

NEW MEDIA, NEW NARRATIVE: LEXICAL CHOICES AND LANGUAGE RESISTANCE IN THE POST-1975 VIETNAMESE DIASPORA

TRUYỀN THÔNG MỚI, DIỄN NGÔN MỚI: SỰ LỰA CHỌN TỪ VỰNG VÀ SỰ KHÁNG CỰ NGÔN NGỮ TRONG CỘNG ĐỒNG NGƯỜI VIỆT HẢI NGOẠI SAU 1975

Andrea Hoa Pham^{1,*}, Andrew T. Pham²

DOI: <http://doi.org/10.57001/huih5804.2025.313>

ABSTRACT

A common assertion among overseas Vietnamese communities that since the country's reunification in 1975, vocabulary from North Vietnam ("Northern words") has replaced terms used in South Vietnam ("Southern words"). This study presents the first empirical test of this claim using a corpus-based approach. We analyzed word frequencies in four major online newspapers - two published in Vietnam and two by diaspora communities in the United States - drawing from a published, community-sourced list of contested vocabulary. Our findings confirm that Northern words appear significantly more frequently in Vietnam-based news sources. However, contrary to the narrative of linguistic resistance, there was no significant difference in the overall usage of Southern words between the two corpora. Moreover, the data reveal a counter-intuitive pattern where some Southern words are used *less* frequently than their Northern equivalents, even in the US-published newspapers.

Keywords: Vietnamese diaspora, lexical choice, language resistance, corpus linguistics, language use, online media.

TÓM TẮT

Một nhận định quen thuộc trong cộng đồng người Việt ở Hoa Kỳ là trong tiếng Việt sau năm 1975, một số từ ngữ được cho là chỉ phổ biến từ Quảng Bình trở ra bắc (nhóm "từ Bắc") đã thay thế những từ trước đó thường dùng ở miền Nam (nhóm "từ Nam"). Quan niệm này được nói nhiều trên báo chí, song chưa có một nghiên cứu nào khảo sát một cách khách quan. Nghiên cứu này là phân tích thực nghiệm đầu tiên, dùng phương pháp ngữ liệu thống kê để đánh giá về nhận định trên dựa trên tần số xuất hiện của từ trên văn bản báo chí. Tư liệu được thu thập từ bốn tờ báo lớn, hai tờ của chính quyền xuất bản ở Việt Nam và hai tờ của cộng đồng người Việt xuất bản ở Hoa Kỳ. Kết quả ban đầu cho thấy như dự đoán, các "từ Bắc" xuất hiện thường xuyên hơn nhiều trên báo chí ở Việt Nam. Ngược lại với nhận định trên, không có chứng cứ cho thấy sự khác nhau ở tần số xuất hiện của các "từ Nam" trên báo chí ở Việt Nam cũng như ở Hoa Kỳ, nơi được cho là có kháng cự ngôn ngữ. Thậm chí trên báo chí ở Hoa Kỳ, một số từ ngữ xuất hiện theo chiều hướng không chờ đợi: "từ Bắc" nhưng được dùng nhiều hơn "từ Nam".

Từ khoá: Tiếng Việt ở Hoa Kỳ, chọn lựa từ ngữ, kháng cự ngôn ngữ, ngôn ngữ học dữ liệu, sử dụng ngôn ngữ.

¹University of Florida, USA

²Mississippi University for Women, USA

*Email: apham@ufl.edu

Received: 18/6/2025

Revised: 25/8/2025

Accepted: 28/8/2025

1. INTRODUCTION

Language is a powerful tool for shaping opinions, constructing social identity, and defining reality.

Language and social-cultural identity are inseparable (e.g., Ige, Talbot et al. [1, 2]. As Weedon [3] states, "Language is also the place where our sense of our self

and our subjectivity is constructed". The use of particular words can foster a sense of belonging within a cultural community, reflecting how a speaker positions themselves and their views on current issues. In this capacity, word choice is constantly at play. As Wardhaugh [4] notes "How we say something is as important as what we say" (p. 274). For example, in the United States of America, the choice between the term "undocumented immigrants" and "illegal immigrants" often reveals the speaker's stance on people entering the country without legal authorization. When a social role or cultural identity is imposed on individuals or groups through language use, this act is often met with constant negotiation, contestation, and even resistance [2, 5, 6]. This paper will explore such linguistic resistance within the Vietnamese diaspora in the United States.

After the fall of US-allied Republic of Vietnam (South Vietnam) to the forces of the Democratic Republic of Vietnam (North Vietnam) in 1975, large waves of people from the South fled the country with most ultimately resettling in the United States. A significant body of scholarship has since examined the Vietnamese American experience, focusing on theme of war, loss, family separation, displacement, and the construction of new identities in a new homeland [7, 8].

