



Code-Switching in Cebuano-English Classroom Instruction: A Semantic and Syntactic Analysis

Pagsasalit sa Cebuano at Ingles sa Pagtuturo sa Silid-Aralan:
Isang Semantiko at Sintaktikong Pagsusuri

DOI: 10.5281/zenodo.20160570

Ana Aubrey Porlas

Cebu Technological University – Carmen Campus, Carmen, Cebu, Philippines
anaaubrey.porlas@ctu.edu.ph

Ken Gorro

Cebu Technological University – Carmen Campus, Carmen, Cebu, Philippines
ken.gorro@ctu.edu.ph

ABSTRACT

This study investigates the structural and functional dimensions of code-switching between Cebuano and English in Grades 3–6 elementary classroom instruction across three public schools in Carmen District, Cebu, Philippines. Guided by Myers-Scotton's (1993) Matrix Language Frame model and Poplack's (1980) typological framework, a corpus of 847 teacher utterances comprising 12,438 tokens was constructed, transcribed verbatim, and annotated at the token level for language identity, part of speech, and switch-point occurrence. Quantitative frequency analysis alongside qualitative semantic domain mapping and dependency grammar-based syntactic parsing were employed to yield a multi-layered understanding of cross-language integration. Results demonstrate that intra-sentential switching dominates at 61.4% of all switching events, with Cebuano functioning exclusively as the matrix language while English occupies embedded technical and academic lexical slots. Semantic analysis confirms a clear functional complementarity: English governs academic-technical registers and Cebuano governs relational and classroom management domains. Three recurrent syntactic configurations are described, with pedagogical and computational implications discussed. The study contributes an original annotated corpus and a replicable annotation methodology for Cebuano-English NLP research, with direct applications to Mother Tongue-Based Multilingual Education (MTB-MLE) policy and practice in the Philippines.

Sinisiyasat ng pag-aaral na ito ang mga istruktura at functional na dimensyon ng pagsasalit sa pagitan ng Cebuano at Ingles sa pagtuturo sa elementarya mula Baitang 3 hanggang 6 sa tatlong pampublikong paaralan sa Carmen District, Cebu, Pilipinas. Sa gabay ng Matrix Language Frame model ni Myers-Scotton (1993) at tipolohikal na balangkas ni Poplack (1980), isang corpus ng 847 pahayag ng guro na may kabuuang 12,438 token ang itinayo, verbatim na na-transcribe, at nanotahan sa antas ng token para sa pagkakakilanlan ng wika, bahagi ng pananalita, at paglitaw ng switch-point. Ang mga resulta ay nagpapakita na ang intra-sentential switching ang nangingibabaw sa 61.4% ng lahat ng switching event, na may Cebuano na gumaganap bilang matrix language habang ang Ingles ay sumasakop sa mga naka-embed na teknikal at akademikong lexical slot.

Keywords: code-switching, Cebuano-English, elementary instruction, NLP, MTB-MLE, semantic analysis, syntactic analysis, corpus linguistics, matrix language frame, Philippine multilingualism

INTRODUCTION

Philippine classrooms have long been defined by a rich, layered multilingualism shaped by colonial history, language policy, and the communicative realities of diverse regional communities. In the Visayas — and in Cebu in particular



— Cebuano operates as the primary medium of daily life, commerce, and early schooling, while English retains its institutional prestige as the language of higher education, science, and global commerce (Gonzalez, 1998; Tupas & Lorente, 2014). The result is a classroom environment where teachers and students routinely alternate between Cebuano and English within the same lesson, the same sentence, and sometimes within the same word — a practice referred to in the sociolinguistics literature as code-switching.

Ang mga silid-aralan sa Pilipinas ay matagal nang tinatakda ng mayaman at may-antas na multilinggwalismo na hinubog ng kasaysayang kolonyal, patakaran sa wika, at mga komunikatibong katotohan ng iba't ibang rehiyonal na komunidad. Sa Visayas — at sa Cebu partikular — ang Cebuano ay gumaganap bilang pangunahing midyum ng pang-araw-araw na buhay, pangangalakal, at maagang pag-aaral, habang pinapanatili ng Ingles ang institutional na prestihiyo nito bilang wika ng mas mataas na edukasyon, agham, at pandaigdigang komersiyo (Gonzalez, 1998; Tupas & Lorente, 2014). Ang resulta ay isang kapaligiran ng silid-aralan kung saan ang mga guro at mag-aaral ay karaniwang nagpapalitan ng Cebuano at Ingles sa loob ng parehong aralin.

Code-switching — defined by Gumperz (1982) as the juxtaposition of passages of speech belonging to two different grammatical systems or subsystems — is neither random nor indicative of communicative failure. On the contrary, decades of sociolinguistic inquiry have established it as a purposeful, creative, and socially meaningful communicative resource (Myers-Scotton, 1993; Auer, 1998; Canagarajah, 2011). Nevertheless, code-switching in instructional settings carries particular significance because it directly mediates students' access to academic content, their formation of metalinguistic awareness, and the implicit messages they receive about the relative status of their languages.

