LACDPH) to see if perhaps I could interview someone who oversaw the production of the PSAs. This endeavor was useful in establishing

whether the language used in the respective media was based on empirical evidence, focus groups, or other ways of informing the messaging. This inquiry helped me to understand whether the differences were made above or below the level of consciousness on the part of producers.

After approximately three months of correspondence, I received responses to my questions from the DHHS. The statement I received from a strategic planner at the Office of the Assistant Secretary for Public Affairs at the DHHS was:

> As part of the COVID-19 Public Education Campaign, we work closely with a contractor, CDC and others across HHS to develop various materials to reach different audiences, including the Latino community. We strive to ensure that campaign materials are culturally tailored and developed in accordance with plain language standards. The campaign is informed by science, aligned with evidence-based best practices, and guided by research and evaluation. (email correspondence, 03/27/2023)

While I was thankful for their response and did learn about the different elements that go into producing PSA materials, this response seems rather vague and is missing details such as who the contractor is that the DHHS works with, how they reach different audiences and ensure culturally tailored materials, and what phrases like “plain language standards” and “research and evaluation” actually mean. However, the respondent also directed me to pages published by the DHHS about how materials are developed for the COVID-19 Public Education Campaign. This included a page with documents outlining specific audience approaches "to motivate behavior change…specially designed to reach diverse populations that are most likely to take action to protect their health and those disproportionately affected by COVID” (COVID-19 Public

Education Campaign, 2022). Among the documents are ones for reaching Asian American/Native Hawaiian/Pacific Islander audiences, Black audiences, Native American/Alaska Native audiences, rural audiences, Latino audiences, and "general" audiences. I found it interesting which audiences were included in these documents, with audiences such as immigrant audiences, MENA audiences, and even white audiences never mentioned. This could suggest an ideology about the types of people who are considered "general" (read: white/caucasian) and those who are not.

The document titled "Campaign Approach to Reaching Latino Audiences" outlines the disproportionate effect COVID-19 has had on the Latino population in the U.S. and the misinformation that has spread within Latino communities. They point to "the premise that messaging must engage the Latino audience culturally, linguistically, and emotionally, taking into account traditions and family values as well as recognizing challenges and sacrifices—as they have been disproportionately affected by COVID—but also tapping into a sense of hope and optimism". This harkens back to findings from the sentiment analysis that suggests a trend of higher positivity scores in Spanish than in English, suggesting that this type of messaging was purposefully planned and not an unconscious decision. The document also suggests that Latino audiences are more likely to be motivated by celebrity figures they see online, which could perhaps be a reason why speaker introductions and familiarity with audiences appear magnified in Spanish versions of the DHHS PSAs, since they are constructed as celebrity-type doctors. They noted that from April to November 2021, they implemented PSA campaigns (some of which have been included in this study) that addresses elderly Latinos and the value of resilience in emotion-charged ways:

> Since Latinos ages 65 and older have been so disproportionately affected by the COVID-19 pandemic, the messaging in Spanish was both emotional and informative, reassuring them that they are appreciated and valuable, that COVID-19 vaccines are safe and effective, and that vaccination offers the best protection available for our "abuelos" who are at much greater risk of suffering serious complications from a COVID-19 infection, including hospitalization and death. (...) The COVID-19 pandemic is testing resilience, one of the core values of Hispanic communities.

Overall, the DHHS reports relying on focus groups, surveys, and "social listening" to develop what it considers to be the most effective messaging for Latino communities. However, they do not write extensively about the differences in approach for multilingual PSAs for Latino audiences, only stating that they created "culturally appropriate" materials in Spanish, English, and Spanglish.

It is useful to compare the Latino audience approach to the general audience approach, which the DHHS declares is aimed at audiences from diverse backgrounds across the nation. Despite this declaration, there were no separate documents for reaching white audiences, which might suggest that this general audience is a stand-in for "standard", which could imply white audiences. Their statement on language in general audience English-language PSAs is:

> Campaign and external research indicates that it is important to use credible, science-based information in communicating with all audiences about vaccination and to acknowledge that it is normal for people to have questions about the vaccines. The research also indicates that fact-based responses to questions about

> safety and effectiveness are important. Since minimizing people's concerns about the vaccine can increase resistance, the Campaign avoids judgmental language when talking about people who are concerned about getting the vaccine.

