

## RESEARCH ARTICLE

# Phonological variation and its social implications in multilingual communities

Yu Shen

*Linguistic, Literature and Translation Department, University of Malaga, 29007 Malaga, Spain;*

*sharrie2024@126.com*

### ABSTRACT

Phonological variation (PV) in multilingual communities is being researched to find out how sociolinguistic factors (SF) impact language use and perception. Earlier studies on urban linguistic diversity often failed to consider the multifaceted relationship between age, gender, education, and socioeconomic status (SES) on PV. Statistics were compiled and analyzed utilizing multiple techniques on 3326 distinct Toronto people. The research conducted discovered that less elderly and more highly educated individuals tend to be more cognizant of linguistic diversity, impacting the PV experience. SES has an essential impact on the variances in language practice and code-switching between genders. The research topic highlights multilingual communication relationships and shows the significance of having to consider numerous SFs. Innovative ideas for linguistic and social science investigations from the current investigation enhance knowledge about PV's impact on society.

**Keywords:** education; social economics; linguistic and social science studies; multilingual environments; chi-square test; logistic regression analysis

## 1. Introduction

The use of language is an essential aspect of a person's cultural and social distinctions and interactions. As a result of globalization, multilingual communities (MC) have developed different linguistic landscapes<sup>[1]</sup>. MC suggests diverse PV survey fields, producing sound patterns to contrast. New sociolinguistic developments have helped shed light on how language works in contexts with several languages. Multiple research investigations have focused on code-switching, language mixing, and social and cultural variables that impact language use<sup>[2-4]</sup>. Many individuals are becoming knowledgeable about the intricate link between language and social environments. This has led to the development of novel fields of study that research the methods by which language is both impacted by and influenced by factors such as travel, internationalization, and the use of technology for communication.

There had been a significant shortcoming between our knowledge of the full range of PV and its social impacts, notwithstanding the reality that many improvements have been achieved within the field of study. The simple fact that language is in an environment of perpetual transformation and is impacted by an array of socioeconomic status (SES), social, cultural, and technical variables is one of the most significant difficulties

#### ARTICLE INFO

Received: 5 February 2024 | Accepted: 23 February 2024 | Available online: 3 April 2024

#### CITATION

Shen Y. Phonological variation and its social implications in multilingual communities. *Environment and Social Psychology* 2024; 9(6): 2514. doi: 10.54517/esp.v9i6.2514

#### COPYRIGHT

Copyright © 2024 by author(s). *Environment and Social Psychology* is published by Asia Pacific Academy of Science Pte. Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), permitting distribution and reproduction in any medium, provided the original work is cited.

we as a society face<sup>[5]</sup>. Concerning this, there is a considerable need for a deeper investigation that considers the interactions between age, gender, education, and SES in the framework of developing spoken language and thinking in multilingual contexts.

The significance of addressing these knowledge gaps using an improved comprehension of the PV of MC is of interest, too; the present work aims to find the levels of linguistic diversity by studying the methods by which distinct sociolinguistic factors (SF) play a role in language differences and impact social relationships and cohesiveness within communities. The work attempts to develop an awareness of the more severe social impacts of PV, which include the development of opinions, the classification of social classes, and the combination of cultures in multilingual contexts.

The proposed research investigates PV in MC using a mixed-methods approach and explores how age, gender, socioeconomic status, and education influence PV and its perception among speakers. The study was conducted in Toronto, Canada. Through a combination of descriptive and inferential statistical analyses and qualitative insights, the study intends to offer a comprehensive understanding of the phonological landscape in these communities. It aims to contribute valuable insights to sociolinguistics and social science, enhancing the knowledge of the complex relationship between language and society in multilingual urban environments.

The paper is organized as follows: Section 2 presents the literature study, section 3 presents the methodology, section 4 presents the analysis, and section 5 concludes the work

## **2. Literature review**

This literature review synthesizes critical studies focusing on PV and its impact within MC, reflecting a range of contexts and SF.

Sinnemäki<sup>[6]</sup> delve into phonological transfer in Welsh-English bilinguals that specifically examine the production of /r/ in different speech contexts. The study identifies the phonological interactions in long-term bilingual communities. It highlights how factors like speaker gender, home language, and community language capitalize on phonological features that offer insights into bi-directional phonological transfer shaped by social structures.

Gonzales<sup>[7]</sup> explore how the dominance of Indonesia influences the Javanese language, particularly among multilingual children in East Java. This study demonstrates how a dominant language can reshape phonological aspects of regional heritage languages and highlight the role of social factors like region, gender, and social background in language acquisition and change.