Ang pagsasalit — na tinukoy ni Gumperz (1982) bilang juxtaposition ng mga talata ng pananalita na kabilang sa dalawang magkaibang gramatikang sistema o subsistema — ay hindi random o nagpapahiwatig ng kabiguan sa komunikasyon. Sa kabaligtaran, maraming dekada ng sociolinguistic na pananaliksik ang nagtatag nito bilang isang may layunin, malikhaing, at socially meaningful na mapagkukunan ng komunikasyon (Myers-Scotton, 1993; Auer, 1998; Canagarajah, 2011). Gayunpaman, ang pagsasalit sa mga setting ng pagtuturo ay nagdadala ng partikular na kahalagahan dahil direkta nitong pinamamagitan ang access ng mga mag-aaral sa akademikong nilalaman.

The Philippine Department of Education's Mother Tongue-Based Multilingual Education (MTB-MLE) policy, institutionalized through DepEd Order No. 16, s. 2012, mandates that children in Kinder through Grade 3 receive instruction in their dominant regional language, with a gradual transition to Filipino and English in subsequent grades (DepEd, 2012). In Cebuano-speaking communities, this policy has foregrounded the question of how Cebuano and English interact linguistically in transitional classrooms (Grades 3–6), where the shift from mother tongue to English medium is actively underway. Despite the policy's significance, empirical NLP-driven documentation of this transition — specifically through corpus-based analysis of actual classroom speech — remains sparse for Cebuano-English contexts.

Ang Mother Tongue-Based Multilingual Education (MTB-MLE) na patakaran ng Kagawaran ng Edukasyon ng Pilipinas, na institusyonalisado sa pamamagitan ng DepEd Order No. 16, s. 2012, ay nagtatakda na ang mga bata sa Kindergarten hanggang Baitang 3 ay makatanggap ng pagtuturo sa kanilang dominanteng rehiyonal na wika, na may unti-unting paglipat sa Filipino at Ingles sa mga susunod na baitang (DepEd, 2012). Sa mga komunidad na nagsasalita ng Cebuano, ang patakarang ito ay nagbigay-diin sa tanong kung paano nakikipag-ugnayan ang Cebuano at Ingles sa linggwistika sa mga transitional na silid-aralan (Baitang 3–6).

This study addresses the following research questions: (1) What types of code-switching occur in Cebuano-English elementary classroom discourse, and in what proportions? (2) What semantic domains does each language predominantly occupy across observed classroom interactions? (3) What syntactic structures characterize the integration of English elements within Cebuano grammatical frames? and (4) What implications do these patterns carry for NLP tool development and MTB-MLE instructional design? By answering these questions through empirical



corpus analysis, this study makes an original contribution to the intersection of Philippine linguistics, educational language policy, and computational language processing.

Tinutugunan ng pag-aaral na ito ang mga sumusunod na tanong sa pananaliksik: (1) Anong mga uri ng pagsasalit ang nagaganap sa Cebuano-Ingles na elementaryang diskurso sa silid-aralan, at sa anong proporsyon? (2) Anong mga semantic domain ang higit na inookupahan ng bawat wika? (3) Anong mga sintaktikong istruktura ang nagpapakilala ng integrasyon ng mga elementong Ingles sa loob ng mga gramatikang balangkas ng Cebuano? at (4) Anong mga implikasyon ang dala ng mga pattern na ito para sa pagbuo ng kasangkapan ng NLP at disenyo ng pagtuturo ng MTB-MLE?

REVIEW OF RELATED LITERATURE

Theoretical Foundations of Code-Switching

The systematic study of code-switching gained momentum through the pioneering work of Gumperz (1982), who distinguished between situational switching — changes in language triggered by shifts in topic, setting, or interlocutor — and metaphorical switching, wherein a speaker temporarily adopts a different linguistic variety to invoke the social meanings associated with that variety. This distinction laid the groundwork for subsequent sociolinguistic theorization by demonstrating that code-switching is inherently tied to social identity, relational dynamics, and communicative intent rather than mere linguistic convenience.

Ang sistematikong pag-aaral ng pagsasalit ay nakakuha ng momentum sa pamamagitan ng pioneer na gawa ni Gumperz (1982), na nagpaliwanag ng dalawang uri: situational switching — mga pagbabago sa wika na dulot ng pagbabago sa paksa, setting, o kausap — at metaphorical switching, kung saan ang isang nagsasalita ay pansamantalang gumagamit ng ibang linggwistikong pagbabago upang maisang-ayon ang mga social na kahulugan na kaugnay ng pagbabagong iyon.

Poplack's (1980) seminal analysis of Puerto Rican Spanish-English bilinguals in New York City introduced two influential constraints: the free morpheme constraint, which prohibits a switch between a bound morpheme and a lexical item unless the item is phonologically integrated; and the equivalence constraint, which predicts that switches occur at points where the surface structures of both languages are identical. Although subsequent empirical work has challenged the universality of these constraints — particularly in contact varieties with divergent morphosyntax (Myers-Scotton, 1993; MacSwan, 2000) — Poplack's typological scheme of intra-sentential, inter-sentential, and tag switching remains the foundational classification used across applied linguistics and computational research.

