ACADEMIA International Journal for Social Sciences
Volume 5, Issue 2(a), 2026 ISSN-L (Online): 3006-6638

Automatic Detection of Code-Switching Patterns in Pakistan ESL Classrooms Using NLP Techniques

Dr. Abdul Khaliq
abdulkhaliq@cuvas.edu.pk
Assistant Professor, Department of Social and Allied Sciences, Cholistan University of Veterinary and Animal Sciences, Bahawalpur, Pakistan

Muhammad Zohair
22-cuvas-0146@student.cuvas.edu.pk
BSCS scholar, Department of computer science and information technology, cholistan University of veterinary and animal sciences, Bahawalpur

Shoaib Yaseen
22-cuvas-0122@student.cuvas.edu.pk
BSCS scholar, Department of computer science and information technology, cholistan University of veterinary and animal sciences, Bahawalpur

Corresponding Author: * Dr. Abdul Khaliq abdulkhaliq@cuvas.edu.pk

Received: 25-12-2025 **Revised:** 09-01-2026 **Accepted:** 23-01-2026 **Published:** 09-02-2026

ABSTRACT

This paper will look at the code-switching practices in ESL classrooms in Pakistan, and its automatic identification using Natural Language Processing (NLP) techniques. Using quantitative research design, the data were gathered on classroom discourse in selected universities in the public sector in the cities of Bahawalpur and Rahim Yar Khan. The data, which included teacher explanations, student interactions, and group discussions, were transcribed and analyzed with the help of the token-level language identification, n-gram analysis, and machine learning models. The results indicate that code-switching is common and contextual with the greatest frequency being observed in collaborative learning context. According to functional analysis, it is mostly applied in terms of description and clarification, which facilitates understanding and involvement. The statistical outcomes show that there is a significant positive correlation between code-switching and learning outcomes. Also, hybrid models developed based on NLP showed the best accuracy in detecting patterns of code switching and this underlines their reliability and applicability in education research. The research concludes that code-switching is an important pedagogical tool and that NLP is a scalable, data-driven methodology of studying multilingual classroom discourse.

Keywords: *code-switching, ESL classrooms, natural language processing, multilingual education, language detection.*

INTRODUCTION

Introduction of research

Language learning in multilingual situations tends to be dynamic, that is, more than one language is used in one interaction. In Pakistan, where English is used as a second language, and Urdu and regional languages are used as primary communication means, the phenomenon of code-switching in the ESL classroom is quite natural and common in that country. Code-switching is the process of switching between two or more languages during a conversation or utterance and is often used in a strategic manner to help establish meaning and communication (Poplack, 1980).

In classrooms, teachers and students often alternate between English and Urdu to elaborate on complicated topics, clarify meaning, and keep the conversation going. This is a practice that does not

only reflect linguistic competence but also a communicative strategy that supports learning (Macaro, 2001; Shamim, 2008). Multilingual education research has indicated that code switching can clarify, decreasing cognitive load, and providing a more inclusive learning experience (García and Wei, 2014).

Historically, the research of code-switching has been based on qualitative research methods including classroom observations and manual transcription. These methods offer very valuable insights, but they tend to be time-consuming, subjective and may be small in scale. The development of Natural Language Processing (NLP) enables researchers nowadays to analyze large amounts of text and spoken information more easily and objectively. Automated code-switching pattern detection using NLP techniques, such as language identification and text classification, opens new opportunities to conduct data-driven educational research (Solorio & Liu, 2008; Rajdhani et al., 2017).

Thus, the research aims at using the methods of NLP to identify and examine the patterns of code-switching in ESL classrooms in Pakistan. By combining computational techniques with research in the field of language education, the study will help to offer a more systematic insight into the interaction in bilingual classrooms.

Statement of the Problem

Although the phenomenon of code-switching is highly widespread in ESL classrooms in Pakistan, it still has little systematic analysis. Most current research studies are based on qualitative observations, which, despite their valuable insights, are not scalable and objective. Such traditional methods are usually limited to small sample sizes and researcher bias, and it is hard to generalize the findings of such traditional methods in various settings (Shamim, 2008).

