Exploring E-Cigarette Perception among Spanish and English-Speaking Users from Social Media

Chen-Jui Chen, Anik De, Wonha Shin, Yadi Zhang, Chen Zhang

Team 11, DS 483, 2024 Spring
Goergen Institute for Data Science
University of Rochester

Revision code: Anik, Wonha, Raymond

Introduction

E-cigarette use has become a significant public health concern in the United States, particularly among youth and Hispanic/Latino populations [1,2]. The increasing prevalence of vaping highlights the importance of understanding the sociocultural factors and linguistic variations shaping public perceptions and attitudes. Social media, a central platform for e-cigarette promotion and discussion, offers a unique opportunity to analyze these perceptions across languages [3]. Since English and Spanish are widely used in online conversations, studying these languages is crucial for creating effective public health strategies.

Despite the growing body of research on social media's role in public health, few studies have systematically analyzed social media data in Spanish and English [4]. Even fewer have explored how translations between these languages might influence the interpretation of public attitudes, potentially overlooking critical cultural nuances and linguistic shifts [4]. This gap limits our understanding of how diverse communities perceive e-cigarettes and poses challenges for designing inclusive and effective health interventions.

Our study employs the natural language processing (NLP) models to examine public perceptions of e-cigarette-related tweets in English and Spanish. Additionally, by incorporating an analysis of original Spanish tweets and their English translation, we aim to evaluate whether the translation process alters meanings or interpretations, with implications in public health research. By exploring cultural nuances and linguistic shifts, we strive to identify how these factors influence attitudes, ensuring that health messaging is relevant and effective for diverse communities. Our findings will inform culturally tailored health interventions and campaigns and contribute to best research practices using multilingual datasets.

Our research objectives include:

1. Comparing public perceptions and discussions about e-cigarettes on Twitter/X in English and Spanish.
2. Evaluating the impact of translation on the interpretation of e-cigarette-related tweets by comparing original Spanish tweets with their English translations.
3. Identifying sociocultural and linguistic factors that shape public discourse across and within languages
4. Providing actionable insights to develop culturally sensitive health interventions and targeted public health campaigns that address e-cigarette use among diverse populations.

Data Preprocessing & Methods

The analyses included two datasets of English (n=1,1000,000) and Spanish tweets (n=34,000). A third-party organization downloaded designed tweets related to e-cigarettes through keyword filtering.

Data Cleaning

To ensure the quality and relevance of the datasets, we implemented a systematic data-cleaning process. First, duplicate document IDs were identified and removed to eliminate redundant entries, ensuring each tweet in the dataset was unique. Repost content types, such as retweets, were excluded to focus exclusively on original content. Next, columns containing only null (NaN) values were eliminated, contributing no meaningful information to the analysis. The datasets have also been streamlined by removing non-essential columns deemed irrelevant to the research objectives. With these cleaning steps completed, the datasets were significantly reduced, with the English dataset containing approximately 505,000 tweets and the Spanish dataset containing about 10,000 tweets.

Full-Text Crawling

After the initial data-cleaning process, the English dataset was reduced to 505,000 tweets, while the Spanish dataset contained 10,000 tweets. However, a significant limitation of these datasets was that many tweets included only partial text, which posed challenges for comprehensive analysis. Therefore, web crawling was employed to retrieve the full text of the tweets. Due to the extensive size of the English dataset, a random sample of 100,000 tweets was selected for crawling. This sampling approach provided a representative subset while optimizing resource usage.

NLP Preprocessing

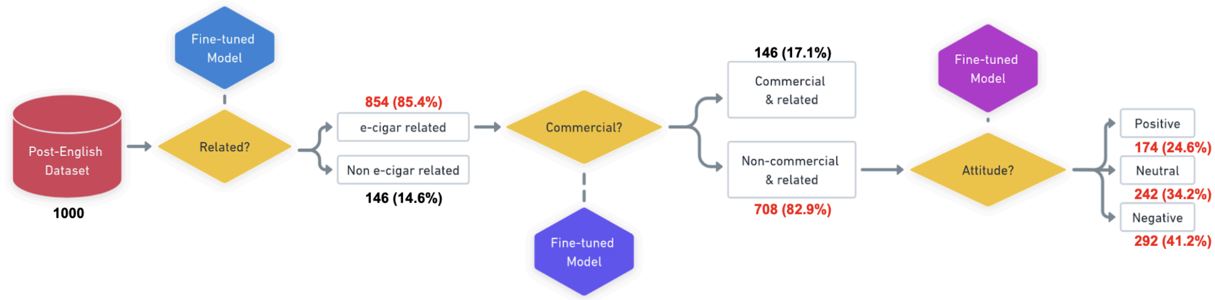
Following the full-text retrieval, NLP techniques were applied to standardize and prepare the textual data for subsequent analysis. This process included lowercasing, removal of stop words, emojis, special characters, and lemmatization. First, all text was converted to lowercase to ensure consistency and eliminate case-sensitive variations. Common stop words, such as "the" and "and," do not add significant meaning to the analysis, were removed to focus on the core semantic content of the tweets. Emojis and other non-alphanumeric characters were stripped from the text to ensure a clean dataset. Lemmatization was then applied, reducing words to their base or root forms, thus standardizing variations of the same word and enhancing the semantic accuracy of the data. These preprocessing efforts resulted in a refined dataset, with the English dataset reduced to 51,000 tweets and the Spanish dataset to 7,000 tweets, which provided a cleaner, more complete foundation for subsequent labeling and model fine-tuning.