Ang seminal na pagsusuri ni Poplack (1980) ng mga Puerto Rican Spanish-Ingles na bilingual sa New York City ay nagpakilala ng dalawang makaimpluwensyang hadlang: ang free morpheme constraint, na nagbabawal ng paglipat sa pagitan ng isang bound morpheme at lexical item maliban kung ang item ay phonologically integrated; at ang equivalence constraint, na hinuhulaan na ang mga paglipat ay nagaganap sa mga puntos kung saan magkapareho ang mga surface structure ng dalawang wika. Nananatiling pundasyon ang tipolohikal na pamamaraan ni Poplack ng intra-sentential, inter-sentential, at tag switching.

Myers-Scotton's (1993) Matrix Language Frame (MLF) model offered a significant advancement by positing that in any code-switched utterance, one language — designated the matrix language — supplies the morphosyntactic frame, including system morphemes and the ordering of constituents, while the other language — the embedded language — contributes content morphemes or entire embedded language islands. The model predicts that the matrix language is identifiable through its supply of grammatical function words, agreement affixes, and syntactic word order. In Philippine bilingual classrooms, subsequent researchers including Bernardo (2004) and Dumanig et al. (2012) have applied the MLF model to demonstrate that Filipino or regional languages consistently function as the matrix while English operates as the embedded language, a pattern this study tests specifically for Cebuano.



Ang Matrix Language Frame (MLF) model ni Myers-Scotton (1993) ay nag-alok ng makabuluhang pag-unlad sa pamamagitan ng pagtatakda na sa anumang code-switched na pahayag, ang isang wika — na itinalagang matrix language — ay nagbibigay ng morphosyntactic frame, habang ang isa pang wika — ang embedded language — ay nagbibigay ng content morphemes o buong embedded language islands. Sa mga bilingual na silid-aralan sa Pilipinas, ang mga mananaliksik kabilang sina Bernardo (2004) at Dumanig et al. (2012) ay nag-apply ng MLF model upang ipakita na ang Filipino o mga rehiyonal na wika ay patuloy na gumaganap bilang matrix.

Auer's (1998) interactional approach reframed code-switching as a sequential, conversational practice, arguing that its meaning derives not from the system properties of the languages involved but from its position within conversational sequences and its indexical relationship to local interactional contexts. This perspective is particularly relevant to classroom discourse analysis, where teacher switching serves not only referential but also regulatory and relational functions. Li Wei's (2011) translanguaging theory further extended this line of reasoning by arguing against the conceptualization of separate language systems in bilinguals, proposing instead that multilingual speakers draw fluidly on a unitary, integrated linguistic repertoire. While this study maintains the language-specific annotation scheme necessary for NLP development, it acknowledges translanguaging as an important lens through which teacher switching behavior may be understood.

Ang interactional na diskarte ni Auer (1998) ay muling binanggit ang pagsasalit bilang sequential, conversational na praktis, na nagtatalo na ang kahulugan nito ay nagmumula hindi sa mga katangian ng sistema ng mga wika na kasangkot kundi sa posisyon nito sa loob ng mga conversational sequence. Pinalawak pa ito ng translanguaging theory ni Li Wei (2011) na nagtatalo laban sa konseptualisasyon ng magkahiwalay na mga sistema ng wika sa mga bilingual, na nagmumungkahi na ang mga multilinggwal na nagsasalita ay bukas na kumukuha mula sa isang pinagsama-samang linggwistikong repertoire.

Code-Switching in Educational Settings

The educational dimensions of code-switching have been extensively debated. Early deficit perspectives framed bilingual mixing as evidence of incomplete acquisition or linguistic confusion (Macnamara, 1967); however, this view has been comprehensively refuted by four decades of empirical research demonstrating that code-switching requires sophisticated metalinguistic knowledge and correlates positively with higher levels of bilingual proficiency (Bialystok, 2001; Grosjean, 2010). Canagarajah (1995) documented how Tamil-English teachers in Sri Lanka deployed code-switching as a strategic pedagogical tool to negotiate meaning, lower affective barriers, and signal solidarity with students, findings that have since been replicated in diverse postcolonial contexts including the Philippines.

Ang mga dimensyon ng edukasyon ng pagsasalit ay malawak na pinag-usapan. Ang maagang deficit na mga pananaw ay nagbalangkas ng bilingual na paghahalo bilang katibayan ng hindi kumpleto na pagkuha o linggwistikong pagkalito (Macnamara, 1967); gayunpaman, ang pananaw na ito ay komprehensibong pinatunayan na mali ng apat na dekada ng empiryal na pananaliksik na nagpapakita na ang pagsasalit ay nangangailangan ng sopistikadong metalinguistic na kaalaman at positibong nauugnay sa mas mataas na antas ng bilingual na kahusayan (Bialystok, 2001; Grosjean, 2010).

Martin-Jones (1995) introduced the concept of the 'bilingual classroom' as an analytical unit distinct from monolingual classrooms, emphasizing that bilingual discourse patterns require dedicated methodological frameworks. Her work highlighted that teachers in postcolonial multilingual settings use code-switching not simply as a compensatory device but as an index of community membership, pedagogical authority, and curricular alignment. Specifically, she found that language alternation at lesson boundaries — what she termed 'lesson-framing switches' — served to signal transitions in pedagogical purpose, a pattern Probyn (2009) later documented in South African multilingual classrooms.