Today, with over 1.5 million speakers, Vietnamese is the sixth most common non-English languages spoken in the United States (United States Census Bureau) [9]. To preserve their language and cultural identity for future generations, many Vietnamese communities organize language classes for children. These programs are often held after school or on Sundays in public schools and religious facilities such as Buddhist temples and churches. This commitment to language preservation extends to higher education, where the number of Vietnamese language courses offered at US colleges has grown steadily over several decades, with programs now available at more than 40 institutions.

In the decades following 1975, a linguistic divergence emerged between the Vietnamese spoken in Vietnam and in the diaspora. This was driven by two parallel processes: in the United States, the language developed independently in multilingual settings, largely disconnected from media in Vietnam. Concurrently, the language within Vietnam continued to evolve. Due to this disconnection from the homeland, the Vietnamese used and taught in the US is often described as a "frozen" form that is also influenced by English. Within post-

reunification Vietnam, a set of vocabulary prevalent in the North became standard. These words were new and unfamiliar to many southerners, including those who had fled the country. Because these terms were associated with the North, they were often labeled "communist words" by southern-identified people in the diaspora. Examples include the Northern term *trung bình* replacing the Southern *bình quân* (S.) for 'average'; *cách ly* for *cô lập* (S.) 'isolation'; *đăng ký* for *ghi danh* (S.) 'to register'; and *phần mềm* for *nhu liệu* (S.) 'software'. One notable example is *hoành tráng*, which replaced several Southern words for 'grand' or 'magnificent', e.g., *thịnh soạn*, *linh đình*, *trọng thể* [10-12].

The following paragraph illustrates this division between what are often termed "overseas Vietnamese words" (i.e., non-communist) and "words used in Vietnam" (i.e., communist). The specific words in question are presented in bold type.

"*Chẳng hạn như trong các cộng đồng người Việt ở Mỹ, chúng ta có thể đọc câu này trên báo chí địa phương tiếng Việt: "Để biết thêm **chi tiết** về **sinh hoạt** giới thiệu một số **nhà vẽ kiểu** thời trang, xin **liên lạc** qua số điện thoại ... và **ghi danh** tham dự."* Cùng nội dung đó, một tờ báo trong nước chắc sẽ viết như thế này: "Để biết thêm **thông tin** về **sự kiện** giới thiệu một số **nhà thiết kế** thời trang, xin **liên hệ** qua số điện thoại ...] và **đăng ký** tham dự." (Trần) [12] ("For example, in Vietnamese communities in the United States, we might see this sentence in the local Vietnamese press: "For further **details** about the **activities** introducing some fashion **designers**, please **contact** the phone number ... and **register** to attend." With the same content, a newspaper in Vietnam would probably write like this: "For further **information** about the **event** introducing some fashion **designers**, please **contact** the phone number ... and **register** to attend." (English translation does not show differences between the northern and southern words).

A prevalent assumption has emerged that many words used in South Vietnam before 1975 have been replaced by vocabulary from the North. Within diaspora communities, these Northern terms are often perceived as incorrect, impoverished, overly Sinicized, or politicized. Consequently, it is believed that Southern-identifying Vietnamese people actively resist using these so-called "communist" words. This paper uses a corpus-based approach to empirically test the validity of this assumption.

In general, Vietnamese American political leaders are not directly involved in language politics; maintenance of the language occurs mainly at home and within community institutions (Dao & Bankston) [13]. However, within large Vietnamese American communities, language instructors must carefully select "appropriate" teaching materials, as parents often express concern about their children learning "communist terminology" or being exposed to "inappropriate content." As a result, many community language schools avoid using textbooks produced in Vietnam, and the use of such materials has, in some cases, led to community boycotts. A notable example occurred in Westminster, California. After the mayor discovered that a textbook used in his daughter's class, *Let's Speak Vietnamese* by Lê Phạm Thuý Kim, contained what he considered "communist" words and symbols, a local controversy erupted. The objectionable content included Northern vocabulary, such as *đăng ký* ('to register') instead of the Southern *ghi danh*, as well as an image of the current Vietnamese flag [12]. After the mayor voiced his concerns to district officials about the textbook's "inappropriate" and "irrelevant" content, the Westminster School District removed the book from its language classes the very next day. The first author of this paper encountered a similar situation involving the same textbook, specifically over its use of the Northern term *hộ chiếu* ('passport') instead of the Southern equivalent, *thông hành* (Pham) [14].