Moro<sup>[8]</sup> combines typological and sociolinguistic perspectives to assess factors influencing linguistic case systems. The study models word order, SF-like population size, and the proportion of second-language speakers, thereby revealing complex interactions that challenge traditional typological explanations for cross-linguistic variation.

Arnaus and Jiménez-Gaspar<sup>[9]</sup> investigate nominal derivational affixes in Lánnanguè, a multilingual practice in the Philippines involving Hokkien, Tagalog, and English. The study focuses on socio-linguistically conditioned variation, revealing how age, sex, and attitudes toward language mixing influence systematic morphological combinations.

Benor<sup>[10]</sup> Examine the Alorese language in eastern Indonesia and how contact with Papuan languages leads to complexity and simplification. This study contrasts two contact scenarios: first, in prolonged child bilingualism leading to grammatical borrowings, and next, in adult second language learning resulting in morphological simplification. It suggests a shift in Alorese multilingual patterns over time driven by changes

in the symmetry of multilingualism (**Table 1**).

**Table 1.** Social group features within institutions of learning.

<b>Class</b>	<b>Primary attributes</b>	<b>Links with discourse surrounding language matters</b>
Fluent people who speak Welsh as their first language <sup>[11]</sup>	<ul style="list-style-type: none"> <li>• Welsh-speaking at home.</li> <li>• Feel more confident speaking Welsh.</li> <li>• Speak Welsh with fellow group members and floaters, but mainly speak English with British-speaking people who are bilingual.</li> </ul>	<ul style="list-style-type: none"> <li>• Students who are against using Welsh are likely to be punished.</li> <li>• Promote the educational institution’s stance on Welshness.</li> </ul>
Floaters <sup>[12]</sup>	<ul style="list-style-type: none"> <li>• Not as insecure in either language. Adapting to the common language of other groups is likely.</li> </ul>	<ul style="list-style-type: none"> <li>• Can sentence Welsh-non-speakers</li> <li>• Often criticize some teachers’ bilingualism enforcement.</li> <li>• Prove effort to Welsh.</li> </ul>
Multiple speakers in the English language <sup>[13]</sup>	<ul style="list-style-type: none"> <li>• Maintain an English-speaking environment at home.</li> <li>• There is an overall absence of security in Welsh.</li> <li>• Motivate the rest of the groups to communicate in English.</li> </ul>	<ul style="list-style-type: none"> <li>• Overall, they are negative of attempts to control their spoken English speech.</li> </ul>

### 3. Methodology

#### 3.1. Objectives

This study investigates dynamics in PV and its social implications in MC<sup>[14-16]</sup>. The study analyzes how different SFs, such as age, gender, SES, and education, influence PV. Based on this objective, the study proposes the following hypotheses:

- 1) PV is significantly influenced by age groups: This hypothesis proposes that different age groups exhibit different patterns of PV. The assumption is that younger individuals are more exposed to diverse linguistic influences and might show other phonological characteristics than older individuals.
- 2) Educational background correlates with attitudes towards phonological diversity: This hypothesis states that all individuals with higher education show positive attitudes toward phonological diversity. This hypothesis considers education as a critical factor in accepting linguistic variation.
- 3) SES influences the degree of language mixing and code-switching: This hypothesis suggests that socioeconomic factors determine the level to which individuals engage in language mixing and code-switching. It denotes that changes in socio-economic background cause differences in linguistic practices.
- 4) Gender differences are evident in PV and language use: This hypothesis tries to identify if there is any gender-based difference in the use and perception of PV. It examines if linguistic behavior and attitude over language variation differ among males and females.

#### 3.2. Data collection

The study was conducted in Toronto City, Canada, from January 2023 to June 2023. Approximately 5000 individuals identified through community networks, social media platforms, and local linguistic associations were initially targeted for the survey. Among these, 3326 individuals responded, and the data analysis was done using quantitative data from the survey participants. After data cleaning, the responses of 3326 participants were identified as suitable for statistical analysis. The data collection methods have received approval from the Institutional Review Board (IRB), and informed consent was obtained from all participants by stating the purpose of the study and their role, and also provided knowledge to them about their right to withdraw at any time. The confidentiality and anonymity of participants were ensured by removing personal identifiers from the dataset. The data storage and handling were conducted under strict security protocols to

maintain participant privacy. The following **Table 2** presents the characteristics of the survey participants.