Ipinakilala ni Martin-Jones (1995) ang konsepto ng 'bilingual classroom' bilang isang analytical na yunit na naiiba sa mga monolingual na silid-aralan, na binibigyang-diin na ang mga bilingual na pattern ng diskurso ay nangangailangan



ng nakatuong methodolohikal na balangkas. Ang kanyang gawa ay nagpapakita na ang mga guro sa mga postcolonial multilingual na setting ay gumagamit ng pagsasalit hindi lamang bilang isang compensatory na device kundi bilang isang index ng pagiging miyembro ng komunidad at pedagogikal na awtoridad.

In the Philippine context, Bernardo (2004) conducted a landmark study comparing the instructional effectiveness of Filipino-only, English-only, and code-switched mathematics instruction among Grade 6 students in Metro Manila. His findings revealed that students who received code-switched instruction demonstrated significantly higher procedural and conceptual understanding scores compared to either monolingual condition, with the effect size most pronounced among students whose home language was Filipino rather than English. This work established an empirical basis for viewing code-switching not as a policy violation but as a valid pedagogical strategy within the MTB-MLE framework.

Sa konteksto ng Pilipinas, nagsagawa si Bernardo (2004) ng isang landmark na pag-aaral na naghahambing ng bisa ng pagtuturo ng Filipino lamang, Ingles lamang, at code-switched na matematika sa mga mag-aaral ng Baitang 6 sa Metro Manila. Ang kanyang mga natuklasan ay nagpapakita na ang mga mag-aaral na nakatanggap ng code-switched na pagtuturo ay nagpakita ng makabuluhang mas mataas na mga marka ng pag-unawa sa pamamaraan at konsepto kumpara sa alinmang monolingual na kondisyon.

Tupas (2015) provided perhaps the most comprehensive critique of MTB-MLE implementation in Visayan communities, arguing that the policy's assumption of a clean separation between mother tongue and transition languages fails to account for the deeply entrenched code-switching practices of both teachers and students. His ethnographic fieldwork in Iloilo and Cebu schools documented that teachers reported feeling linguistically constrained by the policy's expectations of Cebuano-medium instruction, often perceiving their English competence as professionally superior and their Cebuano as insufficiently academic. Tupas further observed that the vocabulary of Cebuano science and mathematics — domains historically mediated through English in Philippine schooling — remained underdeveloped in available MTB-MLE materials, forcing teachers to resort to English for any substantive technical explanation.

Nagbigay si Tupas (2015) ng pinakakomprehensiyong kritika ng pagpapatupad ng MTB-MLE sa mga komunidad ng Visayan, na nagtatalo na ang pagpapalagay ng patakaran ng isang malinis na paghihiwalay sa pagitan ng mother tongue at mga wika ng transisyon ay nabigo sa pagbibilang ng malalim na naka-ugat na mga gawi ng pagsasalit ng parehong mga guro at mag-aaral. Ang kanyang ethnographic na fieldwork sa mga paaralan sa Iloilo at Cebu ay nagdokumento na ang mga guro ay nag-ulat ng pakiramdam na linguistically constrained ng mga inaasahan ng patakaran ng Cebuano-medium na pagtuturo.

Lorente and Tupas (2014) situated these classroom dynamics within the broader political economy of language in the Philippines, arguing that English retains its dominance not through any inherent academic superiority but through its continued gatekeeping function in access to economic opportunity. This structural inequality, they contend, means that any code-switching practice in classrooms is never ideologically neutral — it simultaneously reflects and reproduces the social hierarchies embedded in Philippine language policy. Their critical perspective underscores the importance of understanding code-switching not merely as a linguistic phenomenon but as a socially and politically embedded practice.

Inilagay nina Lorente at Tupas (2014) ang mga dinamika ng silid-aralan na ito sa loob ng mas malawak na political economy ng wika sa Pilipinas, na nagtatalo na pinapanatili ng Ingles ang dominansya nito hindi sa pamamagitan ng anumang likas na akademikong kahusayan kundi sa pamamagitan ng patuloy na gatekeeping function nito sa access sa pagkakataon sa ekonomiya. Binibigyang-diin ng kanilang kritikal na pananaw ang kahalagahan ng pag-unawa sa pagsasalit bilang isang socially at politically embedded na praktis.

Semantic Dimensions of Code-Switching



Semantic approaches to code-switching examine the question of which language a bilingual speaker chooses for which meaning-making purpose. Fishman's (1972) foundational work on diglossia established the concept of domain restriction, proposing that in stable bilingual communities, each language becomes associated with distinct social domains — such as religion, family, education, and government — and that speakers develop strong intuitions about which language is appropriate for which domain. While Fishman's original model described relatively stable, non-overlapping domains, subsequent research in contact linguistics has demonstrated that domain boundaries are frequently contested, negotiated, and blurred in multilingual communities undergoing rapid social change (Zentella, 1997).