The view that Northern words have replaced "more proper" Southern equivalents in Vietnam, thereby impoverishing the language, is still widely held today. While acknowledging the natural evolution of a living language, critics of "improper" word usage tend to focus exclusively on the variety of Vietnamese spoken in Vietnam (Kính Hoà, Huỳnh) [15, 16]. For instance, in an interview with Radio Free Asia (RFA), when overseas Vietnamese journalist Trần Phong Vũ was asked about the linguistic differences between Vietnam and the diaspora, he replied:

Nhưng ở Việt nam ngày nay thì có những cái chữ như là phản lại ngôn ngữ. Có những chữ đã có nghĩa nguyên gốc của nó rồi, tôi lấy ví dụ như cái chữ vô tư, chữ này trên căn bản của ngôn ngữ Việt nam là nói đến một người vô tư lự, một người không quan tâm lo lắng gì cả, một thái độ hồn nhiên, thì bây giờ chữ vô tư được dùng như một thái độ tự nhiên. Ví dụ như người ta mời dùng bữa thì lại nói là các anh cứ vô tư đi! Tức là muốn nói các anh cứ dùng bữa tự nhiên. Như vậy dùng chữ như thế nó lạc hết, bởi vì nó mất cái tiếng

gốc của nó. Chưa kể là những nguy hiểm của những chữ nó đi ngược lại với đạo lý của mình, với truyền thống khiêm tốn của mình, toàn là những từ đao to búa lớn mà chẳng có ý nghĩa gì cả, mà bây giờ đang phổ biến rộng rãi. Vì vậy tôi sợ rằng một ngày nào đó tiếng Việt của chúng ta đi xa quá nếu không có sự quan tâm. (But in Vietnam today, there are words that seem to go against the language. There are words that already have their original meaning, for example, the word "vô tư", the basic meaning of the word in the Vietnamese language refers to a carefree person, a person who does not care about anything, a carefree attitude, but now the word "vô tư" is used as a natural attitude. For example, when someone invites you to a meal, you say "Các anh cứ vô tư đi" (You guys just do not care about anything)! That means you can just eat naturally. Using words like that is completely out of place, because it loses its original meaning. Not to mention the dangers of words that go against our morality, our tradition of modesty, all of which are big words that have no meaning but are now widely used. Therefore, I am afraid that one day our Vietnamese language will go too far if we do not pay attention.) [15].

The dynamics of negotiation and resistance in language use among Vietnamese people in North America have been explored by scholars such as Pierce [6]. More specifically, researchers like Lam [17] and Pham [14, 18] have discussed the stigma that equates the Northern dialects of Vietnamese language teachers with "communism." This tension is particularly evident in heritage language classrooms. When instructors speak a contemporary, post-1975 Northern dialect, students from Central or Southern family backgrounds are often advised to adopt the teacher's dialect and vocabulary from the selected textbook, a pedagogical approach that can result in parental resentment. The two examples below, from Lam [17], illustrate this stigma.

"... a heritage student of Da Nang central Vietnamese family background takes a Beginning Vietnamese course with an instructor of Hanoi northern-accented Vietnamese whose family migrated to South Vietnam in 1954. This student returns to her family during the summer break speaking the pre-1975 northern language form adopted in the classroom, much to the dismay, criticism and alienation of her primarily central-accented speaking family (the first of her primary speech communities). She questions whether this language instruction will actually help her to build stronger bonds with her immediate and extended family if the family

feels that her education will further repress her regional cultural heritage, and she chooses not to continue with Intermediate Vietnamese the following academic year" (Lam, 8-9) [17].

In the second case, a heritage student of a southern refugee family, in a Beginning Vietnamese class uses the southern words *viết* 'a pen' and *mướn* 'to rent' instead of the northern equivalents, *bút* and *thuê*, respectively. For that, "...he is chastised for using "improper," "outdated," or at worst "backward" diction. With much shame and hesitation, the student makes the necessary adjustments throughout the year, and returns in the summer to a family for whom his new language unearths and further inflicts the trauma of exile and loss they have agonized to overcome. A family upset by this dramatic change in their child's intellectual and academic development and unfamiliar with the possible arbitrariness of the U.S. educational bureaucracy and policies questions the college curriculum and leadership with its seemingly "communist" imposition upon their generation's youth and instructs the student not to return to Intermediate Vietnamese the following year. (Lam, 9) [17].

However, despite the persistence of this belief, the claim that "communist words" (Northern vocabulary) are replacing "non-communist words" (Southern vocabulary) has never been systematically studied. To date, no data-driven research has been conducted to confirm or challenge this assumption, particularly as it pertains to Vietnamese media. While the related issue of "outdated" Southern words also lacks empirical analysis, that topic is beyond the scope of this paper.

This study, therefore, presents the first empirical investigation into this long-standing claim. Using a corpus of online newspapers from both Vietnam and the United States, we statistically examine the usage patterns of contested "Northern" and "Southern" words to test the validity of this assumption.

2. METHODOLOGY

To test the claim about replacement of "southern" words with "northern" words, we formulate two hypotheses and employ a quantitative, corpus-based approach. This method involves comparing the frequency of specific words in Vietnamese newspapers published in Vietnam versus those published in the United States. Our analysis is based on a curated word list that categorizes equivalent vocabulary into "Northern" and "Southern" groups. Obtaining such a word list

requires a source that explicitly distinguishes between these two regional variations. This distinction is a socio-political one, made almost exclusively by Vietnamese speakers who self-identify as Southern and non-communist. Because this group's perspective is not prevalent in mainstream Vietnamese sources, documentation of these vocabulary differences is almost non-existent. Therefore, this study uses a word list compiled by Trần Văn Giang that was published in the US-based magazine *Đặc San Lâm Viên* (ĐSLV, *Lâm Viên's* Special Issue) [10].