Moreover, empirical studies that use sophisticated computational methods, including NLP, to study classroom conversations are missing. Although NLP has been effectively applied in the analysis of multilingual texts and the identification of languages, its use in the learning process, especially in the Pakistani ESL classrooms, has been under-researched (Solorio and Liu, 2008). Consequently, there is a lack of knowledge of the frequency, pattern, and functionality of code-switching in actual classroom settings.

The other urgent challenge is the lack of data-driven instruments that can help the educators monitor and analyze the language use in the classrooms. In the absence of such tools, it becomes difficult to assess the effectiveness of code-switching as a pedagogical intervention or to come up with evidence-based teaching practices. This gap highlights the need for an automated and scalable approach to studying code-switching using NLP techniques.

Research Objectives

The main aims of this research are:

1. Aim. To determine the patterns of code-switching in Pakistani ESL classrooms.
2. 2. To use the NLP techniques to automatically detect code-switching.
3. 3. To examine how code switching contributes towards better understanding and student involvement.

Research Questions

This research paper will seek to answer the following research questions:

1. What is the frequency of code-switching in ESL classrooms in Pakistan?
2. What are the common patterns and types of code switching used by teachers and students?
3. To what extent can NLP be used to identify and categorize cases of code-switching?

Importance of Research

This research has a great scholarly and practical importance. Academically, it adds to existing body of literature by combining linguistic theory and computational approaches, which forms a new approach to studying code-switching in educational settings. Although past studies are descriptively based, this study presents an automated, data-driven framework that is more objective and scalable.

Practically speaking, the research could be very useful to educators and policymakers. By studying code-switching behavior, educators can develop more effective instructional interventions that can utilize bilingual resources to promote learning. Also, NLP tools should be used to help the institutions to analyze classroom interactions in a more efficient way, to make evidence-based decisions.

Moreover, the work contributes to the overall inclusion of technology in classroom by showing how NLP can be used to apply to actual classroom data. This may open the way towards creation of intelligent educational systems that will track and enhance language learning procedures.

Study Limitations

There are some limitations that characterize this study. To begin with, the study is confined to the ESL classrooms in Pakistan, and the study specifically focuses on the code-switching between English and Urdu, which may restrict the overall external validity of the results to other language settings.

Second, the research is based on transcribed classroom data, which might not be well suited to capturing non-verbal communication and situational nuances of verbal communication. Third, the NLP models applied in this research are based on textual analysis and more sophisticated speech-based models are not considered because of limited resources.

The sample size and data collection is also confined to the selected classrooms, which might not be a representative of all ESL settings in Pakistan. Regardless of these constraints, the study offers some valuable information on the use of NLP in the analysis of code-switching patterns in learning institutions.

LITERATURE REVIEW

As a natural and working linguistic phenomenon, code-switching has been widely studied both in the context of bilingual and multilingual environments. According to Poplack (1980), code-switching can be defined as the alternation of two or more languages in one discourse, which is often under grammatical and sociolinguistic constraints. In the teaching context, specifically in the ESL classes, the phenomenon of code-switching is not only a linguistic phenomenon but also a pedagogical process that serves to facilitate understanding, elucidate meaning and support the process of interaction (Macaro, 2001). In multilingual societies such as those in Pakistan where English is used in classrooms along with other languages like Urdu and regional languages, code-switching forms a very deep part of classroom communication. Rahman (2010) points out that the bilingual character of the Pakistani education system promotes the application of mixed language practices, whereas Shamim (2008) observes that the mixed language practices are commonly used by the teachers to clarify the complex concepts and to keep the students engaged.

This study is conceptually based on interactions between classroom discourse and code-switching behavior and automated detection using NLP techniques. The data will mainly be collected in the form of classroom discourse where code-switching can be observed in the lexical, syntactic, and discourse levels. NLP is the analysis system that recognizes and categorizes language alternation patterns and turns qualitative interactions into quantifiable data. The result of this process is an organized knowledge of the frequency, patterns, and functions of code-switching. This conceptual interconnection implies that computational tools may help fill the gap between linguistic theory and empirical analysis, making it possible to conduct scalable and objective research in multilingual education (Solorio & Liu, 2008).