Labeling and Model Fine-Tuning

Labeling Process

The final preprocessed datasets were labeled to distinguish between relevant and irrelevant tweets and between commercial and non-commercial content. These labeled datasets were then used to fine-tune machine learning models for classification tasks. The refined data and models were a foundation for more advanced analyses, including attitude classification, ensuring robust and meaningful insights. The labeling process was critical to the foundation of the subsequent modeling efforts. Initially, individual labeling was conducted, followed by group discussions to establish clear and consistent labeling criteria. This iterative process involved labeling a sample of 1,000 lead data points and further group discussions to refine the criteria and ensure high labeling accuracy. Ultimately, the team achieved a reliable labeled dataset of 1,000 tweets.

The labeled dataset revealed that 85.4% of the tweets were e-cigarette-related, and 82.9% were identified as non-commercial. Further classification of attitudes toward e-cigarettes indicated that 24.6% of the tweets expressed a positive attitude, 34.2% were neutral, and 41.2% conveyed a negative attitude. These findings served as the basis for fine-tuning the model, enhancing its ability to identify relevant, non-commercial, and attitude-specific tweets for deeper analysis. Subsequent sections provide further details on the fine-tuning process and its implications for understanding public perceptions of e-cigarettes.



Fine-Tuning and Classification Performance

The fine-tuning process involved three main classification tasks to analyze the dataset: (a) relevance classification, (b) commercial versus non-commercial classification, and (c) attitude classification. Each task utilized advanced machine learning models to achieve high accuracy and F1 scores, reflecting the robustness of the approach.

The first task is relevance classification, which aimed to determine whether tweets were relevant to e-cigarettes. Three models were employed: Bertweet, Twitter Twin BERT Large, and Zero-shot Vanilla Binary Large. Bertweet and Twitter Twin BERT Large outperformed the others, achieving accuracy scores of 96% and 94% and F1 scores of 0.851 and 0.8768, respectively. These results demonstrate that both models performed exceptionally well, with Bertweet excelling in overall accuracy and Twitter Twin Bert Large showcasing a better balance between precision and recall.

The second task sought to classify tweets as either commercial or non-commercial. The same models—Bertweet and Twitter Twin BERT Large—were fine-tuned for this purpose. Bertweet achieved 93% accuracy with an F1 score of 0.87, while Twitter Twin BERT Large reached 91% accuracy with an F1 score of 0.88. This task identified 50,620 English and 6,700 Spanish tweets as relevant and non-commercial, forming a solid foundation for further analysis.

The final task focused on identifying the attitudes expressed in tweets—positive, neutral, or negative. For this, the Bert Rate model was employed, achieving an accuracy of 92% and an F1 score of 0.89. This classification is essential for understanding public perceptions and sentiments regarding e-cigarettes, particularly across diverse linguistic and cultural groups.

Overall, the fine-tuning results across the three tasks demonstrate the efficacy of the selected models for analyzing Twitter data. Bertweet showed outstanding performance in relevance classification and commercial intent tasks, while the Bert Rate Model excelled in handling sentiment-related tasks. The high accuracy and F1 scores across all tasks underscore the robustness and suitability of transformer-based models for social media data classification. These fine-tuning efforts demonstrate the models' effectiveness in extracting meaningful insights from the dataset and lay the groundwork for subsequent analyses and interventions.