Throughout this paper, "Northern words" refers to vocabulary primarily associated with usage in North Vietnam, while "Southern words" refers to vocabulary primarily associated with usage in South Vietnam before 1975.

Based on the central claim, we formulated the following two hypotheses:

• **Hypothesis 1 (H1):** Northern words on the list will appear with significantly greater frequency in newspapers published in Vietnam than in those published in the United States.

• **Hypothesis 2 (H2):** Southern words on the list will appear with significantly greater frequency in newspapers published in the United States than in those published in Vietnam.

The lexicon examined in this study includes words such as *bình quân* 'average', *cách ly* 'to separate', *chất lượng* 'quality', and *chủ trì* 'to chair', among others. For example, on the list, *bình quân* is labeled as a "Northern word", and *trung bình* a "Southern word" because *trung bình* was said to be common in South Vietnam before 1975.

To construct the two corpora for this study, articles were sourced from prominent newspapers in Vietnam and the United States. The Vietnam-based corpus was drawn from two major state-run news outlets, *Nhân Dân* (The People) [19] and *Thông Tấn xã Việt Nam* (Vietnam News Agency) [20]. The US-based corpus was compiled from two leading diaspora newspapers, *Người Việt* [21] and *Cali Today* [22], which serve the large Vietnamese American population. Established in 1978, *Người Việt* is the oldest and largest continuously operating Vietnamese-language newspaper in the United States. *Cali Today* (or *Nhat Bao Calitoday*), founded in 1994, is a media and publishing company based in San Jose, California, that serves the Vietnamese community. The

data collection was restricted to articles published mainly in 2024 to ensure contemporaneity and comparability. The selection process was guided by specific criteria: articles were required to be original, human-authored content intended for a general audience, consistent word counts, and written in a standard journalistic register. To maintain consistency within the corpora, texts of extreme length (e.g., very short posts or extensive reports) and those with excessive slang were excluded.

To ensure a breadth of topics, articles were selected from four categories: Economics, Education, Technology, and Sports. We aimed for a roughly equivalent number of articles from each category for both the Vietnam-based and US-based corpora. We noted that the US-published newspapers in our sample tend to feature shorter articles than their Vietnam-published counterparts. To achieve comparable total word counts, we selected more articles from US-published sources, as they are typically shorter (550-1,000 words) than their Vietnam-published counterparts (1,000-1,600 words).

For data analysis, articles from each country (Vietnam-published vs. USA-published) were combined into two distinct corpora. The Vietnam-publishing corpus comprised 45 articles totaling 48,436 words, while the USA-published corpus comprised 70 articles totaling 48,434 words.

The analysis uses two parallel word lists: one containing Northern terms and the other their corresponding Southern equivalents. These were derived from a master list published in *Đặc San Lâm Viên* (ĐSLV), which contained over 300 Northern/Southern word pairs. To ensure a clear one-to-one correspondence for our quantitative analysis, we excluded any pairs where a single term was associated with multiple equivalents. This filtering process resulted in a final set of 180 distinct word pairs. From these pairs, we constructed the two lists used in our study: a 180-item Northern word list and a corresponding 180-item Southern word list, with care taken to ensure no duplicates existed within each list. The complete word lists are provided in the Appendix.

3. FINDINGS AND ANALYSES

Each word in a word list (Northern or Southern) was treated as a separate observation, and each word had a score on 2 variables: number of appearances in the 48,436-word Vietnam news group, and number of appearances in the 48,434-word USA news group. A two-

tailed paired-samples t-test was conducted to compare differences between these two variables: the rate of appearance of Northern words in Vietnam-published news articles vs. the rate of appearance of Northern words in USA-published news articles. Results showed that Northern words appeared more frequently in Vietnam-published ($M = 4.72$) news articles than in USA-published ($M = 2.67$) news articles, $t(179) = 3.06$, $p < 0.01$, $d = 0.23$. In other words, Northern words were found at an average frequency of 4.72 in Vietnamese news articles, significantly more often than the 2.67 in USA-published news articles.

A similar test was conducted, but with Southern words instead of Northern words. Results showed no evidence of a difference in how often Southern words appeared in Vietnam-published ($M = 6.28$) or USA-published ($M = 6.38$) news articles, $t(179) = -0.10$, $p = 0.921$, $d = 0.0074$. In other words, Southern words were found at an average frequency of 6.28 in Vietnam-published news articles and 6.38 in USA-published articles, a non-significant difference.

It is important to note that most word tokens in both corpora consist of high-frequency grammatical words such as *và* 'and', *các* 'plurality marker', *của* 'of', *trong* 'in', *là* 'to be', *với* 'with', *được* 'able to/in favor', and *cho* 'to, for'. These function words, which constitute the bulk of any text, are not specific to either the Northern or Southern lexicons. The frequency differences we observed for our target words are, therefore, present despite being "diluted" by this large volume of neutral grammatical words. We can infer that an analysis focusing only on the contested Northern vs. Southern vocabulary - thereby reducing this statistical "noise" - would likely yield an even larger effect size and a more pronounced difference than the one reported above.