The conceptual framework of this study is informed by Sociolinguistic Theory, Interactionist Theory, and Computational Linguistics model. The sociolinguistic theory describes code-switching as a socially oriented behavior based on the circumstances, identity, and communicative needs (Gumperz, 1982). The interactionist theory highlights the importance of language interaction in enabling learning and proposes that code-switching enables meaning negotiations and comprehension during acquisition of second language (Long, 1996). Also, computational linguistics offers a methodological framework of studying multilingual text based on machine learning and language recognition algorithms (Jurafsky and Martin, 2020). Combined, these frameworks legitimize the combination of linguistic and technological methods in studying code-switching.

Some studies have been conducted on the code-switching and language practices of ESL classrooms in the Pakistani context. The research question of Rahman (2010) is language policy and classroom practices, and the aim is to understand the dynamics of bilingual education using qualitative analysis and concluding that code-switching is an inevitable and functional practice although lacking in quantitative measurement. In their research on the trends of teaching English language, Shamim (2008) used classroom observations and discovered that teachers are overly dependent on the use of Urdu to explain the material to the students. In her study, Shamim (2008) did not use computational tools. Khan and Mansoor (2017) investigated the language preferences of students using a questionnaire and reported the findings of the survey, but their methodology did not involve the analysis of the discourse. Ali and Hamid (2018) used qualitative methods to study the patterns of classroom interaction and concluded that code-switching increased interaction levels but did not quantify switching frequency. Raza et al. (2019) investigated the attitude of teachers towards code-switching, and they found that teachers had positive attitudes towards its pedagogical value, although their study was not data-driven, but instead perception-based. Iqbal and Akhtar (2020) evaluated the efficacy of code-switching in ESL learning and found that code-switching is an effective tool to understand the material, but they did not examine linguistic patterns. Hussain et al. (2021) explored the interaction between students in multilingual classrooms and found that mixed language use enhances interaction, but there was no computational analysis. Zafar and Qureshi (2022) explained the potential of digital tools in language teaching but did not specifically refer to the NLP-based detection. Tariq et al. (2023) examined online ESL classrooms and found that code-switching was common, but they did not discuss the analysis further. Lastly, Malik et al. (2023) qualitatively analyzed classroom discourse and justified the usefulness of code-switching but did not use automated methods.

Studies of code-switching and NLP have improved a great deal across national boundaries. Poplack (1980) offered some of the preliminary work of structural constraints of code-switching by linguistic analysis. In the research by Macaro (2001), the researcher utilized observational techniques to study code-switching in language classrooms and concluded that code-switching is helpful in promoting understanding. Gumperz (1982) examined code-switching in conversations and highlighted its social functions. Long (1996) also emphasized the importance of interaction in the process of language learning, which in turn indirectly supports the use of bilingual communication. Solorio and Liu (2008) proposed machine learning methods to predict code switching points in multilingual texts, thus showing that automated detection is possible. Nguyen and Dogruoza (2013) used NLP to examine bilingual conversations and discovered that there was high accuracy when identifying language switches. Rijhwani et al. (2017) introduced computational models of the detection of code-switching in

multilingual datasets, demonstrating better performance with hybrid methods. In the study of Barman et al. (2014) language identification in code-mixed text in social media was the subject of the study. The researchers achieved consistent classification results. Pratapa et al. (2018) modeled the patterns of code-switching using neural models and showed the efficacy of deep learning strategies. Lastly, Winata et al. (2019) used multilingual embeddings in code-switching recognition, and they obtained state-of-the-art results, which emphasizes the potential of the sophisticated NLP methods.

An overview of national and international literature indicates that, although code-switching is a widely accepted concept of a useful pedagogical technique, most research in Pakistan is at most qualitative or survey-based, limiting its scalability and objectivity. Conversely, global studies have shown that NLP techniques are effective in identifying and examining multilingual patterns but is usually done in non-classroom or high-tech environments. It is evident that no research has included the application of NLP-based analysis and ESL classroom discourse in Pakistan.

Thus, the most significant gap in the literature is the lack of empirical, quantitative studies which implement NLP techniques to identify and analyze patterns of code-switching in Pakistani ESL classrooms. The research proposed fills this gap by integrating sociolinguistic theory with computational approaches that will offer a new perspective on the analysis of bilingual classroom interactions and will be of value to both language education and NLP research.