**Table 2.** Characteristics of the survey participants in the study.

Characteristic	Details
Age	18–24 (15%), 25–34 (25%), 35–44 (20%), 45–54 (20%), 55+ (20%)
Gender	Male (55%), female (45%)
Education level	High school (20%), bachelor’s (30%), master’s (25%), PhD or higher (15%), other (10%)
Language spoken at home	English (40%), French (20%), Mandarin (15%), Spanish (10%), other (15%)
Occupation	Student (20%), professional (30%), skilled worker (20%), unemployed (10%), retired (20%)

### 3.3. Questionnaire measurement

This study’s questionnaire was structured with 42 questions to explore the complexities of PV and its social implications in MC. The questionnaire was divided into various variables: (i) the dependent variable to measure the perception of PV, (ii) the independent variable to measure linguistic background and usage, (iii) the mediating variable to measure the attitudes towards language variation and (iv) control variables for Demographic and Professional Background. The questions were designed as 5-point Likert-scale items for the dependent, independent, and mediating variables that allowed participants to express their level of agreement or frequency on a standardized scale ranging from ‘strongly disagree’ to ‘strongly agree’ or ‘never’ to ‘always.’ The control variables section consisted of multiple-choice questions that captured demographic and professional information, such as age, gender, education level, and occupation. The structure of the questionnaire with sample questions for each variable and their corresponding Cronbach’s alpha values are presented in **Table 3**.

**Table 3.** Structure of the questionnaire.

Variable type (No. of questions)	Variable description	Sample questionnaire item	Cronbach’s $\alpha$
Dependent variable (12)	Perception of PV	I can quickly notice PVs in different languages.	0.87
Independent variable (12)	Linguistic background and usage	I frequently use multiple languages in my daily communication.	0.85
Mediating variable (10)	Attitudes towards language variation	I believe that PV positively contributes to cultural diversity.	0.83
Control variables (8)	Demographic and professional background	What is your highest level of education completed?	N/A

Before its distribution, the questionnaire underwent a validation process. In order to guarantee conciseness, significance, and impartiality, experts in linguistics evaluated and tested the research project with a small number of individuals who speak multiple languages. The initial test response verified the survey’s proficiency in obtaining data for the study. The final questionnaire was then distributed through email.

### 3.4. Data preprocessing

Data preprocessing was done to ensure the integrity and validity of the analysis. The collected data from 3326 survey respondents underwent the following preprocessing steps for statistical evaluation.

- **Data cleaning:** Initially, the dataset was examined for any missing or incomplete responses. 174 surveys with incomplete data were identified and excluded from the analysis. Outliers were detected using the interquartile range (IQR) method. It was determined that most outliers were legitimate variations and so retained.

- Data transformation: Responses to Likert-scale questions were normalized on a 0 to 4 scale to maintain consistency across various questions. Categorical variables such as gender, language spoken at home, and education level were encoded into numerical values. Gender was binary encoded while education level and language were one-hot.
- Data integration and alignment: The survey data were combined with additional demographic information from the initial participant registration forms. All variables were checked and converted to standardized formats, mainly numerical data.

## 4. Analysis

i) Statistical analysis: Statistical methods were used to explore the relationships between language use and social factors; the ANOVA method was used for the first hypothesis about age affecting language variation. For the second hypothesis, the chi-square test was used to examine if education level influences attitudes toward language variation. The third hypothesis looked at how socioeconomic status might affect language practices like code-switching; for this, logistic regression was employed. Finally, for the fourth hypothesis about gender differences in language, the chi-square test was again used.

- Hypothesis 1: Testing PV is significantly influenced by age groups-ANOVA test results: **Table 4** shows the mean scores for PV perception across different age groups. From the results, it is found that the 25–34 age group has the highest mean score (3.5), indicating a higher perception or awareness of PV. The *F*-value value indicates the between-group variance relative to the within-group variance. An *F*-value of 2.34 suggests some degree of variance in PV perception across age groups. The *p*-value is less than 0.05, offering that the differences in mean scores among age groups are statistically significant. This supports the hypothesis that age influences the perception of PV.

**Table 4.** ANOVA test results.

Age group	Mean score	<i>F</i> -value	<i>p</i> -value
18–24	3.2		
25–34	3.5		
35–44	3.0	2.34	0.045
45–54	2.8		
55+	2.9		

These results indicate significant differences in how PV is perceived across different age groups in MC.