4. DISCUSSION

While the statistical findings confirm Hypothesis 1, a closer examination of the data reveals several nuances worth discussing. First, although the Northern word list as a group was significantly more frequent in the Vietnam corpus, it is notable that many individual Northern words also appeared in the US corpus. Table 1 presents the frequencies of occurrence for several of these words. This cross-usage suggests that journalists for US-published newspapers do not exclusively use Southern vocabulary and, in some cases, still employ the Northern equivalents, e.g., items b-g in Table 1.

Table 1. Frequencies of Northern Words in the Vietnam and US Corpora

Northern words	News in VN	News in USA	Gloss
a. công trình	10	5	construction
b. động viên	19	10	to encourage
c. hiện đại	9	6	modern
d. khả năng	17	14	ability
e. phí	75	52	fee
f. quá trình	23	12	progress, process
g. quan chức	11	7	official
h. quy hoạch	30	3	zoning
i. tư vấn	6	3	to consult
g. vận động viên	18	7	athlete
k. xuất khẩu	15	4	to export

Second, the data revealed that not all Northern words followed the general trend. Some appeared with similar frequency across both corpora. For instance, several words showed near-equal usage: *diện* 'category' (Vietnam corpus: 34, US corpus: 36), *huyện* 'district' (Vietnam: 45, US: 41), and *vận hành* 'operate' (Vietnam: 5, US: 5).

More surprisingly, and contrary to our hypothesis, a subset of Northern words occurred even more frequently in the US-published corpus than in the Vietnam-published one, as shown in Table 2. Given that these words are from the Northern list, the expectation was for their frequencies to be higher in the Vietnam corpus. For this subset of words, however, the opposite was observed.

Thus, while the overall trend confirmed that Northern words are used significantly less in the US diaspora press ($M = 2.67$) than in Vietnam's domestic press ($M = 4.72$), this aggregate finding conceals the countertrend of specific lexical items. This was particularly evident for the words listed in Table 2, items c, e, f, g, j, k, l, and m.

Table 2. Frequencies of Northern Words Used More Often in the US Corpus

Northern words	News in VN	News in USA	Gloss
a. khắc phục	6	8	to fix
b. lợi nhuận	9	11	profit
c. quy trình	3	7	procedure
d. sân bay	3	5	airport
e. sự cố	0	12	event
f. thể hiện	2	5	to show
g. thi công	1	7	construction

h. thị phần	0	2	market share
i. thiết bị	3	5	device
j. tiên tiến	4	8	advanced
k. trải nghiệm	0	4	to experience
l. trí tuệ	2	7	wisdom
m. truy cập	0	7	to access

Turning to Hypothesis 2, we expected to find that Southern words occur more frequently in US-published newspapers than in their Vietnam-published counterparts. However, a paired-samples t-test revealed no statistically significant difference in the overall frequency of Southern words between the US corpus and the Vietnam corpus ($p = 0.921$).

While Hypothesis 2 was not supported by the aggregate data, a closer look at individual word usage reveals several interesting patterns. First, consistent with our initial expectation, a number of individual Southern words did appear more frequently in the US corpus than in the Vietnam corpus (see Table 3). This finding provides limited evidence for linguistic resistance, suggesting that while the overall effect is not statistically significant, diaspora journalists may selectively favor certain Southern variants.

Table 3. Frequencies of Southern Words Used More Often in the US Corpus

Southern words	News in VN	News in USA	Gloss
rẻ (N. bèo)	35	50	cheap
tài liệu (N. tư liệu)	3	9	data
kinh nghiệm (N. trải nghiệm)	7	11	experience
điều hành (N. vận hành)	6	10	operating
vấn đề (N. vấn nạn)	20	53	issue

Second, and most surprisingly, the data revealed a pattern directly contrary to our hypothesis: several words from the Southern list occurred less frequently in the US-published corpus than in the Vietnam-published corpus (see Table 4).

Table 4. Frequencies of Southern Words Used More Often in the Vietnam Corpus

Southern words	News in USA	News in VN	Gloss
a. bác sĩ (N. bác sỹ)	25	43	Medical doctor
b. quận (N. huyện)	13	20	district
c. dự án (N. đề án)	17	83	project
d. giá (N. phí)	151	234	fee
e. thực hiện (N. thể hiện)	37	86	to perform

f. <i>thị trường</i> (N. <i>thị phần</i>)	15	43	market
g. <i>xuất sắc</i> (N. <i>tiên tiến</i>)	5	11	excellent
h. <i>kế hoạch</i> (N. <i>quy hoạch</i>)	21	27	plan
i. <i>kiến thức</i> (N. <i>trí tuệ</i>)	1	4	knowledge
j. <i>trung bình</i> (N. <i>bình quân</i>)	11	13	average

This finding indicates that, paradoxically, some words traditionally classified as "Southern" now exhibit more widespread usage in Vietnam's domestic press than in the diaspora press. This counter-intuitive pattern was particularly notable for the words corresponding to items a, c, d, e, f in Table 4.