RESEARCH METHODOLOGY

Research Design

To examine the patterns of code-switching in ESL classrooms by applying the techniques of Natural Language Processing (NLP), the quantitative research design is used in this study. Quantitative approach is suitable as it allows measuring, classifying, and statistically analyzing linguistic data based on classroom discourse (Creswell and Creswell, 2018). The research is cross sectional in nature since data were taken at one point in time and analytical in that it involved identification of patterns and assessment of the performance of the computational models. The study, combining NLP techniques with educational research, offers an objective and scalable framework of language use analysis in the actual classroom environments (Jurafsky and Martin, 2020).

Population and Sample

The sample of this research is composed of students and teachers working in ESL classrooms in the universities of the public sector in Southern Punjab. Particularly, the research was carried out in three state-funded universities of Bahawalpur, such as The Islamia University of Bahawalpur, Cholistan University of Veterinary and Animal Sciences (CUVAS), and Sadiq Women University Bahawalpur, as well as the Engineering University of Rahim Yar Khan. These institutions were chosen because of its multilingual learning environment where English and Urdu are widely spoken and used in the academic interaction.

One of the sampling methods was purposive sampling method whereby classrooms where code-switching was commonly witnessed were selected. The sample involved classroom sessions with the involvement of undergraduate learners and the instructors. This methodology made sure that the data collected was pertinent to the research objectives, and of ESL practices in the sampled institutions (Saunders et al., 2019).

Data Collection

The data were recorded in class by recording their interactions, teacher instructions provided and their interaction with the environment during a regular ESL class. Audio recordings were acquired with

previous approval of the participants to be able to comply with the ethical principles. These tapes were recorded in natural classroom talk, such as lecture, discussion and peer interactions.

The audio data obtained was transcribed into written form to enable them to be analyzed computationally. Transcription was done using standardized conventions to reflect accurately the language alternation between English and Urdu. The ecological validity of the study is maximized by the utilization of actual classroom data because it reflects real-life language use as opposed to controlled experimental conditions (Macaro, 2001).

NLP Techniques Used

This research employs several NLP methods to automatically identify and analyze the patterns of code-switching. To begin with, token- level language identification was done to categorize each word in the text as one of the languages (English or Urdu). This technique can accurately detect the switching points in the sentences (Solorio and Liu, 2008).

Second, n-gram analysis was applied to determine common patterns of using language and language contact. N-grams help in capturing contextual dependencies and common switching structures within classroom discourse (Jurafsky & Martin, 2020).

Third, machine learning classification models were used to automate the process of detecting code-switching. These models were trained on annotated data to identify linguistic patterns and more accurately predict language switches. The rule-based and machine learning methods are combined to increase the rigor and quality of the analysis (Rijhwani et al., 2017).

Data Analysis Techniques

A mix of statistical and computational methods was used to analyze the transcribed and processed data. Frequency analysis was used to find out how common code-switching is in various classroom settings, e.g. lectures, discussions, and explanations. This gives a quantitative estimate of the patterns of language alternation.

Moreover, pattern recognition methods were used to determine the typical ways of switching codes, such as intra-sentential and inter-sentential switching. These patterns were examined to comprehend the functional role of code-switching in assisting the communication and learning processes.

Lastly, the models of the NLP were tested based on accuracy measures, such as precision, recall, and F1-score. These measures determine how effective the models are in identifying correctly the cases of code-switching, which makes sure that the results are reliable (Jurafsky and Martin, 2020).

Ethical Considerations

The research process was carried out under strict ethical guidelines. The purpose of the study was explained to the participants and informed consent was signed before the classroom sessions were recorded. Anonymity and confidentiality were ensured by deleting any identifiable data in the dataset. Data collection was done exclusively on academic grounds, within the framework of accepted research ethics standards (Saunders et al., 2019).

RESULTS AND ANALYSIS

Table 1: Frequency of Code-Switching Across Contexts

Category	Mean	Percentage	Std. Dev
Teacher Speech	3.8	65%	0.62
Student Interaction	3.5	55%	0.70
Group Discussion	4.1	70%	0.58

The results show that the distribution of the code-switching is context dependent, with the highest occurrence in group discussion (70%), teacher speech (65%) and student interaction (55%). This trend indicates that code-switching is not a process done randomly but a strategically planned process in cognitively demanding and socially interactive situations. The increased frequency of group discussions is an indicator to show that learners are using bilingual resources to negotiate meaning, lessen cognitive load, and keep the flow of interaction, which is consistent with interactionist theory (Long, 1996). The comparatively lower percent in the student only interaction (55) means that the student might still be having some linguistic insecurity or or limited proficiency that would limit spontaneous switching in such an environment as compared to guided interaction.The smaller standard deviation in group discussion (0.58) also shows consistency in switching behavior, indicating that collaborative activities naturally cause bilingual communication patterns..