These documents appear to support the findings that emotionally-charged language was used in Spanish-language DHHS PSAs targeting Latino communities while "science-based information" and "fact-based responses" were a priority in the general "standard" PSAs in English. Why is science- and fact-based information not a priority for messaging targeting Latino communities? Could this suggest that the DHHS has evidence of Latinos being less persuaded by science-based information, or could it point to an ideology or stereotype about Latinos being motivated more by their hearts than their heads? It might be useful for public health agencies such as the DHHS to state more clearly their evidence-based messaging practices or publish evidence of motivating approaches so as not to potentially perpetuate unfair ideologies about cultural communities or communities of language speakers.

After initial correspondence and numerous follow-ups, the LACDPH never responded to my questions. Given their more limited use of speaker and messaging types, this was a missed opportunity to understand how agencies at the local level might not have the resources to put as much time and funding behind linguistically and culturally appropriate PSA materials. Then again, the lack of response may also be indicative of a lack of staff dedicated to being public-facing health media officials. Federal and state budget administrators should consider the ways in which a lack of funding for the production of public health media could be a hindrance to access of reliable information, especially in L.A. county, where as of 2022, "the age adjusted death rate among Latino/x residents alone is more than two and a half times as high the rate among White

residents" (Los Angeles County Department of Public Health, Office of the County Health Officer, 2022).

CLOSING THOUGHTS

Key Findings

Overall, there is a difference between the messaging in government-sponsored COVID-19 PSAs in English and Spanish. This finding is especially intriguing given that the PSAs I chose for analysis were presented as comparable *translations* of the same PSA, and not as distinct PSAs in separate languages. These differences were even more pronounced at the federal (DHHS) level than at the local level (LACDPH). In general, the DHHS PSAs in Spanish utilized familiarity, collectivism, and community-specific language more than comparable English PSAs from the same timeframe. This could suggest that when DHHS produced this type of media, they were influenced by beliefs about Spanish-speakers in the United States as being more collectively than individually motivated. Based on their audience-approach documents, it seems as though the DHHS *has* put effort into understanding the best ways to culturally translate COVID-19 material for consumption by a Latino audience. However, they do not cite evidence-based facts about how their choices—while likely well-intentioned and giving the impression of being factual—were made or the effects of those choices on specific audiences' willingness to respond to the DHHS' calls-to-action. The process could be made more transparent to show that the choices are well-informed and effective, to avoid the risk of these choices being grounded in potentially problematic ideologies about communities of language speakers.

The fact that the PSAs produced by the LACDPH lacked many of the cultural translation strategies employed by the DHHS (i.e. reference to family, collective-language use, etc.) shows that there is a discrepancy between the two levels of government when it comes to developing media for specific target audiences, specifically Spanish-speaking and Latino audiences.[8] Whether this difference was due to a difference in funding or otherwise, the results are surprising given that Los Angeles County is nearly 50% Hispanic or Latino and nearly 30% of the population is Spanish-speaking (U.S. Census Bureau, 2022).This is compounded by the fact that, while vaccination rates were overall higher for Latinos in LA County than at the national level by the end of November 2021 (57% and 47.3%, respectively), they were statistically significantly lower than among White, non-Latino populations (Kriss et al., 2022; County of Los Angeles Public Health, 2023). It seems possible that a lack of culturally translated language strategies—if found to be evidence-based— in the LACDPH sample could be doing a disservice to Latino and Spanish-speaking constituents who could benefit from a tailored audience approach to the dissemination of public health information. This discrepancy could highlight the need for more resources to be directed to local initiatives, more emphasis on community-based public health media production, and more preparation for a combined media approach between federal and local level government agencies when major public health crises like COVID-19 occur.

Overall, this analysis shows that even seemingly minor differences in language or linguistic cues employed by speakers can have large effects on how public health messaging is

[8] Since the LACDPH did not respond to my request for comment, it remains unclear how exactly they approached creating media for Spanish-speaking and Latino audiences.

presented to target audiences. This study serves to highlight the linguistic strategies that underly public health media with the purpose of motivating audiences to follow through on a call to action. Language is a principal mediator in PSAs connecting health officials with institutional authority to diverse audiences, including those most vulnerable to the effects of health crises like COVID-19. In this way, it is important to analyze language, even at the smallest levels, to hold accountable the institutions that are responsible for keeping us safe and informed, regardless of language or any other sociocultural differences.

Suggestions for Future Exploration

There are many directions in which this study can be expanded upon or narrowed.