Analytical Framework and Topic Modeling Approach

The analysis was structured systematically to leverage the fine-tuned models and extract meaningful insights from the data. After the dataset was prepared and the models were trained, the analysis was conducted on both the English and Spanish datasets. The attitude model was applied to classify and analyze the expressed attitudes to gain deeper insights, allowing for a detailed comparison between attitude and sentiment analysis. An additional layer of analysis was implemented for the Spanish dataset, where the text was translated into English to investigate any potential loss of information or changes in analytical outcomes. This dual approach ensured a comprehensive understanding of the data while maintaining cross-linguistic comparability.

Topic Modeling Pipeline

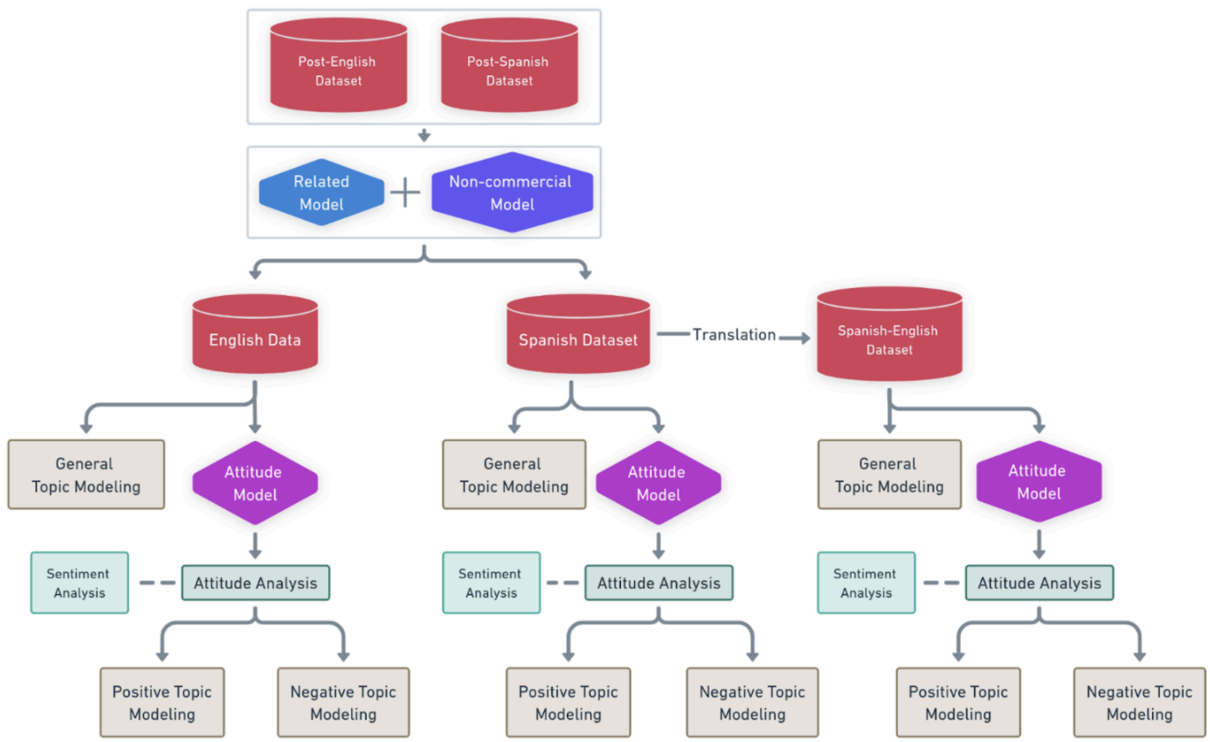
The core of the analytical structure involved topic modeling. A customized pipeline was developed, integrating advanced methodologies to extract coherent and interpretable topics. The process began with embedding generation using Sentence Transformers, followed by dimensionality reduction with UMAP. K-means clustering was then applied to group semantically similar text into distinct topics. Domain-specific stop words related to e-cigarettes and vaping were incorporated, and keywords irrelevant to the analysis were rigorously filtered out to ensure the relevance of the extracted topics.

Multiple models, including LDA, TF-IDF, and BERT Topic, were evaluated for their coherence and interpretability. While some models demonstrated higher coherence scores, the Bar Topic model was ultimately selected due to its superior balance between coherence, interpretability, and domain applicability. This decision was also aligned with discussions and recommendations from the project sponsor.

Topic Clustering and Interpretation

After identifying the optimal topic modeling approach, the number of topics was fine-tuned through iterative runs to achieve the highest coherence scores. The resulting themes were clustered and named using large language models, such as OpenAI's tools, to maintain objectivity and ensure semantic accuracy. Each identified theme underwent 100 iterations for name refinement to provide clarity and relevance.