Table 2: Functions of Code-Switching

Function	Mean	Percentage	Std. Dev
Explanation	3.9	60%	0.65
Clarification	3.7	55%	0.68
Classroom Management	3.2	40%	0.72

The prevalence of explanation (60%), clarification (55%), and proves the idea that code-switching is more of a pedagogical scaffold than a random linguistic habit. Teachers tactically switch to Urdu so that conceptual clarity and understanding can be achieved, especially when addressing abstract or complex material. The relatively low usage in classroom management (40% suggest that regular or routine communication is still largely English-dominant signifying functional separation between instructional and administrative discourse. This functional differentiation advocates Macaro (2001), who states that code-switching is not a replacement to use the target language.

Table 3: Types of Code-Switching

Type	Percentage
Intra-sentential	50%
Inter-sentential	35%
Tag-switching	15%

The high percentage of intra-sentential switching (50%) attests to the relatively high degree of bilingual competence (in this type, the ability to switch languages within the same syntactic structure is required). This echoes the linguistic flexibility and cognitive control by both teachers and students. The fact that inter-sentential switching occurs (35%), indicates that inter-sentential switching is also used at discourse boundaries, and is often done to switch between explanation and elaboration. The low tag-switching (15%) implies that switching is functional and meaningful instead of stylistic or habitual.

Table 4: NLP Model Performance

Model	Accuracy	Precision	Recall	F1 Score
Rule-Based	75%	0.74	0.72	0.73
Machine Learning	85%	0.84	0.83	0.83
Hybrid Model	90%	0.89	0.88	0.88

The comparison of performance shows that hybrid models are more effective than rule-based and standalone machine learning models and are the most effective with the highest accuracy (90%). It means that the process of code-switching detection could be a complicated language task that could be improved by using both linguistic knowledge based on rules and the ability to learn by statistics. The comparatively lower recall in rule-based systems implies that it has less adaptability to the variation of real-world languages, but machine learning models, in turn, are more adaptable to the variation of the real-world languages but may not respond to the context. These shortcomings are well balanced by the hybrid model and hence it is the most appropriate mode to apply to educational NLP applications.

Table 5: Correlation Analysis

Variables	Code-Switching	Engagement	Comprehension
Code-Switching	1	0.62	0.68
Engagement	0.62	1	0.70
Comprehension	0.68	0.70	1

The results of the correlation show that there is a strong positive relationship between code-switching and both engagement ($r = 0.62$) and comprehension ($r = 0.68$). This implies that, code-switching is not only a compensatory strategy but proactively leads to improved learning outcomes. The most intense correlation between engagement and comprehension ($r = 0.70$) suggests that cognitive and behavioral aspects of learning are mutually dependent and that code-switching indirectly enhances this relationship by helping them to understand.

Table 6: Regression Analysis

Variable	Beta	t-value	Sig.
Code-Switching	0.58	7.10	0.001
Engagement	0.42	5.60	0.002

The regression analysis supports the fact that code-switching is a strong predictor of learning outcomes (= 0.58), and the effect of code-switching is stronger than the effect of engagement. This means that the flexibility of language directly impacts comprehension, and not just facilitating engagement. The high t-values and low p-values (<0.01) attest to the strength of the model that supports the view that code-switching is a causal factor in enhancing learning effectiveness.

Table 7: ANOVA Results

Source	F	Sig.
Model	48.25	0.000

The results of the ANOVA show that the model is statistically significant, meaning that the independent variables all explain a significant amount of variance in learning results. This supports the legitimacy of the suggested framework connecting code-switching, engagement, and comprehension

Table 8: Hypothesis Testing

Hypothesis	Statement	Result
H1	Code-switching occurs frequently	Supported
H2	NLP detects code-switching effectively	Supported
H3	Code-switching enhances learning outcomes	Supported

Each of the hypotheses is substantiated and empirically validates both linguistic and computational components of the study. The findings confirm that not only is code-switching widespread, but that it is, in fact, pedagogically and computationally detectable.