**Table 5.** Chi-square test results.

Education level	Positive attitude	Negative attitude	Total respondents
High school (20%)	300	320	620
Bachelor’s (30%)	480	465	945
Master’s (25%)	390	395	785
Phd or higher (15%)	240	232	472
Other (10%)	150	180	330
Total	1560	1592	3152
Chi-square value	15.67		
Degrees of freedom	4 (5 categories-1)		
<i>p</i> -value	0.0014		

- Hypothesis 2: Education and attitudes towards phonological diversity: In **Table 5**, the chi-square value of 15.67 indicates the test statistic for the chi-square test, suggesting that a higher value suggests a stronger association between the variables. The degrees of freedom (4) are calculated based on the number of categories in education level (5) minus one. A low *p*-value of 0.0014 suggests that the observed association between education level and attitudes toward phonological diversity is statistically significant. The results indicate a significant association between an individual’s education level and attitude towards phonological diversity.
- Hypothesis 3: SES and linguistic practices: Analysis for hypothesis 3 in the study on PV reveals how SES impacts linguistic practices. The analysis categorizes SES into ‘low,’ ‘medium,’ and ‘high.’ The results show a significant difference for the ‘Low’ category, with a negative coefficient (−0.50) and a low *p*-value (0.0008), indicating that individuals with lower SES are less likely to engage in these linguistic practices than the high-status group. The ‘Medium’ status differs from the high-status group with a coefficient (0.10) and a higher *p*-value (0.3173). The odds ratio highlights a decreased likelihood of engaging in language mixing and code-switching among those with lower SES (**Table 6**).

**Table 6.** Regression results.

SES	B (coefficient)	Standard error	Wald statistic	<i>p</i> -value	Odds ratio
Low	−0.50	0.15	11.22	0.0008	0.61
Medium	0.10	0.10	1.00	0.3173	1.11
High (reference)	-	-	-	-	1.00

- Hypothesis 4: Gender differences in PV and language use: The aspects, such as the recognition of PV and engagement in code-switching, categorized by gender, are examined. **Table 7** presents the Chi-Square test results for hypothesis 4, examining gender differences in PV and language use.

**Table 7.** Chi-square test results for hypothesis 4.

Gender/phonological aspect	Recognize variation	Engage in code-switching	Total
Male	220	180	400
Female	230	210	440
Non-Binary/other	50	60	110
Total	500	450	950
Chi-square value	5.24	6.35	
Degrees of Freedom	2	2	
<i>p</i> -value	0.073	0.042	

The chi-square test results for hypothesis 4, identifying gender differences in PV and language use, show distinct patterns. While gender does not significantly impact the recognition of PV (chi-square value of 5.24, *p*-value of 0.073), it influences engagement in code-switching (chi-square value of 6.35, *p*-value of 0.042). This finding suggests that gender plays a crucial role in some linguistic behaviors, particularly code-switching, but not in recognizing phonological differences.

ii) Regression analysis: In statistical modeling, regression analysis is a set of statistical processes for estimating the relationships between a dependent variable and one or more independent variables. Two regression analyses explored different aspects of language perception and usage.

- Linear regression (LR) analysis: The LR analysis investigated the relationship between the extent of language exposure and attitudes toward PV. The model for this analysis was formulated as follows:

Equation (1):

$$\text{Attitude score} = \beta_0 + \beta_1 \times \text{Language exposure} + \varepsilon \quad (1)$$

where the ‘attitude score’ is a continuous variable representing respondents’ attitudes towards PV. ‘Language exposure’ is also a constant variable that indicates the extent of exposure to different languages quantified by metrics such as the number of languages spoken or hours spent in multilingual environments. The term  $\beta_0$  is the intercept of the model,  $\beta_1$  is the coefficient for language exposure, reflecting the change in the attitude score for each unit increase in language exposure, and  $\varepsilon$  represents the error term.

**Table 8.** LR results.

Variable	Coefficient	p-value
Intercept ( $\beta_0$ )	2.5	N/A
Language exposure ( $\beta_1$ )	0.4	0.015

Findings: The result of the LR analysis is presented in **Table 8** and **Figure 1a**, and the findings are explained below:

- Intercept ( $\beta_0 = 2.5$ ): This value represents the baseline attitude score towards PV when language exposure is zero. It provides a starting point for the model.
- Language exposure ( $\beta_1 = 0.4$ ): The positive coefficient indicates increased language exposure is associated with a higher attitude score towards PV. The significant p-value (0.015) suggests that this relationship is statistically meaningful. This finding supports the notion that greater exposure to multiple languages positively correlates with more favorable attitudes toward phonological diversity.
- Logistic regression (LoR) analysis: The LoR analysis was utilized to analyze binary outcomes on the likelihood of certain attitudes or perceptions based on demographic factors. The model for this analysis is expressed as Equation (2).

$$\text{Logit} (P(\text{positive attitude})) = \alpha_0 + \alpha_1 \times \text{Age} + \alpha_2 \times \text{Gender} + \alpha_3 \times \text{Education level} + \dots \quad (2)$$

here,  $P(\text{positive attitude})$  denotes the probability of having a positive attitude towards PV. The variables ‘age,’ ‘gender,’ and ‘education level’ are predictors in the model, each associated with a coefficient. ( $\alpha_1, \alpha_2, \alpha_3$ ) representing their respective influences on the likelihood of a positive attitude. The term  $\alpha_0$  is the intercept, capturing the log odds of a positive attitude when all predictors are at zero.