The finalized set of 100 topics provided a robust foundation for further analysis and interpretation. This approach enabled the study to uncover nuanced themes within the data, paving the way for meaningful summarizations and actionable insights. The findings from this comprehensive analysis will be discussed in the subsequent sections.

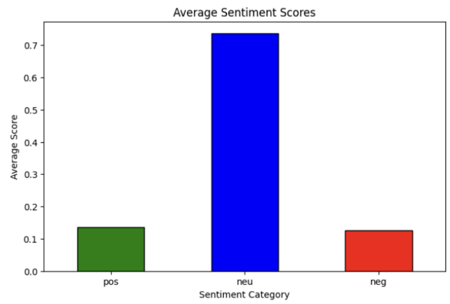


Findings

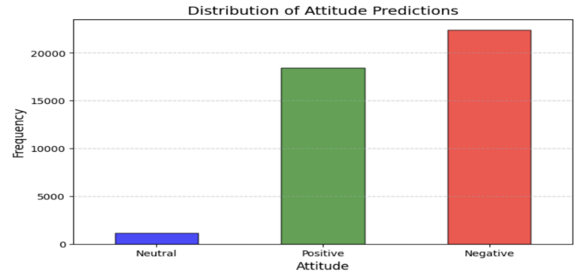
Sentiment vs. Attitude Analysis and Translation Assessment

Sentiment vs. Attitude Analysis

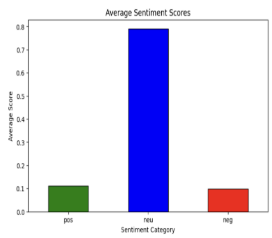
The comparative analysis of sentiment and attitude provided critical insights into the differences between these two dimensions across the dataset. On the left side of the visualization, sentiment analysis was conducted using the VADER sentiment analysis tool, which revealed a dominant neutral sentiment across the dataset. In contrast, positive and negative sentiments were relatively less prominent. In contrast, the right-hand visualization displays the results of the attitude prediction generated by the fine-tuned model. Unlike sentiment, attitude analysis revealed a stronger dominance of positive and negative classifications, with neutral attitudes notably underrepresented. These findings underscore a fundamental divergence between sentiment and attitude metrics, highlighting the nuanced insights that attitude analysis can offer compared to traditional sentiment analysis.



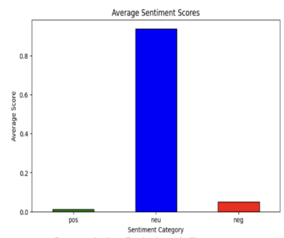
English Dataset (VADER)



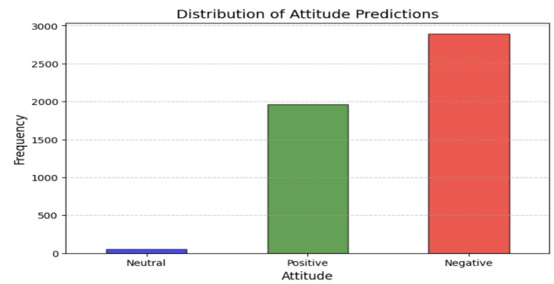
English Dataset (Finetuned)



Spanish Translated Dataset (VADER)



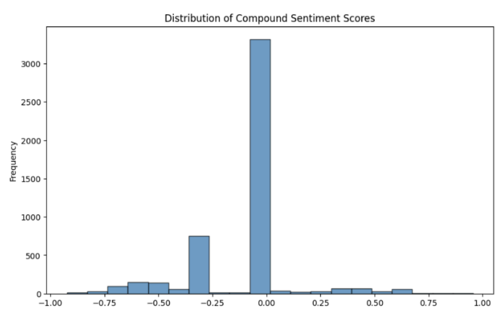
Spanish Original Dataset (VADER)



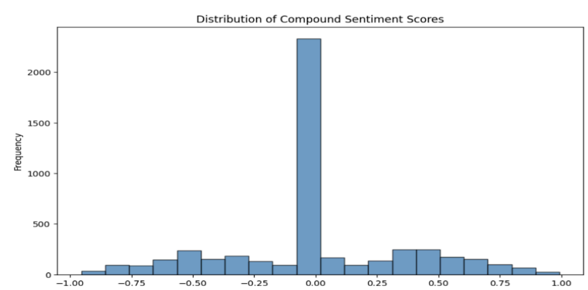
Spanish Translated Dataset (Finetuned)

Spanish Translation Evaluation

An additional analysis was performed to evaluate the potential impact of translating Spanish tweets into English. The compound scores from VADER sentiment analysis were compared for the original Spanish and English-translated datasets. The resulting distributions exhibited high similarity, indicating minimal information loss during translation. This consistency supports the robustness of the translation methodology and ensures that cross-linguistic comparisons within the study remain reliable and valid.