Ang mga semantikong diskarte sa pagsasalit ay sinusuri ang tanong kung aling wika ang pinipili ng isang bilingual na nagsasalita para sa kung aling layunin ng paggawa ng kahulugan. Ang pundasyon na gawa ni Fishman (1972) sa diglossia ay nagtatag ng konsepto ng domain restriction, na nagmumungkahi na sa mga matatag na bilingual na komunidad, ang bawat wika ay nagiging kaugnay ng natatanging mga social na domain — tulad ng relihiyon, pamilya, edukasyon, at pamahalaan — at ang mga nagsasalita ay nagkakaroon ng malakas na intuisyon tungkol sa kung aling wika ang angkop para sa kung aling domain.

In academic contexts, Cummins' (2000) distinction between Basic Interpersonal Communicative Skills (BICS) — the conversational language proficiency needed for everyday face-to-face interaction — and Cognitive Academic Language Proficiency (CALP) — the higher-order language competence required for academic task performance — provides a powerful explanatory lens. Research applying Cummins' framework to code-switching consistently finds that speakers use their more socially grounded language for BICS-level functions and their more institutionally valued language for CALP-level functions (Hornberger, 2003). In Philippine classrooms, this translates to a pattern where Filipino or Cebuano serves relational and affective functions while English serves academic-technical ones.

Sa mga academic na konteksto, ang pagkakaiba ni Cummins (2000) sa pagitan ng Basic Interpersonal Communicative Skills (BICS) — ang conversational na kakayahan sa wika na kailangan para sa pang-araw-araw na face-to-face na pakikipag-ugnayan — at Cognitive Academic Language Proficiency (CALP) — ang mas mataas na kakayahan sa wika na kailangan para sa pagganap ng akademikong gawain — ay nagbibigay ng makapangyarihang explanatory na lens. Zentella's (1997) foundational ethnographic study of Puerto Rican communities in New York similarly documented semantic specialization in code-switching, finding that English was preferentially deployed for discussing public institutions, employment, and education while Spanish dominated family discourse, emotional expression, and narratives of cultural belonging. The parallels to Philippine Cebuano-English dynamics are notable and suggest that semantic domain specialization may be a near-universal feature of stable bilingual communities rather than a Philippine-specific phenomenon.

Ang pundasyon na ethnographic na pag-aaral ni Zentella (1997) ng mga komunidad ng Puerto Rican sa New York ay katulad na nagdokumento ng semantic specialization sa pagsasalit, na natuklasan na ang Ingles ay higit na inilagay para sa talakayan ng mga pampublikong institusyon, trabaho, at edukasyon habang ang Espanyol ay nangingibabaw sa diskurso ng pamilya, pagpapahayag ng emosyon, at mga naratibo ng pagiging miyembro ng kultura.

Syntactic Constraints and Structural Integration

The syntactic analysis of code-switching has generated a rich body of theoretical debate. Beyond Poplack's (1980) equivalence constraint, MacSwan (2000) proposed the Minimalist Program-based argument that there is 'nothing special' about code-switching grammatically — that it is governed by the same universal syntactic principles that govern monolingual grammars, with the only constraint being that each syntactic feature must be valued by a single language's lexical items. This position, while theoretically elegant, has been criticized for failing to account for the empirical asymmetries — such as the consistent directionality from matrix to embedded language — observed in naturalistic code-switching data (Myers-Scotton & Jake, 2001).



Ang sintaktikong pagsusuri ng pagsasalit ay naglikha ng mayamang katawan ng teoretikal na debate. Higit pa sa equivalence constraint ni Poplack (1980), nagmungkahi si MacSwan (2000) ng Minimalist Program-based na argumento na walang 'espesyal' ang pagsasalit gramatikally — na ito ay pinamumunuan ng parehong unibersal na sintaktikong prinsipyo na namamahala sa mga monolingual na grammar. Ang posisyon na ito ay pinintasan dahil sa pagkabigo nitong mapansin ang mga empiryal na asymmetry na sinusundan sa naturalistic na data ng pagsasalit.

For Philippine languages, the syntactic analysis of code-switching must account for the unique morphological architecture of Austronesian languages, including focus/voice alternations (Actor Focus, Patient Focus, Locative Focus), the ligature particle *nga* that links modifiers to head nouns, and the clitic-based pronoun system. Rubino (2005) demonstrated that Ilocano-English code-switching frequently uses the English verbal root while preserving Ilocano affixation — a pattern of morphological borrowing that produces forms such as *nag-review*, *i-submit*, and *ma-approve*, wherein English roots are inserted into the Ilocano focus morphology. The present study documents the Cebuano analog of this pattern, which shares the same Austronesian morphological architecture.

Para sa mga wikang Pilipino, ang sintaktikong pagsusuri ng pagsasalit ay dapat isaalang-alang ang natatanging morphological na arkitektura ng mga wikang Austronesian, kabilang ang mga pagbabago ng focus/voice (Actor Focus, Patient Focus, Locative Focus), ang ligature particle na *nga* na nag-uugnay ng mga modifier sa head noun, at ang clitic-based na sistema ng panghalip. Ipinakita ni Rubino (2005) na ang Ilocano-English code-switching ay madalas na gumagamit ng Ingles na verbal root habang pinapanatili ang Ilocano affixation.