DISCUSSION

The results of the study are very strong empirical evidence that code-switching is not a peripheral or incidental aspect of ESL classrooms in Pakistan. The much higher rate of occurrence of code-switching in group discussions allows seeing that bilingual language use becomes most conspicuous in contexts that involve interaction and are cognitively challenging, whereby learners actively negotiate meaning. This is in line with the Interactionist Theory (Long, 1996) which postulates that acquisition of language takes place through meaningful interaction and negotiation. Code-switching in these environments seems to serve as a cognitive support tool, allowing learners to overcome linguistic barriers and maintain the dialogue without a breakdown.

Moreover, the high positive correlations between the use of code switching, engagement, and comprehension indicate that the use of the language code switching not only leads to behavioral participation, but also to cognitive processing and understanding. These results build on the work by Macaro (2001), and Shamim (2008), which emphasized the facilitative effect code-switching has, by providing quantitative data demonstrating the significant positive impact code-switching produces on learning. Notably, the results of the regression show that there is a direct and significant predictive effect of code-switching on comprehension, meaning that the role of the former extends beyond engagement and has a direct and significant predictive effect on comprehension. This helps to support the argument that code-switching is a scaffolding tool, which is in line with the sociocultural theory of learning as proposed by Vygotsky (1978), where language is a mediational tool of learning.

The functional analysis further indicates that code-switching is mostly applied in terms of explanation and clarification, which makes it more instructional. This implies that teachers effectively use bilingual discourse to make complex ideas easy to understand and to be inclusive even to learners who have different levels of proficiency. Their relatively less frequent use in classroom management reflects that code-switching is a goal-oriented process as opposed to a weakness in language use.

Computationally, the research gives a meaningful contribution in presenting the usefulness of NLP techniques in the analysis of classroom talk. The fact that the hybrid model (90% accuracy) has been much more reliable than either of the two approaches alone. This is not the first to demonstrate the intricacy of data in code switching and the necessity of integrative methods (Solorio and Liu, 2008; Rijhwani et al., 2017). The high precision of this study confirms the possibility of using NLP tools in real-life educational settings, not only on controlled datasets but also on the real classroom interactions.

Furthermore, the fact that NLP was successfully applied in this case is indicative of the potential of the method as a scalable and objective alternative to traditional qualitative approaches. In contrast to the

manual observation and transcription which are both time-consuming and can be easily biased, the NLP-based analysis allows large quantities of data to be processed and the language patterns to be consistently classified. This is a methodological improvement of the research on language education, which fills the gap between linguistic theory and the analysis of data.

All in all, it can be concluded that code-switching can be regarded as a tool that helps to increase understanding, communication, and interaction. Simultaneously, the combination of the NLP methods provides a strong paradigm of systematically researching multilingual classroom dynamics. Collectively, these results support the idea of the balanced and informed approach, in which the code-switching is to be implemented strategically and to be technologically analyzed to enhance the results of both teaching and learning in ESL.

CONCLUSION

This paper aimed at analyzing the code-switching behavior in ESL classrooms in Pakistan and to determine whether NLP methods are effective in automatically detecting the behavior. The results show that code-switching is common, predictable, and intentional with the most frequent cases occurring in the context of interactions, e.g. in group discussions. It is primarily applied to explain and clarify and, therefore, is a pedagogical scaffold and not a random, compensatory behavior.

Quantitative analyses also indicate that code-switching is significantly positively correlated with engagement and comprehension, and regression analysis confirms that code-switching makes a strong predictor of learning outcomes. These findings conform to interactionist and sociocultural views and indicate that the discourse of bilingualism facilitates meaning making and reduces cognitive load in the second language learning environment (Long, 1996; Vygotsky, 1978).

In terms of the methodology, the research will demonstrate that NLP- especially hybrid methods based on a combination of rule-based and machine learning techniques- can effectively identify code-switching in natural classroom data. This indicates that it is possible to use computational tools to transcend the small scale, subjective analyses of classroom discourse to scalable and objective analysis.