Spanish Dataset (Original)



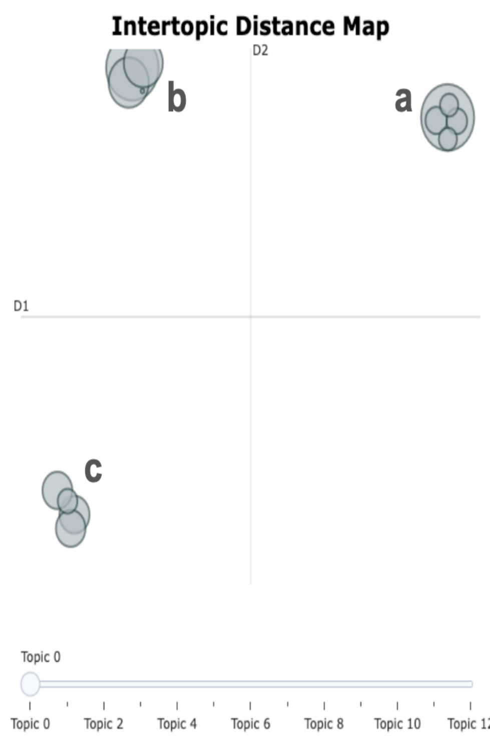
Spanish Dataset (Translated to English)

These analyses provide a deeper understanding of public discourse surrounding e-cigarettes while affirming the validity of cross-linguistic approaches to sentiment and attitude evaluations.

Topic Analysis Across Linguistic Groups

English Dataset: General Themes

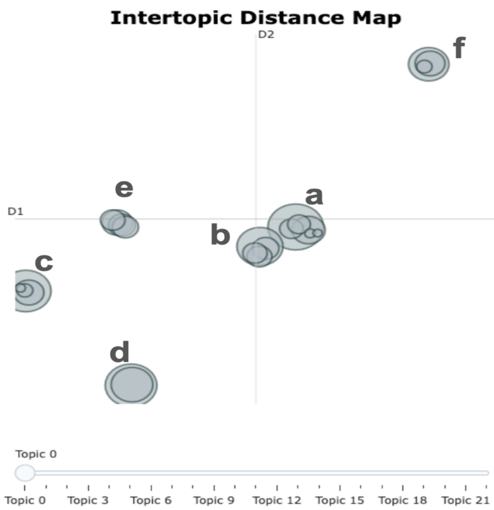
The topic modeling analysis of English-language tweets revealed three primary areas of focus. First, there was considerable discussion surrounding the **use and regulation of tobacco and nicotine**, highlighting concerns about policy enforcement and legal boundaries. Second, conversations delved into **youth culture associated with vaping**, particularly the popularity of vaping among younger demographics. Finally, the analysis identified behavioral concerns, explicitly addressing **school vaping-related issues**. Notably, the clustering of these discussions among English-speaking users was less distinct, suggesting a broader, more generalized style of conversation compared to the Spanish-language dataset.



Cluster	Theme
a	Tobacco and Nicotine Use and Regulation – encompasses the multifaceted nature of discussions addressing both the consumption patterns and health implications of tobacco and nicotine products, such as smoking, vaping, and their association with lung cancer and addiction.
b	Youth Culture and Vaping – the integration of vaping into the broader youth identity and contemporary trends — such as "vapes," "pods," and "Juul"—along with elements like "tiktok," "zodiac," "pronouns,
c	Youth Vaping and Behavioral Issues in Schools – the widespread issue of vaping among young students also along with "behavior," "crime," and "arrested," consistently references efforts by schools, parents, and the community through terms like "parents," "detectors," "help," and "quit,"