Future research should explore whether differences vary by a variety of codes including speaker gender, speaker type, PSA topic, and month published. I coded for these factors in my data, but was unable to carry out their comparative analysis without losing focus of the main language differences. However, it would be interesting to answer questions such as "Are female speakers more likely to use collective-coded language in PSAs than male speakers? How does this vary across languages?" using both the quantitative and qualitative methods employed in this paper. This could potentially lead to interesting findings about the relationship between language use and social factors.

Another suggestion for follow-up work is expanding the quantitative sentiment analysis work that was introduced here. As a later addition to the overall analysis, the sentiment section could be better fleshed out by finding models that are trained on the same language data. It would also be interesting to see what comes of translating the texts and running them through the opposite language's pre-trained model. For example, translating the Spanish text into English

and running that translated text through the English model to look at if any major differences emerge because of model-type. However, this would require careful consideration of translation strategies since translation can be a tricky process in itself.

Additionally, quantitative methods could be utilized to better understand the qualitative variables outlined in the sections above. Concerning the *Family and Community* variable, quantitative word frequency analysis showed that *community* had 4 tokens in Spanish and 6 tokens in English, which amounts to 50% more frequency in English. Meanwhile, the word *family* had a frequency of 7 in Spanish and 3 in English, which is 133% more frequent in Spanish. A puzzle that remains is how to measure the frequency of *you/we/us* language in AntConc given the grammatical structure of Spanish, which is a pro-drop (pronoun dropping) language and grammatically adds number and person to the verb, which English does not have save for plurals. This makes it difficult to find all instances of collective language in Spanish using a corpus manager. Future research might consider methods to account for this grammatical difference in keyword analysis.

Next, while unknown at this time how government-sponsored PSAS *specifically* impact vaccination rates for Spanish-speaking communities, an experimental study that asks participants to make decisions about a speaker or message after watching a PSA could be a way to elucidate the effect of the PSA itself. Testing could include trials across government agencies *and* across PSAs with different funding sources.

As a final suggestion for future exploration, this study would have been illuminated further by a statement from the LACDPH. However, input from a variety of public health institutions could point to trends that would make suggestions for future media production in

English and Spanish more generalizable. In all, work in this area ought to prioritize the transparency of public health media production and equitable strategies for cultural and linguistic translation.

References

Abbas, A.H. (2021). Politicizing COVID-19 vaccines in the press: A critical discourse analysis. *International Journal for the Semiotics of Law, 35*, 1167-1185. https://doi.org/10.1007/s11196-021-09857-3.

African American Research Collaborative. (2021). *COVID-19 Vaccine Poll Methodology*. https://africanamericanresearch.us/covid-poll-methodology/.

Anthony, L. (2022). AntConc (Version 4.2.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software.

Assistant Secretary for Public Affairs. (2022). *COVID-19 vaccines*. U.S. Department of Health and Human Services. https://www.hhs.gov/coronavirus/covid-19-vaccines/index.html.

Bird, S., Klein, E., & Loper, E. (2009). Natural Language Processing with Python. O'Reilly Media Inc.

Byrne, L.M., Cook, S.K., Kennedy, N., Russell, M., Jerome, R.N., Tan, J., Barajas, C., Wilkins, C.H., & Harris, P.A. (2020). Opening doors to clinical trial participation among Hispanics: Lessons learned from the Spanish translation of ResearchMatch. *Journal of Clinical and Translational Science, 5*, e46, 1-7. https://doi.org/10.1017/cts.2020.539.

Cardellino, C. (2016) Spanish Billion Words Corpus and Embeddings. Available from https://crscardellino.github.io/SBWCE/.

Cigler, Beverly A. (2021). Fighting COVID-19 in the United States with Federalism and Other Constitutional and Statutory Authority. *Publius: The Journal of Federalism, vol. 51 (*no. 4), pp. 673-692. https://doi:10.1093/publius/pjab02.

Centers for Disease Control and Prevention. (2022). *Unfair and Unjust Practices and Conditions Harm Hispanic and Latino People and Drive Health Disparities.* Office on Smoking and Health, National Center for Chronic Disease Prevention and Health Promotion. https://www.cdc.gov/tobacco/health-equity/hispanic-latino/unfair-and-unjust.html.

Cooper, C.P., Gelb, C.A., Jameson, H., Macario, E., Jorgensen, C.M., & Seeff, L. (2005). Developing English and Spanish television public service announcements to promote colorectal cancer screenings. *Health Promotion Practice, 6*(4), 385-393. https://doi.org/10.1177/1524839905278759.