In summary, while our findings confirm that Northern words as a group are more prevalent in the Vietnam corpus, the key takeaway is the significant overlap in vocabulary use. Both Northern and Southern words appear in the corpora from both countries, albeit at varying frequencies. Most notably, many words identified as "Southern" were not only present in the Vietnam corpus but sometimes appeared there more frequently than in the US corpus.

This suggests a potential blurring of linguistic boundaries in the digital age. Increased access and exposure to each other's media may be eroding the once-clear distinctions between these regional vocabularies for both overseas Vietnamese and those in the homeland. If this trend continues, advocacy for a "pure, non-communist" lexicon may become less relevant over time. Consequently, the pedagogical tensions imposed on Vietnamese American heritage students, such as those described by Lam [17], may also begin to fade.

5. CONCLUSION

The common claim that "Northern words" are replacing "Southern words" in the Vietnamese language appears to be based on intuition rather than empirical data. Our study provides the first data-driven analysis of this claim. We found that, as expected, Northern words are significantly more frequent in newspapers published in Vietnam than in those published in the United States. However, our analysis revealed no statistically significant difference in the overall usage of Southern words between the two corpora. Furthermore, these aggregate trends conceal complex patterns where certain words defy expectations, highlighting the nuances of real-world language use.

While we believe the large and varied corpora used in this study help mitigate potential selection bias, our

research has limitations. First, the primary limitation stems from the source word list, as the inclusion criteria used by *Đặc San Lâm Viên* are unclear. For instance, some pairs represent minor orthographic preferences rather than lexical differences (e.g., *bác sĩ* (S) vs. *bác sỹ* (N) for 'medical doctor'). Other pairs problematically compare slang to standard vocabulary; the list includes the Northern slang term *bèo* ('cheap'), which did not appear in our data, while the standard term *rẻ* was prevalent in both corpora (see Table 3). Furthermore, many so-called "Northern words" such as *đề xuất* 'to suggest', *hiện đại* 'modern', *hữu nghị* 'friendship' appear in dictionaries published in Saigon (e.g., *Việt-Nam tự-điển* by Lê Văn Đức, 1970) and vice versa. While such inconsistencies weaken the source's credibility, the list from *Đặc San Lâm Viên* is the only US-based published source of its kind we could identify that categorizes words as "communist" or "non-communist". To minimize bias in our article selection, we implemented a two-person system. The first selector is fluent in Vietnamese and possesses a deep understanding of the cultural context. The second selector (the second author) is not fluent in Vietnamese and unaware of potential words which might be labeled as "communist", thereby preventing hidden political bias from influencing the final selection of articles.

Regardless of the drawbacks of the word list from *Đặc San Lâm Viên*, the practice of categorizing words into "communist" and "non-communist" persists among overseas Vietnamese. This sentiment is reflected in the cited articles and observed on social media both within and outside the United States. The recency of data, with an extensive selection of articles from 2024, underscores the contemporary relevance of the linguistic phenomena and community beliefs central in this study. Our purpose is to add a quantitative perspective to what has been an overwhelmingly qualitative discussion. While our results do not settle the argument, we hope they will stimulate future research and offer deeper insight.

To address the issues of word list validity, future research will aim to develop a new list based on participant ratings. Such a tool, created with clear and consistent criteria, would allow for an even more precise examination of lexical variation and resistance in the Vietnamese diaspora.

ACKNOWLEDGEMENTS

We are grateful to an anonymous reviewer for their insightful feedback. We alone are responsible for any errors.

APPENDIX (from *Đặc San Lâm Viên*)**NORTHERN GROUP**

Bắc bộ
Trung bộ
Nam bộ
bác sỹ
ca sỹ
bài nói
bang
mũ bảo hiểm
Bảo tàng
bèo
biên chế
biến thái

biểu diễn
bình quân
bo mạch

cá cược
cách ly
cái a lô
căn hộ
căng
cầu lông
chất lượng
chế độ
chiêu đãi
chỉnh chu
chủ đạo
chủ trì
chữa cháy
chức năng
chui
chứng minh nhân dân
cơ bản
cơ bắp
co cụm
cơ trưởng
cơ phó