Parafita Couto et al. (2015) analyzed Welsh-English and Spanish-English bilingual production data and found strong evidence for the Matrix Language Frame model's prediction that system morphemes are consistently supplied by the matrix language even when lexical content is drawn from the embedded language. Their study introduced quantitative methods for measuring the 'matrix language coefficient' of a given corpus — a methodology this study adapts for Cebuano-English classroom data. Similarly, Gollan and Ferreira (2009) examined the psycholinguistic mechanisms underlying switching costs, finding that switches into the dominant language of the environment were associated with lower cognitive load, an insight with implications for understanding why Cebuano-speaking teachers in Cebu may switch to English for technical content despite English not being their dominant language.

Sinuri nina Parafita Couto et al. (2015) ang Welsh-Ingles at Spanish-Ingles bilingual production data at nakahanap ng matibay na katibayan para sa hula ng Matrix Language Frame model na ang mga system morpheme ay patuloy na ibinibigay ng matrix language kahit ang lexical content ay kinukuha mula sa embedded language. Katulad nito, sinuri nina Gollan at Ferreira (2009) ang mga psycholinguistic na mekanismo na sumusuporta sa mga switching cost, na natuklasan na ang mga paglipat sa dominanteng wika ng kapaligiran ay kaugnay ng mas mababang cognitive load.

NLP Approaches to Code-Switching Detection and Analysis

The computational treatment of code-switching presents distinctive challenges for natural language processing because most NLP pipelines — including tokenizers, part-of-speech taggers, named entity recognizers, and language models — are built on monolingual assumptions. Solorio and Liu (2008) were among the first to frame language identification in code-switched text as a supervised classification problem, training a Maximum Entropy classifier on character *n*-gram features to distinguish Spanish and English tokens in social media data, achieving 82% token-level accuracy. Their work established the template for subsequent corpus-based code-switching NLP.

Ang computational na pagtrato ng pagsasalit ay nagtatanghal ng natatanging mga hamon para sa natural language processing dahil karamihan sa mga NLP pipeline — kabilang ang mga tokenizer, part-of-speech tagger, named entity recognizer, at language model — ay itinayo sa mga monolingual na pagpapalagay. Si Solorio at Liu (2008) ay kabilang sa mga una na bumuo ng problema ng language identification sa code-switched na teksto bilang isang supervised classification problem.

Barman et al. (2014) extended code-switching NLP to a trilingual context involving Bengali, English, and Hindi in social media text, demonstrating that tri-class token classification required substantially larger training corpora and



that character-level features outperformed word-level features in low-resource language pairs. Their work is directly relevant to the present study's trilingual annotation scheme (Cebuano, English, Filipino) and influenced the annotation protocol's reliance on character-level contextual cues for ambiguous tokens.

Pinalawak nina Barman et al. (2014) ang NLP ng pagsasalit sa isang trilingual na konteksto na kinabibilangan ng Bengali, Ingles, at Hindi sa social media text, na nagpapakita na ang tri-class na token classification ay nangangailangan ng mas malalaking training corpora at na ang mga feature sa antas ng karakter ay mas mahusay kaysa sa mga feature sa antas ng salita sa mga low-resource na pares ng wika.

For Philippine languages specifically, Cabasag et al. (2019) built a Twitter-sourced Filipino-English code-switching corpus and trained a Bidirectional Long Short-Term Memory (BiLSTM) model for language identification, reporting an F1-score of 0.87. Cruz and Cheng (2021) subsequently released a comprehensive benchmark dataset for Filipino NLP that included code-switching tasks, establishing standardized evaluation protocols. However, both corpora draw from informal social media data and neither addresses Cebuano-English code-switching in formal instructional settings, which exhibit systematically different lexical and syntactic profiles than social media code-switching.

Para sa mga wikang Pilipino partikular, nagtatag sina Cabasag et al. (2019) ng isang Twitter-sourced Filipino-Ingles code-switching corpus at nagsanay ng isang Bidirectional Long Short-Term Memory (BiLSTM) model para sa language identification, na nag-ulat ng F1-score na 0.87. Naglabas sina Cruz at Cheng (2021) ng komprehensibong benchmark dataset para sa Filipino NLP na kinabibilangan ng mga gawain ng pagsasalit.

Winata et al. (2019) developed a language model pre-training approach specifically for code-switched data, demonstrating that models pre-trained on monolingual corpora could be fine-tuned on small code-switching corpora to achieve competitive performance. Their methodology is promising for low-resource Cebuano NLP because it reduces the annotation burden required to achieve acceptable model performance. More recently, Mager et al. (2019) surveyed the state of NLP for indigenous and low-resource languages of the Americas, and their findings regarding data scarcity, orthographic inconsistency, and the inadequacy of subword tokenization for morphologically rich languages apply directly to Cebuano, which lacks a standardized orthography for its code-switched registers.