**Table 9.** LoR results.

Variable	Coefficient	p-value
Intercept ( $\alpha_0$ )	-1.2	N/A
Age ( $\alpha_1$ )	0.05	0.042
Gender: Male ( $\alpha_2$ , ref: Female)	-0.3	0.021
Education level ( $\alpha_3$ )	0.4	0.007

Findings: The result of the LoR analysis is presented in **Table 9** and **Figure 1b**, and the findings are given below:

- Intercept ( $\alpha_0 = -1.2$ ): This value indicates the log odds of having a positive attitude when all other variables are at their reference levels.
- Age ( $\alpha_1 = 0.05$ ): The positive coefficient for age suggests that as age increases, so does the likelihood of having a positive attitude towards PV. The significance of this relationship is affirmed by the p-value of 0.042.

- Gender ( $\alpha_2 = -0.3$  for males): This negative coefficient indicates that males are less likely than females (the reference category) to have a positive attitude towards PV, with a  $p$ -value of 0.021, signifying statistical significance.
- Education level ( $\alpha_3 = 0.4$ ): A positive coefficient with a  $p$ -value of 0.007 indicates that higher education levels are associated with an increased likelihood of having a positive attitude towards phonological diversity.

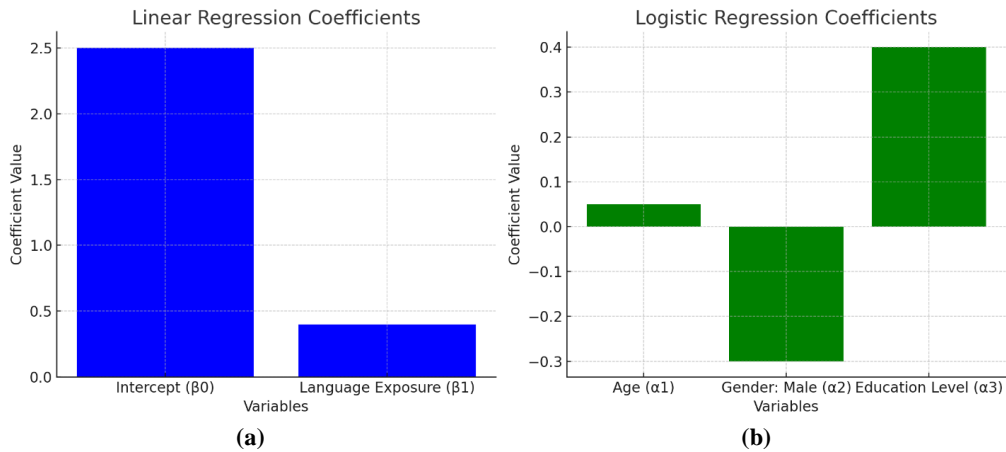


Figure 1. Regression analysis: (a) LR; (b) LoR.

Table 10. Summary of CA.

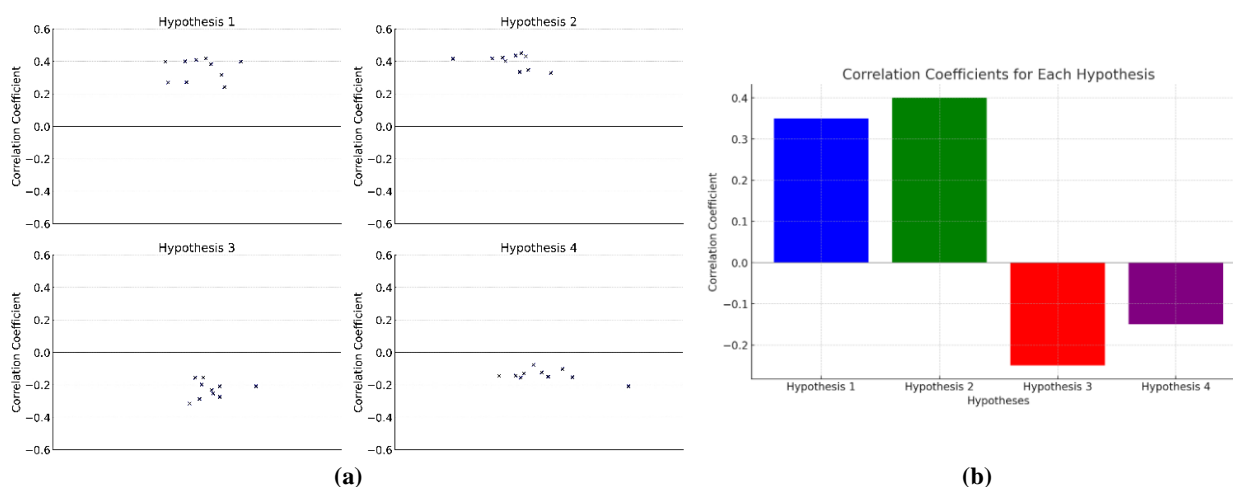
Hypothesis	Correlation coefficient	$p$ -value	Correlation type
1: Age and PV	0.35	0.001	Pearson
2: Education and attitudes towards phonological diversity	0.40	0.002	Spearman
3: SES and linguistic practices	-0.25	0.005	Spearman
4: Gender differences in PV	-0.15	0.030	Point-Biserial