Spanish Dataset: General Themes

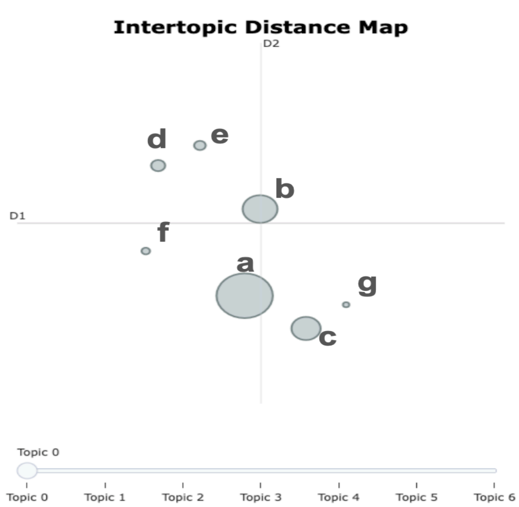
In contrast, the Spanish-language dataset revealed a richer diversity of themes. Key topics included **everyday life and habits**, reflecting the integration of vaping into routine activities and **vaping as part of a party lifestyle**, often discussed in the context of social gatherings. Additionally, significant emphasis was placed on **regulations and the health impacts of e-cigarettes**, underscoring concerns about societal and individual consequences. The discussions in the Spanish dataset were deeply rooted in cultural practices, highlighting aspects such as social interactions, family dynamics, and personal habits, which distinguished these conversations from their English counterparts.



Cluster	Theme
a	Shopping and Personal Preferences in Spanish-speaking Context
b	Everyday Life and Habits in Spanish-speaking Cultures
c	Regulation and Impact of Electronic Cigarettes – reinforced by frequent mentions of terms like regulation, law, prohibition, and legislation. And the significant concerns regarding health effects, especially among youth
d	Smoking and Tobacco Consumption – "fumar" (smoking), "cigarro" (cigarette), and health-related terms like "pulmones" (lungs), along with mentions of "dejar" (quit)
e	Spanish Language Social Media Conversations
f	Spanish Vaping and Party Lifestyle

Spanish-English Dataset: General Themes

The inter-topic distance map for the English dataset revealed distinct clusters of discussion within both positive and negative attitudes. Specifically, positive tweets formed four clusters, while negative tweets exhibited six. Upon summarization, these clusters consolidated into two overarching themes: **Behavior and Habit** and **Public Health and Policy**. The thematic analysis of subtopics revealed that while positive and negative tweets addressed similar themes, their perspectives and emphasis diverged significantly. Positive tweets leaned toward normalization and constructive solutions, whereas negative tweets focused on adverse consequences and regulatory lapses. This duality underscores the complexity of public discourse on vaping, highlighting the importance of addressing both constructive and critical viewpoints in public health campaigns and regulatory strategies.



Cluster	Theme
a	Vaping and Smoking Habits - behaviors, decisions, and social dynamics
b	Spanish Language and Vaping (General) - the repeated emphasis across explanations on the term "vapes"
c	Electronic Cigarettes and Public Health - consistently emphasizes the dual focus on health implications and regulatory
d	Hospitality and Culinary Experience
e	Effects of Smoking and Vaping on Health - Physical and psychological impact on an individual's health
f	Illicit Substance Use and Its Impact on Youth - the impact of illegal substance use, particularly among teenagers.

Comparative Insights

The comparative analysis of English and Spanish datasets reveals significant thematic differences, emphasizing the impact of cultural and linguistic contexts on public discourse surrounding e-cigarettes. English-speaking users predominantly focus on regulatory measures, public health implications, and youth behavioral concerns related to vaping. In contrast, Spanish-speaking users highlight social and cultural dimensions, including everyday habits, family dynamics, and lifestyle practices associated with vaping. These distinctions underscore the necessity for culturally tailored public health messaging and policy development approaches. Effectively addressing the unique concerns of diverse linguistic and cultural groups is critical for creating interventions and campaigns that resonate across different communities.

Aspect	English Dataset	Spanish Dataset	Translated Spanish Dataset
Focus	Broad societal trends and public health.	Lifestyle and culturally rooted themes.	Mix of global and cultural/local themes.
Health and Regulation	Universal focus on e-cigarette health risks and regulations.	Same, with localized regulatory references.	Same, with added cultural dimensions.
Youth Themes	Behavioral concerns, youth culture, and vaping in schools.	Youth vaping within celebratory and cultural contexts.	Youth vaping as part of social/cultural identity.
Cultural /Lifestyle	Minimal cultural/lifestyle emphasis.	Strong focus on lifestyle and cultural habits.	Balanced global and cultural perspective.
Social Media	Not prominent.	Strong focus on informal social media use.	Integrated into cultural discussions.

Subtopic Analysis of the Original Spanish Dataset

This section delves into the subtopic modeling of the original Spanish dataset, examining both positive and negative clusters to uncover thematic distinctions. The positive dataset comprises five clusters, whereas the negative dataset is divided into four clusters, each providing unique insights into public discourse on vaping.