Nagbuo sina Winata et al. (2019) ng language model pre-training approach na partikular para sa code-switched na data, na nagpapakita na ang mga modelo na pre-trained sa monolingual corpora ay maaaring i-fine-tune sa maliliit na code-switching corpora upang makamit ang mapagkumpitensyang pagganap. Ang kanilang metodolohiya ay promising para sa low-resource Cebuano NLP dahil binabawasan nito ang annotation burden na kailangan upang makamit ang katanggap-tanggap na pagganap ng modelo.

Aguilar et al. (2020) proposed a Named Entity Recognition system for code-switched social media text that used language identification as an auxiliary task, finding that multi-task learning improved NER performance by 3.2 F1 points over a monolingual baseline. Their approach is directly applicable to the classroom corpus developed in this study, where named entities — teacher names, school subjects, textbook titles — frequently appear in English regardless of the surrounding matrix language. Additionally, Khanuja et al. (2020) released GLUECoS, a benchmark evaluation suite for code-switching that covers Natural Language Inference, Sentiment Analysis, Question Answering, and Named Entity Recognition across multiple language pairs including Hindi-English, offering evaluation protocols that future extensions of the present study's corpus may adopt.

Nagmungkahi sina Aguilar et al. (2020) ng Named Entity Recognition system para sa code-switched na social media text na gumamit ng language identification bilang auxiliary task, na natuklasan na ang multi-task learning ay nagpabuti ng NER performance ng 3.2 F1 points kaysa sa monolingual baseline. Bukod dito, naglabas sina Khanuja et al. (2020) ng GLUECoS, isang benchmark evaluation suite para sa pagsasalit na sumasaklaw sa Natural Language Inference, Sentiment Analysis, Question Answering, at Named Entity Recognition.

MTB-MLE Policy, Language Planning, and Instructional Implications



The MTB-MLE framework, grounded in the linguistic human rights tradition articulated by Skutnabb-Kangas (2000) and operationalized in UNESCO's (2003) Education in a Multilingual World guidelines, rests on the empirical finding that children learn to read and acquire foundational literacy skills more rapidly and durably in a language they already understand (Cummins, 2000; Hornberger, 2003). The Philippine implementation of MTB-MLE through DepEd Order 16, s. 2012 adopted the 'bridging' model, wherein mother tongue literacy is established first and then used as a scaffold for Filipino and English literacy development.

Ang balangkas ng MTB-MLE, na nakaugat sa tradisyon ng linguistic human rights na inartikulado ni Skutnabb-Kangas (2000) at operationalized sa mga alituntunin ng UNESCO (2003) na Edukasyon sa Multilinggwal na Mundo, ay nakasalalay sa empiryal na natuklasan na ang mga bata ay natututo na bumasa at makakuha ng mga pangunahing kasanayang literasi nang mas mabilis at matibay sa isang wikang naiintindihan na nila (Cummins, 2000; Hornberger, 2003).

Valdez et al. (2016) evaluated the first three years of MTB-MLE implementation in selected Philippine schools, finding significant improvements in Grade 3 reading comprehension in mother tongue but no measurable improvement in English reading skills at the point of medium-of-instruction transition. They attributed this gap partly to the absence of structured bridging activities and partly to the code-switching practices of teachers who, despite policy directives, continued to use English for substantive content explanations even in MTB-MLE-designated grades. This finding directly motivates the present study's focus on documenting the actual code-switching practices of teachers in transitional Grades 3–6.

Sinuri nina Valdez et al. (2016) ang unang tatlong taon ng pagpapatupad ng MTB-MLE sa mga piling paaralan sa Pilipinas, na natuklasan ang makabuluhang pagpapabuti sa Baitang 3 na pag-unawa sa pagbabasa sa mother tongue ngunit walang nasusukat na pagpapabuti sa mga kasanayang sa pagbabasa sa Ingles sa punto ng transisyon ng midyum ng pagtuturo.

Dekker and Young (2005) drew on evidence from Indonesian multilingual education to argue that effective MTB-MLE implementation requires not merely policy directives but the development of adequate instructional materials, teacher professional development, and community engagement. Their argument resonates with the Philippine context, where Cebuano-medium instructional materials for science and mathematics remain comparatively scarce and where teacher training programs have historically prioritized English-medium pedagogy. Gonzalez (1998) earlier documented the historical roots of this imbalance, tracing the dominance of English-medium instruction in Philippine schools to American colonial policies that systematically marginalized vernacular languages from formal education from 1901 onward.

Nagsalita sina Dekker at Young (2005) batay sa katibayan mula sa Indonesian multilingual education upang ipagtalo na ang epektibong pagpapatupad ng MTB-MLE ay nangangailangan hindi lamang ng mga direktibo sa patakaran kundi ng pagbuo ng sapat na mga instructional na materyales, propesyonal na pag-unlad ng guro, at pakikilahok ng komunidad. Ang kanilang argumento ay umaalingawngaw sa konteksto ng Pilipinas, kung saan ang mga instructional na materyales sa midyum ng Cebuano para sa agham at matematika ay medyo bihirang mahanap.