SOUTHERN GROUP

Bắc phần
Trung phần
Nam phần
bác sĩ
ca sĩ
diễn văn
tiểu bang
mũ an toàn
viện bảo tàng
rẻ
chính ngạch
suy nghĩ hành động
lệch lạc trái với bình thường
trình diễn
trung bình
mạch điện chính trong computer
đánh cá
cô lập
cái điện thoại
căn nhà
căng thẳng
vũ cầu
phẩm chất tốt
quy chế
thiết đãi
chỉnh tề và chu đáo
chính ngạch
chủ tọa
cứu hoả
khả năng và nhiệm vụ
lén lút
thẻ căn cước
căn bản
bắp thịt
thu hẹp
phi công chính
phi công phụ

công đoàn

công nghiệp

công trình

cụm từ

cứu hộ

diện

diễu binh

diễu hành

dũng cảm

đẳng cấp

đào tị

để án

để xuất

đi làm suốt

điểm nhấn

đỉnh

đội ngũ

đội tàu

động viên

đột xuất

đường băng

đường cao tốc

gia công

giá hữu nghị

giá mềm

giảm tốc

giản đơn

giao dịch

giáo trình

hạch toán

hải quan

hâm

từng

hàng không dân dụng

hát đôi

hát tốp

hạt nhân

hậu cần

hệ quả

hiện đại

nghị viện

kỹ nghệ

công tác

nhóm chữ

cứu cấp

thành phần

diễn binh

diễn hành

mạnh mẽ

giai cấp

tị nạn

dự án

để nghị

đi làm suốt ngày

điểm nổi bật

hay

hàng ngũ

hạm đội

khuyến khích

bất ngờ

phi đạo

xa lộ

làm ăn công

giá tượng trưng

giá rẻ

giảm tốc độ

đơn giản

thương thảo

hương trình giảng dạy

kế toán

quan thuế

khùng

mát giây

hàng không dân sự

song ca

hợp ca

nguyên tử

tiếp liệu

hậu quả

tối tân

hộ chiếu	sổ thông hành	phản biện	phản đối
visa	giấy chiếu khán	phản cảm	gây khó chịu
hộ khẩu	tờ khai gia đình	phản cứng	cương liệu
hộ lý	dâm nô	phản mềm	nhu liệu
hồ hỏi	phấn khởi	phát sóng	phát thanh
học vị	bằng cấp	phí	giá
hội chữ thập đỏ hội	Hồng thập tự	phiên bản	bản
hữu hảo	tốt đẹp	Phó tiến sĩ	cao học
hữu nghị	thân hữu	phục hồi nhân phẩm	hoàn lương
huyện	quận	quá độ	qua khỏi thời gian cũ để chuyển sang thời gian mới
kênh	băng tần		
khả năng	có thể xảy ra	quá trình	những điều đã xảy ra theo một trình tự
khắc phục	tìm cách giải quyết		
khẩu trang	miếng vải che mặt	quan chức	nhân viên có thẩm quyền
khủng	rất lớn		
kiểu hối	ngoại tệ	quân hàm	cấp bậc
kinh qua	trải qua	quán triệt	hiểu rõ
làm gái	làm điểm	quảng trường	công trường
lầu năm góc	Ngũ giác đài	quy hoạch	kế hoạch
nhà trắng	toà Bạch ốc	quy trình	tiến trình
lễ tân	ngghi lễ	sân bay	phi trường
Liên Hợp quốc	Liên Hiệp quốc	sơ tán	tản cư
lính gái	nữ quân nhân	sự cố	trở ngại
lính thuỷ đánh bộ	thuỷ quân lục chiến	sư	sư đoàn
linh tinh	vớ vẩn	sức khoẻ công dân	y tế công cộng
lợi nhuận	lợi tức	tài chủ nước lạ	Tàu cộng xâm lăng
lược tóm	tóm lược	tập đoàn	công ty
Mĩ	Hoa kỳ	doanh nghiệp	tổ hợp
minh hoạ	tượng trưng	tàu vũ trụ	phi thuyền
nghệ danh	tên dùng ngoài tên thật	tên lửa	hoả tiễn
nghĩa vụ quân sự	đi quân dịch	tham gia lưu thông	lưu hành
ngghiêm túc	ngghiêm chỉnh	tham quan	thăm viếng
nhà giàn	giàn khoan	thân thương	thân mến
nhà khách	khách sạn	thể hiện	thực hiện
nhân thân	thân nhân	thi công	làm
nhất trí	đồng lòng	thị phần	thị trường
nổi niềm	vẻ suy tư	thiết bị	trang bị
nội thất	trong nhà	thu nhập	lợi tức
nội y	quần áo lót	tiến công	tấn công
phản ánh	phản ảnh	tiến độ	tiến trình

tiên tiến	xuất sắc
tiêu dùng	tiêu thụ
tổ chất	tư chất
tổ lái	phi hành đoàn
tổp ca	ban hợp ca
tờ rơi	truyền đơn
trải nghiệm	kinh nghiệm
tranh thủ	cố gắng
trí tuệ	kiến thức
triển khai	khai triển
truy cập	vào xem
tư duy	suy nghĩ
tư liệu	tài liệu
tư vấn	cố vấn
tuyến đường	con đường
tuyển	đội tuyển
ùn tắc	tắc nghẽn
vận động viên	lực sĩ
vận hành	điều hành
vấn nạn	vấn đề
viện ung bướu	viện ung thư
vô tư	tự nhiên
xác tín	chính xác
xe con	xe du lịch
xe khách	xe đò
xuất khẩu	xuất cảng
nhập khẩu	nhập cảng