The analysis of the positive dataset reveals themes centered around the integration of vaping and smoking into social and recreational contexts. These include: 1) **Social Gatherings and Recreational Activities**: Highlighting vaping as a component of communal experiences, emphasizing its role in youth culture and casual interactions. 2) **Party Culture**: Illustrating vaping's prevalence in celebratory settings and its association with lifestyle trends. 3) **Cravings and Consumer Preferences**: Reflecting on individual preferences and consumption habits related to vaping. 4) **Casual Indulgences and Personal Life**: Showcasing vaping's role in informal social dynamics and everyday routines. 5) **Health Impacts and Usage Patterns**:

Acknowledging vaping's perceived integration into personal identity and culture while downplaying risks.

In contrast, the negative dataset focuses on the detrimental aspects of vaping, including: 1) **Health-Related Consequences**: Emphasizing the risks and health implications of vaping, including addiction and respiratory issues. 2) **Public Health and Regulatory Concerns**: Addressing electronic cigarette use's societal and governmental challenges. 3) **Substance Use and Addiction**: Highlighting the issues of dependency and substance abuse, particularly within specific demographics. 4) **Regional Contexts in Mexico**: Underscoring unique concerns in Mexico, such as regulatory gaps and public health emergencies tied to vaping.

The thematic analysis highlights contrasting narratives between the positive and negative clusters. While positive discussions frame vaping as a lifestyle choice intertwined with social and recreational activities, negative discussions underscore its potential for harm, focusing on health risks and societal challenges. The emphasis on regional and cultural nuances, particularly in Mexico, reflects the need for context-specific interventions and policies. These findings underscore the importance of a nuanced approach to addressing vaping-related public health concerns in diverse communities.

Subtopic Analysis of Spanish-English Translated Dataset

Analyzing the Spanish-English translated dataset provided insights into the thematic clusters of positive and negative attitudes toward e-cigarettes. Three key clusters emerged after conducting topic modeling on the positive sentiment data. The first one is **Gift-Giving and Consumer Behavior**. This cluster highlights the intersection between gift-giving practices and consumer preferences for e-cigarettes. Discussions within this theme often focused on purchasing habits, product preferences, and the appeal of flavored products such as cherry-flavored e-cigarettes. Our findings reflect the role of e-cigarettes in social interactions and their value as consumer goods. Then, the **Consumer Lifestyle and Trends** cluster examines lifestyle choices where users perceive e-cigarettes as a healthier alternative to traditional cigarettes. Conversations within this theme underscore trends and preferences that position vaping as part of a modern, health-conscious lifestyle. The final cluster, **Smoking and Vaping as Social Cultures**, within the positive sentiment analysis centers on the cultural and social identity associated with vaping. Keywords such as "friends" and "social interactions" suggest that e-cigarettes are embedded in daily social life and reflect cultural norms and group identities.

In contrast, two primary clusters were identified within the negative sentiment dataset. Firstly, the **Regulation and Health Concerns Related to E-Cigarettes** focused on health implications, including concerns over chemical exposure, potential risks such as cancer, and general body health concerns. Users highlighted the need for stricter regulations to mitigate these health risks. Additionally, the cluster of **Smoking and Vaping Culture and Their Impact on Lifestyle Choices** explores the negative aspects of vaping culture, emphasizing issues such as peer pressure to adopt vaping behaviors within specific cultural groups. Additionally, vaping is viewed as a gateway behavior that reinforces unhealthy habits and lifestyles. This theme reflects a critique of the societal normalization of e-cigarette use.

The findings illustrate a dichotomy in public discourse surrounding e-cigarettes. While positive attitudes emphasize their role in enhancing social interactions and promoting alternative lifestyle choices, negative perspectives critique their health risks and cultural normalization. This contrast underscores the complexity of e-cigarette perceptions across different cultural and linguistic contexts. Further exploring these themes can provide actionable insights for health campaigns and policy interventions tailored to diverse communities.

Comparative Analysis of Spanish Original and Translated Datasets

The comparative analysis of the Spanish original dataset and its English-translated counterpart revealed thematic consistencies and significant differences. Shared themes across both datasets included discussions on vaping, smoking, social interactions involving e-cigarettes, and the associated health impacts. However, there were notable variations in emphasis and the introduction of new themes, reflecting the influence of translation on topic modeling. Both datasets discussed broad topics such as social interactions involving e-cigarettes, health consequences, and general vaping and smoking behaviors.