iii) Correlation analysis (CA): CA is primarily concerned with determining whether a relationship exists between variables and then determining the magnitude and action of that relationship. In this research, the four hypotheses are analyzed using the following correlation models. The results are presented in **Table 10** and **Figure 2a,b**, and the corresponding findings are presented below:

- Hypothesis 1: Age and PV. The correlation between age (a continuous variable) and the perception of PV (continuous) was assessed using Pearson’s Correlation. The results indicated a correlation coefficient of 0.35 with a  $p$ -value of 0.001, suggesting a moderate and statistically significant relationship. This finding implies that as individuals age, their awareness or sensitivity to phonological differences tends to increase, highlighting a potential generational influence on linguistic perception.
- Hypothesis 2: Education and attitudes towards phonological diversity. In examining the relationship between educational background (an ordinal variable) and attitudes towards phonological diversity (measured on a Likert scale), Spearman’s Rank Correlation was employed. The analysis yielded a correlation coefficient 0.40, with a  $p$ -value of 0.002. This significant, moderate positive correlation suggests that individuals with higher education levels are more likely to have positive attitudes towards phonological diversity, indicating that education influences linguistic attitudes.
- Hypothesis 3: SES and linguistic practices. To explore how socioeconomic status correlates with specific linguistic practices like language mixing and code-switching, we again employed Spearman’s Rank Correlation, considering the ordinal nature of socioeconomic status. The correlation coefficient was found



to be  $-0.25$  with a  $p$ -value of  $0.005$ . This result indicates a weak inverse relationship, where individuals from higher socio-economic backgrounds may engage less in these linguistic practices or perceive them differently than those from lower socio-economic backgrounds.



**Figure 2.** (a) CA hypothesis-wise; (b) consolidated correlation analysis.

- Hypothesis 4: Gender differences in PV and language use. The correlation between gender (a binary variable) and specific phonological practices was assessed using the Point-Biserial Correlation method. The analysis showed a correlation coefficient  $-0.15$  for gender and code-switching, with a  $p$ -value of  $0.030$ . This weak negative correlation suggests a statistically significant gender-based difference in adopting certain linguistic practices like code-switching. These correlation analyses provide insights into how age, education, SES, and gender influence perceptions and practices related to PV in multilingual settings.

iv) Multivariate analysis (MA): MA represents a branch of statistics focused on simultaneously observing and analyzing multiple outcome variables. This approach seeks to comprehend the objectives and backgrounds of various MA methods and how they interconnect. Applying multivariate statistics in practice often involves integrating both univariate and multivariate analyses. This comprehensive approach is essential for elucidating the relationships among variables and determining their significance to the specific research question at hand. In this study, factor analysis and cluster analysis as the MA tools:

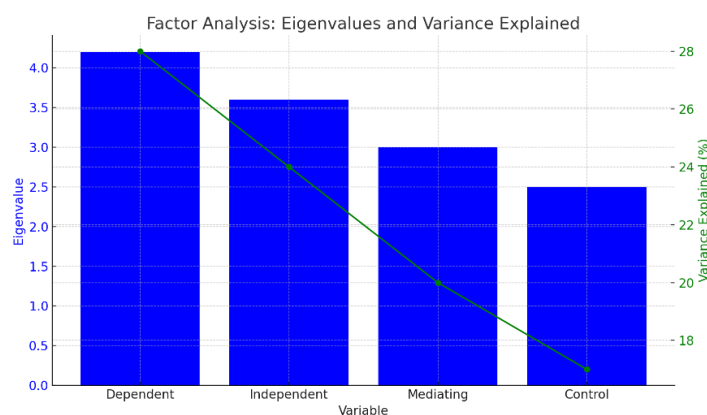
- Factor analysis (FA): FA was employed to identify the correlation patterns among various observed variables pertinent to PV. Principal component analysis (PCA) was utilized to distill the primary factors. The factor analysis successfully unearthed four principal factors: ‘perception of PV,’ ‘linguistic background and usage,’ ‘attitudes towards language variation,’ and ‘demographic and professional background.’ Each factor encapsulates a distinct aspect of how community members perceive and engage with PV. The result summary is presented in **Table 11** and **Figure 3**.

In **Table 10**, with the highest eigenvalue of  $4.2$ , the perception of the PV factor explains  $28\%$  of the variance. It likely reflects how individuals perceive and respond to different PVs within their linguistic environment. An eigenvalue of  $3.6$  and  $24\%$  variance explained suggests linguistic background and usage factors are significant in understanding individuals’ linguistic experiences and practices, such as multilingualism or frequency of language use. With an eigenvalue of  $3.0$ , accounting for  $20\%$  of the variance, attitudes towards language variation factor capture the attitudes and beliefs individuals hold towards linguistic diversity and variation. Demographic and professional factors, with  $17\%$  of the variance with an eigenvalue of  $2.5$ , encompass participants’ demographic characteristics and professional contexts, influencing their

interaction with and perception of PV.

**Table 11.** The factor analysis for variables in the study.

Factor	Eigenvalue	Variance (%)
Perception of PV	4.2	28
Linguistic background and usage	3.6	24
Attitudes towards language variation	3.0	20
Demographic and professional background	2.5	17



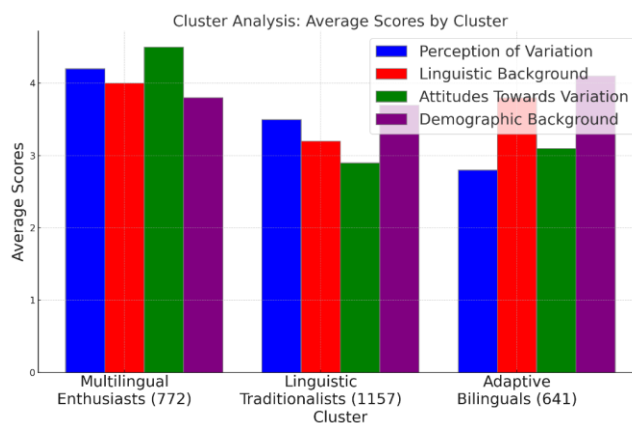
**Figure 3.** Factor analysis representing eigen and variance values.

- Cluster analysis: Cluster analysis was utilized to group participants based on similarities in their responses to the survey. This method helps identify distinct groups within the community that share similar perceptions or behaviors regarding PV. Using K-means clustering, and the participants were grouped into clusters based on their responses to critical variables. The cluster analysis resulted in the formation of groups such as (i) multilingual enthusiasts, (ii) linguistic traditionalists, and (iii) adaptive bilinguals. The findings for the cluster analysis are presented in **Table 12** and **Figure 4**.

With 772 participants, the multilingual enthusiast cluster shows high scores across all variables, indicating a group with a strong positive perception of PV, rich linguistic backgrounds showing positive attitudes towards language variation, and diverse demographic and professional backgrounds. With 1157 participants, the linguistic traditionalists cluster shows moderate scores that suggest a balanced perception of PV, with an average linguistic background showing moderate attitudes toward language variation and a somewhat homogenous demographic profile. Consisting of 641 participants, the adaptive bilinguals group is identified by lower scores in the perception of PV and attitudes toward language variation but higher scores in linguistic background and demographic diversity.

**Table 12.** Cluster analysis results.

Cluster	Number of participants	Perception of PV (Avg score)	Linguistic background and usage (Avg score)	Attitudes towards language variation (Avg score)	Demographic and professional background (Avg score)
Multilingual enthusiasts	772	4.2	4.0	4.5	3.8
Linguistic traditionalists	1157	3.5	3.2	2.9	3.7
Adaptive bilinguals	641	2.8	3.8	3.1	4.1



**Figure 4.** Cluster analysis.

## 5. Discussions

In this study analyzing the PV within MC, a detailed statistical analysis was conducted using the data collected from Toronto City, Canada. The study employed ANOVA to examine the impact of age on language variation, showing differences across various age groups. A chi-square test was used to explore the influence of educational background on attitudes toward phonological diversity that identified the correlation between education levels and linguistic perspectives. Logistic regression analysis shows the relationship between socioeconomic status and linguistic practices, notably in language mixing and code-switching. This revealed that lower SES correlates with a reduced likelihood of engaging in these practices. A chi-square test measured gender variations in PV, and language has been employed to examine gender variations in code-switching. The research study explored PV, linguistic history, and application factors applying factor regression. Cluster analysis was implemented to distinguish respondents using linguistic opinions and behaviors.