County of Los Angeles Public Health. (2023). *COVID-19 Vaccine: Vaccinations in LA County*. County of Los Angeles Public Health Department of Communications and Public Affairs. http://publichealth.lacounty.gov/media/coronavirus/vaccine/vaccine-dashboard.htm.

COVID-19 Public Education Campaign. (2022). *COVID-19 Campaign Audience Approaches.* U.S. Department of Health and Human Services. https://wecandothis.hhs.gov/covid-19-campaign-strategy-and-approach.

Mair, C., Ludwigs, A., Universität Freiburg & Leech, G. (2009) *The Freiburg-Brown Corpus ('Frown') (POS-tagged version).* University of Lancaster. Available from https://varieng.helsinki.fi/CoRD/corpora/FROWN/.

Gee, J.P. (1999). *An introduction to discourse analysis: Theory and method*. Routledge.

Gomez-Aguinaga, B., Oaxaca, A.L., Barreto, M.A., & Sanchez, G.R.. (2021). Spanish-language news consumption and latino reactions to COVID-19. *International Journal of Environmental Research and Public Health, 18*, 9629. https://doi.org/10.3390/ijerph18189629.

Hernández-García, I. & Giménez-Júlvez, T. (2020). Characteristics of YouTube videos in Spanish on how to prevent COVID-19. *International Journal of Environmental Research and Public Health, 17*, 4671. https://doi.org/10.3390/ijerph17134671.

Instituto Cervantes. (2021). *El español: Una lengua viva: Informe 2021*. https://cvc.cervantes.es/lengua/espanol_lengua_viva/pdf/espanol_lengua_viva_2021.pdf.

Knuesel, S., Chuang, W., Olson, E., & Betancourt, J. (2020). Language barriers, equity, and COVID-19: The impact of a novel Spanish language care group. *Journal of Hospital Medicine, 16*(2). https://doi.org/10.12788/jhm.3542.

Knuppenburg, R.M. & Fredericks, C.M. (2021). Linguistic Affect: Positive and Negative Emotion Words are Contagious, Predict Likability, and Moderate Positive and Negative Affect. *Inquiries Journal, 13*(3). http://www.inquiriesjournal.com/a?id=1884.

Kriss, J.L, Hung, M., Srivastav, A., et al. (2022). COVID-19 Vaccination Coverage, by Race and Ethnicity—National Immunization Survey Adult COVID Module, United States, December 2020-November 2021. *MMWR Morbidity and Mortality Weekly Report, 71*, pp. 757-763. https://dx.doi.org/10.15585/mmwr.mm7123a2.

Krogstad, J.M. & Lopez, M.H. (2017). *Use of Spanish declines among Latinos in major U.S. metros.* Pew Research Center. https://www.pewresearch.org/fact-tank/2017/10/31/use-of-spanish-declines-among-latinos-in-major-u-s-metros/.

Lihang, X., Peishan, C., Fang, L., Zhenna Z., & Yuning, Z. (2018). *Keyword List.* AntConc Manual. https://antconc-manual.readthedocs.io/en/latest/keyword_list.html.

López-Cevallos, D.F., Harvey, S.M., Warren, J.T. (2014) Medical mistrust, perceived discrimination, and satisfaction with health care among young-adult rural latinos. *Journal of Rural Health, 30(4),* pp. 344-51. https://doi.org/10.1111/jrh.12063.

Los Angeles County Department of Public Health, Office of the County Health Officer. (2022) Addressing the Needs of Communities Most Impacted by COVID-19: Strategies in Service of Los Angeles County's Latino/x Residents. http://publichealth.lacounty.gov/media/coronavirus/docs/COVIDServingLatinxCommunities.pdf.

Luzón, M.J. (2022). "Coronavirus Explainers" for Public Communication of Science: Everything the Public Needs to Know. In Musolff, A., Breeze, R., Kondo, K., & Vilar-Lluch, S. (Ed.). (2022). *Pandemic and crisis discourse: Communicating COVID-19 and public health strategy.* (Chapter 6). Bloomsbury Publishing.

Mehring, W.M., Kim, J.S., Hendel, C., & Hochman, M.. (2021). COVID-19 exposure risk by speakers of Spanish and English using a web-based self-assessment tool. *Journal of General Internal Medicine, 36*(6), 1835-1836. https://doi.org/10.1007/s11606-021-06756-y.

Monte, L.M. (2021). *Who are the adults not vaccinated against COVID?: Household pulse survey shows many don't trust COVID vaccine, worry about side effects.* United States Census Bureau. https://www.census.gov/library/stories/2021/12/who-are-the-adults-not-vaccinated-against-covid.html.