Overall, the study also contributes to language education as well as computational linguistics because it demonstrates that code-switching promotes learning, and that NLP offers a strong framework to study code-switching in a systematic way. Instead of negative discouragement of code-switching, a less confrontational strategy of acknowledging the pedagogical usefulness of code-switching is justified in multilingual ESL settings.

RECOMMENDATIONS

According to the results, the following recommendations are offered:

To start with, teachers must embrace a planned and deliberate application of code-switching in ESL classrooms. As the results suggest that it has a very strong positive impact on comprehension and engagement, teachers can intentionally use the bilingual explanations to support the learners when introducing a complicated topic, though English should be used as a primary instructional language.

Second, teachers need to be trained to employ good bilingual pedagogy. Instructors should be guided by professional development programs to know when and how to use code-switching in an appropriate way that will not be excessive and counter-productive (Macaro, 2001).

Third, learning institutions ought to think about adopting NLP-based tools to analyze and monitor classrooms. These tools can be used to assess instructional practices, determine patterns of language use, and create data-driven insights to enhance instructional approaches.

Fourth, curriculum developers must also include multilingual learning models as they provide resources but not obstacles to learning. This has the potential to make the process more inclusive and accommodate learners with different levels of proficiency.

Fifth, further studies are needed to consider the development of more advanced NLP models, such as deep learning and speech-based systems, to achieve more accurate detection and expand analysis to oral discourse in real time. Generalizability would also be enhanced by expanding the dataset to various parts of the world and different levels of education.

Lastly, policymakers ought to encourage a balanced language policy to recognize the reality of multilingual classrooms and still focus on promoting English proficiency. By taking a more flexible approach and combining the two languages, more successful and contextually appropriate language education can be achieved.

REFERENCES

Barman, U., Das, A., Wagner, J., & Foster, J. (2014). Code mixing: A challenge for language identification in social media. In *Proceedings of the first workshop on computational approaches to code switching* (pp. 13–23).

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approach* (5th ed.). SAGE Publications.

García, O., & Wei, L. (2014). *Translanguaging: Language, bilingualism and education*. Palgrave Macmillan.

Gumperz, J. J. (1982). *Discourse strategies*. Cambridge University Press.

Jurafsky, D., & Martin, J. H. (2020). *Speech and language processing* (3rd ed.). Draft version. https://web.stanford.edu/~jurafsky/slp3/

Khan, M. A., & Mansoor, S. (2017). Code-switching in Pakistani ESL classrooms: A sociolinguistic perspective. *Journal of Language Studies, 12*(2), 45–60.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press.

Macaro, E. (2001). Analysing student teachers' code-switching in foreign language classrooms: Theories and decision making. *The Modern Language Journal, 85*(4), 531–548. https://doi.org/10.1111/0026-7902.00124

Malik, S., Ahmed, R., & Tariq, H. (2023). Classroom discourse and bilingual practices in Pakistan. *Pakistan Journal of Education, 40*(1), 75–92.

Nguyen, D., & Dogruöz, A. S. (2013). Word level language identification in online multilingual communication. In *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing* (pp. 857–862).

Poplack, S. (1980). Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching. *Linguistics, 18*(7–8), 581–618. https://doi.org/10.1515/ling.1980.18.7-8.581

Pratapa, A., Bhat, G., Choudhury, M., & Bali, K. (2018). Language modeling for code-mixing: The role of linguistic theory. In *Proceedings of ACL* (pp. 112–122).

Rahman, T. (2010). *Language policy, identity, and religion: Aspects of the civilization of the Muslims of Pakistan and North India*. Oxford University Press.

Rijhwani, S., Bali, K., & Bansal, M. (2017). Code-switching detection in multilingual text. In *Proceedings of the CoNLL 2017 Shared Task* (pp. 31–39).

Shamim, F. (2008). Trends, issues and challenges in English language education in Pakistan. *Asia Pacific Journal of Education, 28*(3), 235–249. https://doi.org/10.1080/02188790802267324

Solorio, T., & Liu, Y. (2008). Learning to predict code-switching points. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 973–981).

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Winata, G. I., Madotto, A., Wu, C.-S., & Fung, P. (2019). Code-switching language modeling using multilingual embeddings. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 2947–2952). https://doi.org/10.18653/v1/P19-1284