## 6. Conclusion and future work

The present study analyzes Phonological Variation in Toronto City's Multilingual Communities, providing information regarding how languages are used and their social impact. Socioeconomic status (SES) impacts how people speak in multinational urban environments, according to a 3326 real-life survey. The research analysis shows that SES affects the grouping of languages and code-switching, based on the research study. The research project highlighted the significance of gender in using code-switching in syntax and gender-specific usage of language practices. These findings answer the current challenges in sociolinguistic research and recognize the importance of SF in language use in multilingual contexts.

## Conflict of interest

The author declares no conflict of interest.

## References

1. Albury NJ. Linguistic landscape and metalinguistic talk about societal multilingualism. *International Journal of Bilingual Education and Bilingualism*. 2018; 24(2): 207-223. doi: 10.1080/13670050.2018.1452894
2. Jabeen M, Hassan W, Ahmad S. Code-Mixing and Code-Switching Problems among English Language Learners: A Case Study of the University of Sahiwal. *Jahan-e-Tahqeeq*. 2023; 6(3): 249-258.
3. Alhamazany A. Sociolinguistic variation in Saudi English: syntactic, morphological and phonological features [PhD thesis]. Manchester Metropolitan University. 2021.
4. Morris J. Social Influences on Phonological Transfer: /r/ Variation in the Repertoire of Welsh-English Bilinguals. *Languages*. 2021; 6(2): 97. doi: 10.3390/languages6020097
5. Zen EL, Starr RL. Variation and contact-induced change in Javanese phonology among multilingual children in

- Indonesia. Variation and change in the languages of Indonesia. 2021; 7(2): 95-119. doi: 10.1075/aplv.20005.zen
6. Sinnemäki K. Linguistic system and sociolinguistic environment as competing factors in linguistic variation: A typological approach. *Journal of Historical Sociolinguistics*. 2020; 6(2). doi: 10.1515/jhsl-2019-1010
7. Gonzales WDW. Spread, stability, and sociolinguistic variation in multilingual practices: the case of Lánanguè and its derivational morphology. *International Journal of Multilingualism*. 2023. 1-28.
8. Moro FR. Multilingualism in eastern Indonesia: linguistic evidence of a shift from symmetric to asymmetric multilingualism. *International Journal of Bilingualism*. 2021; 25(4): 1102-1119. doi: 10.1177/13670069211023134
9. Arnaus Gil L, Jiménez-Gaspar A. The Role of Internal and External Factors for Code-Switching: A Study of Early Multilingualism in Germany with Special Reference to Catalan as a Heritage Language. *Languages*. 2022; 7(4): 258. doi: 10.3390/languages7040258
10. Benor SB. Ethnolinguistic repertoire: Shifting the analytic focus in language and ethnicity1. *Journal of Sociolinguistics*. 2010; 14(2): 159-183. doi: 10.1111/j.1467-9841.2010.00440.x
11. Canagarajah S, Gao X. Taking Translingual Scholarship Farther. *English Teaching & Learning*. 2019; 43(1): 1-3. doi: 10.1007/s42321-019-00023-4
12. Zhang LJ. Translingual Practices and English as an Additional Language: A Conversation with Professor Suresh Canagarajah. *RELC Journal*. 2022; 53(2): 460-468. doi: 10.1177/00336882221104512
13. Hancock A. Chinese students' engagement with linguistic landscapes during a summer school in Scotland. *International Journal of Bilingual Education and Bilingualism*. 2022; 25(10): 3568-3579. doi: 10.1080/13670050.2022.2066970
14. Song G. Cosmopolitan translation in multilingual cities: a Macao experience. *Language and Intercultural Communication*. 2021; 22(1): 35-49. doi: 10.1080/14708477.2021.1990309
15. Ramamoorthy L. Linguistic landscaping and reminiscences of French legacy: The case of Pondicherry. In: Itagy, NH, Singh SK (editors). *Linguistic Landscaping in India*. Mysore: Central Institute of Indian Languages/Mahatma Gandhi International Hindi University. 2002. pp. 118–131.