Muelas-Gil, M. (2022).'COVID Warriors: An Analysis of the Use of Metaphors in Children's Books to Help them Understand Covid-19. In Musolff, A., Breeze, R., Kondo, K., &

Vilar-Lluch, S. (Ed.). (2022). *Pandemic and crisis discourse: Communicating COVID-19 and public health strategy.* (Chapter 7). Bloomsbury Publishing.

Musolff, A., Breeze, R., Kondo, K., & Vilar-Lluch, S. (Ed.). (2022). *Pandemic and crisis discourse: Communicating COVID-19 and public health strategy*. Bloomsbury Publishing.

National Center for Health Statistics. (2022). *COVID-19 mortality overview.* Centers for Disease Control and Prevention. https://www.cdc.gov/nchs/covid19/mortality-overview.htm.

National Center for Health Statistics. (2022). *Health disparities: Provisional death counts for coronavirus disease 2019 (COVID-19).* Centers for Disease Control and Prevention. https://www.cdc.gov/nchs/nvss/vsrr/covid19/health_disparities.htm.

Papamanoli, A. & Kaniklidou, T.P. (2022) COVID-19 Representations in Political Statements: A Corpus-Driven Analysis. In Musolff, A., Breeze, R., Kondo, K., & Vilar-Lluch, S. (Ed.). (2022). *Pandemic and crisis discourse: Communicating COVID-19 and public health strategy.* (Chapter 3). Bloomsbury Publishing.

Pew Research Center. (2021). *Social media fact sheet*. https://www.pewresearch.org/internet/fact-sheet/social-media/#panel-3814afe3-3f3c-4623-910b-8a6a37885ab8.

Purdue Global Academic Success and Writing Resource Center and Blog (2023). *Creating a Public Service Announcement (PSA)*. Purdue University Global. https://purdueglobalwriting.center/creating-a-public-service-announcement/.

Salama, A.H.Y. (2022). Exploring the Multimodal Representation of COVID-19 on the Official Homepage of World Health Organization (WHO): A Social-Semiotic Approach. In

Musolff, A., Breeze, R., Kondo, K., & Vilar-Lluch, S. (Ed.). (2022). *Pandemic and crisis discourse: Communicating COVID-19 and public health strategy.* (Chapter 2). Bloomsbury Publishing.

Skinner, J. (2023). Sublime Text [Computer Application]. Woollahra, Sydney: Sublime HQ Pty Ltd. Available from http://www.sublimetext.com/.

Statista Research Department. (2022). *United States: Number of online users 2018-2027.* https://www.statista.com/statistics/325645/usa-number-of-internet-users/.

United States Census Bureau. (2022). *QuickFacts: United States*. https://www.census.gov/quickfacts.

United States Census Bureau. (2022). *QuickFacts: Los Angeles County, California.* https://www.census.gov/quickfacts/fact/table/losangelescountycalifornia/PST045222. https://data.census.gov/table?t=Language+Spoken+at+Home&g=040XX00US06_050XX00US06037&tid=ACSST1Y2021.S1601.

R&D BluePrint. (2020). *COVID-19 public health emergency of international concern (PHEIC): Global research and innovation forum: Towards a research roadmap.* World Health Organization & Global Research Collaboration for Infectious Disease Prevention. https://www.who.int/publications/m/item/covid-19-public-health-emergency-of-international-concern-(pheic)-global-research-and-innovation-forum.

WHO Director-General. (2020). *WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020.* World Health Organization. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020.

Wilkin, H.A. & Ball-Rokeach, S.J. (2006). Reaching at risk groups: The importance of health storytelling in Los Angeles Latino media. *Journalism, 7*(3), 299-320. https://doi.org/10.1177/1464884906065513.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., & Brew, J. (2020). *Transformers: State-of-the-art natural language processing.* In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations (pp. 38-45).

YouTube. (2022). *YouTube for press*. https://blog.youtube/press/.

Zorzi, V. (2022). Political Comedy and the Challenges of Public Communication during the Covid-19 Crisis: A Corpus-Based Study of Last Week Tonight's Coverage of the Pandemic. In Musolff, A., Breeze, R., Kondo, K., & Vilar-Lluch, S. (Ed.). (2022). *Pandemic and crisis discourse: Communicating COVID-19 and public health strategy.* (Chapter 9). Bloomsbury Publishing.