It was also revealed that there were notable themes between the Spanish original dataset and its English-translated counterpart, highlighting the challenges of preserving cultural nuances during translation. In the translated dataset, regulatory issues and new topics, such as gift-giving and specific product preferences, were introduced that were not present in the original Spanish dataset. Conversely, the original Spanish dataset focused more prominently on addiction and craving-related behaviors, emphasizing personal and cultural aspects of e-cigarette use. The translation process demonstrated limitations in capturing cultural nuances, including difficulties in interpreting idioms, slang, and expressions unique to Spanish. Additionally, themes such as addiction-related behaviors, which were evident in the original dataset, were diminished or absent in the translated version. At the same time, topics like gift-giving may not authentically reflect the original cultural discourse. Furthermore, specific details tied to cultural context, such as emoji usage and unique textual expressions, were often simplified or misrepresented, leading to a loss of depth and authenticity in the translated analysis.

	Spanish-English (Translated)	Spanish-Spanish (Original)
Similarities	Discuss vaping and smoking; Social interactions; Health impacts	
Difference	· **Emphasize more on regulations** · **Introduce new themes like gift-giving and specific product preferences** Different themes emerge between positive and negative sentiments, indicating variation in topics based on sentiment classification.	· **Lack of explicit discussion on regulation** ·**Focus more on addiction and craving** The clusters for positive and negative sentiments are identical, suggesting similar topics regardless of sentiment.
Implications	The translation procedure may not capture potential cultural nuances or loss of details; it may introduce more themes or shift the focus.	

Challenges in Cross-Linguistic E-Cigarette Discourse Analysis

Understanding public discourse on e-cigarettes across linguistic and cultural boundaries is a complex task that requires robust methodologies and careful interpretation. Despite the strengths of this study in capturing multilingual data from social media platforms, several challenges emerged, highlighting caveats in the analytical process, data representation, and implementation.

A primary challenge stemmed from **linguistic diversity**, as the translation process introduced limitations in accurately conveying cultural nuances and contextual expressions. Subtle meanings embedded in idiomatic phrases, slang, and culturally specific terms were often lost or misrepresented, resulting in gaps in understanding the authentic sentiment and themes of the original Spanish dataset.

Another challenge involved **topic modeling across languages**, where accurately detecting sentiment and clustering topics proved difficult due to the structural and semantic differences between English and Spanish. The diversity in language expressions posed challenges in creating consistent and unbiased topic clusters, highlighting the limitations of current modeling tools in cross-lingual contexts.

Demographic biases also played a significant role in shaping the findings. Social media users, whose opinions and behaviors formed the basis of the analysis, do not represent the broader population. These users may differ from the general public in terms of age, cultural background, and socioeconomic status, which limits the generalizability of the results and their relevance to broader public health initiatives.

Finally, the **disparity in data volume between English and Spanish datasets** created an imbalance that affected the comparative analysis. There was a significant difference in the volume of data between the English and Spanish datasets, with the English dataset containing over 10,000 tweets compared to around 5000 in Spanish. The significantly larger volume of English tweets likely overemphasized themes in the English dataset while underrepresenting critical insights from the Spanish dataset. This imbalance reduced the objectivity and comparability of the findings across the two linguistic groups.

Addressing these issues is essential for advancing the accuracy of cross-linguistic research and ensuring its relevance to diverse communities.

Future Work and Next Steps

Several critical areas of improvement have been proposed to address the challenges identified in this study and enhance the accuracy and depth of future analyses. First, advancing multilingual topic modeling methodologies is essential to capture better culturally and linguistically specific themes, minimize information loss during translation, and ensure reliable cross-linguistic comparisons. Second, creating proportionally balanced datasets will address disparities in data volume between English and Spanish tweets, ensuring equal representation and enhancing the objectivity of cross-language analyses. Finally, longitudinal and temporal analyses are critical for examining the evolution of public discourse over time, particularly in response to policy changes, public health campaigns, and significant events related to e-cigarettes. These steps collectively aim to refine analytical methods, provide more nuanced insights, and support the development of culturally sensitive public health strategies.

References:

[1] Centers for Disease Control and Prevention (CDC). (2021). Youth Tobacco Use: Results from the National Youth Tobacco Survey.

[2] Wang, T. W., Neff, L. J., Park-Lee, E., Ren, C., Cullen, K. A., & King, B. A. (2020). E-cigarette use among middle and high school students — United States, 2020. Morbidity and Mortality Weekly Report (MMWR).

[3] Pew Research Center. (2020). Social media use in 2020: Platforms and demographics.

[4] Desjardins, R. (2017). Translation and Social Media: In Theory, in Training and in Professional Practice. Palgrave Macmillan.