REFERENCES

- [1]. Ige B., "Identity and language choice: 'We equals I'," *Journal of Pragmatics*, 42(11), 3047-3054, 2010. <https://doi.org/10.1016/J.PRAGMA.2010.04.018>.
- [2]. Talbot M., Atkinson K., Atkinson D., *Language and Power in the Modern World*. Edinburgh University Press, 2003.
- [3]. Weedon C., *Feminist practice and poststructuralist theory* (2nd ed.). Blackwell, 1997.
- [4]. Wardhaugh R., *An introduction to sociolinguistics* (6th ed.). Blackwell, 2010.
- [5]. Mills S., "Class, gender and politeness," *Multilingua*, 23, 171-190, 2004. <https://doi.org/10.1515/mult.2004.004>
- [6]. Pierce B. N., "Social Identity, Investment and Language Learning," *TESOL*, 29, 9-31, 1995. <https://doi.org/10.2307/3587803>.

- [7] Truitt A., "Vietnamese in Louisiana," in Dajko N., Walton S. (Eds.), *Language in Louisiana: Community and culture*, 246-258, 2019.
- [8] Hoang Tuan, "The Vietnamese diaspora," Asselin P. (Ed.), *The Cambridge history of the Vietnam War: 3. Legacies*, 496-515, 2025.
- [9] United States Census Bureau, *More than 1 million United States residents speak Vietnamese at home*. Census Bureau Reports, 2013. <https://www.census.gov/dataviz/visualizations/045/>.
- [10]. Đặc san Lâm Viên, "Bản đối chiếu chữ Việt Cộng & chữ quốc gia Việt Nam Cộng Hoà" [A comparison of communist words and words used in South Vietnam until 1975], 2019. <http://www.dslamvien.com/p/ban-oi-chieu-tu-ngu-viet-cong-tu-ngu.html?fbclid=IwAR2IKucB7fpxRjMP574S6lrbmzrfPsZ HDDjuwSfr7huKkNCUOdGaAHQbO 3Y>.
- [11] Người Việt, "Học khu Westminster bỏ sách giáo khoa tiếng Việt có từ ngữ trong nước" [Westminster school district removed Vietnamese textbook containing vocabulary popular in Vietnam], 2015. <https://www.nguoi-viet.com/little-saigon/Hoc-Khu-Westminster-bo-sach-giao-khoa-tieng-Viet-co-tu-ngu-trong-nuoc-5003/>.
- [12]. Người Việt, "Hành trang ngôn ngữ 50 năm tiếng Việt hải ngoại" [50 years of overseas Vietnamese language], 2025, <https://www.nguoi-viet.com/van-hoc-nghe-thuat/hanh-trang-ngon-ngu-50-nam-tieng-viet-hai-ngoai/>
- [13]. Dao V. T., Bankston C., "Vietnamese in the USA," Potowski K. (Ed.), *Language diversity in the USA*, 128-145, 2010.
- [14]. Pham A. H., "Ngôn ngữ luôn luôn biến đổi" [Language constantly changes]. *Nông Nghiệp Việt Nam*, 2024.
- [15]. Kính Hoà, "Giữ gìn văn hoá và tiếng nói rất khó khăn tại hải ngoại: Nhà báo Trần Phong Vũ," [It is very difficult to preserve the language and culture for overseas Vietnamese: Journalist Trần Phong Vũ], *Radio Free Asia*, 2015. https://www.rfa.org/vietnamese/in_depth/to-conserv-langu-n-cultur-10142015132822.html.

- [16]. Huỳnh Hoa, "Tiếng Việt từ trong nước ra hải ngoại" [Vietnamese from inside the country to overseas], *Sài Gòn Nhỏ*. 2022. <https://saigonnhonews.com/thoi-su/ca-ke-chu-nghia/tieng-viet-tu-trong-nuoc-ra-hai-ngoai/>

- [17]. Lam M. B., "The cultural politics of Vietnamese language pedagogy," *Journal of Southeast Asian Language Teaching*, 12, 1-19, 2006.
- [18]. Pham A. H., "The non-issue of dialect in teaching Vietnamese," *Journal of Southeast Asian Language Teaching*, 14, 1-17, 2008.
- [19]. *Nhân dân*, 2025, from <https://nhandan.vn>.
- [20]. *Báo Tin tức*, 2025, from <https://baotintuc.vn/>.
- [21]. *Người Việt*, 2025, from <https://www.nguoi-viet.com/>.
- [22]. *Cali Today*, 2025, from <https://www.baocalitoday.com>.

THÔNG TIN TÁC GIẢ

Andrea Hoa Pham¹, Andrew T. Pham²

¹Đại học Florida, Hoa Kỳ

²Đại học Mississippi for Women, Hoa Kỳ