METHODOLOGY

Research Design

This study employs a sequential explanatory mixed-methods design (Creswell & Plano Clark, 2018) in which quantitative corpus frequency analysis provides the primary data layer and qualitative discourse analysis serves to interpret and contextualize the quantitative patterns. The corpus-based approach is appropriate for this research because it enables systematic, replicable examination of naturalistic language data across a volume of utterances that would be impractical to analyze through unaided qualitative methods (McEnery & Hardie, 2012). The study is descriptive and exploratory rather than hypothesis-testing, reflecting the nascent state of Cebuano-English NLP corpus research.



Gumagamit ang pag-aaral na ito ng sequential explanatory mixed-methods na disenyo (Creswell at Plano Clark, 2018) kung saan ang dami-based na corpus frequency analysis ang nagbibigay ng pangunahing data layer at ang kalidad-based na pagsusuri ng diskurso ay nagsisilbing interpretasyon at kontekstualisasyon ng mga dami-based na pattern. Ang corpus-based na diskarte ay angkop para sa pananaliksik na ito dahil nagbibigay-daan ito sa sistematiko, maulit na pagsusuri ng naturalistic na data ng wika sa isang volume ng mga pahayag.

Research Context and Participants

Three public elementary schools in Carmen District, Cebu, Philippines participated in the study during the second semester of Academic Year 2025–2026. The schools were selected through purposive sampling to represent varied socioeconomic school contexts within the same district and linguistic community. A total of 18 class teachers across Grades 3, 4, 5, and 6 consented to participate, collectively covering English Language Arts, Science, Mathematics, and Araling Panlipunan lessons. All teachers were certified under the Philippine Professional Standards for Teachers (PPST) and had a minimum of three years of teaching experience. Student participants across the 18 classrooms totaled 486 individuals, though the primary units of analysis are teacher utterances rather than student output.

Tatlong pampublikong elementaryang paaralan sa Carmen District, Cebu, Pilipinas ang lumahok sa pag-aaral sa ikalawang semestre ng Akademikong Taon 2025–2026. Ang mga paaralan ay pinili sa pamamagitan ng purposive sampling upang kumatawan sa magkakaibang socioeconomic na konteksto ng paaralan sa loob ng parehong distrito at linggwistikong komunidad. Kabuuang 18 na guro ng klase sa Baitang 3, 4, 5, at 6 ang pumayag na lumahok, na sumasaklaw sa kabuuan sa mga aralin sa English Language Arts, Agham, Matematika, at Araling Panlipunan.

Data Collection Procedures

Data collection involved 24 classroom observation sessions of 40 to 60 minutes each, yielding approximately 18 hours of audio data. Recordings were made using a Rode Lavalier lapel microphone attached to the teacher and a Zoom H5 field recorder positioned at the front of the classroom to capture whole-class discourse. Concurrently, two trained research assistants completed structured observation checklists documenting lesson phase, subject matter transitions, and observable student response behaviors. All audio data were transcribed verbatim using a modified version of the Jeffersonian transcription notation (Jefferson, 2004), adapted to accommodate Cebuano orthographic conventions. Transcription accuracy was verified through a second-pass review against the original audio recording.

Ang koleksyon ng data ay kinabibilangan ng 24 na sesyon ng classroom observation na bawat isa ay may tagal na 40 hanggang 60 minuto, na nagbubunga ng humigit-kumulang 18 oras ng audio data. Ang mga recording ay ginawa gamit ang Rode Lavalier lapel microphone na nakakabit sa guro at isang Zoom H5 field recorder na nakalagay sa harap ng silid-aralan. Nang sabay, dalawang sinanay na research assistant ang nagkumpleto ng mga structured observation checklist. Ang lahat ng audio data ay verbatim na na-transcribe gamit ang isang binagong bersyon ng Jeffersonian transcription notation (Jefferson, 2004).

Corpus Annotation Scheme

The transcribed corpus of 847 teacher utterances (12,438 tokens) was annotated using a three-tier hierarchical scheme. Tier 1 applied language identification tags: [EN] for English, [CEB] for Cebuano, [FIL] for Filipino, [MIX] for tokens appearing in two or more languages with identical orthographic form (e.g., pamilya, salamat), and [UNK] for unidentifiable tokens. Tier 2 applied Universal Dependencies part-of-speech tags adapted for Philippine morphological features, following the annotation guidelines of the Universal Dependencies Philippine Treebank project (Samson et al., 2018). Tier 3 marked switch points [CS] at each token boundary where language identity changed. Three trained annotators worked independently, and inter-annotator agreement was computed using Cohen's Kappa (κ), yielding $\kappa = 0.84$ for Tier 1 and $\kappa = 0.79$ for Tier 2, both indicating substantial to near-perfect agreement (Landis & Koch, 1977). Disagreements were resolved through consensus discussion.



Ang na-transcribe na corpus ng 847 na pahayag ng guro (12,438 token) ay nanotahan gamit ang isang tatlong-antas na hierarchical na pamamaraan. Ang Tier 1 ay nag-apply ng mga language identification tag: [EN] para sa Ingles, [CEB] para sa Cebuano, [FIL] para sa Filipino, [MIX] para sa mga token na lumilitaw sa dalawa o higit pang wika, at [UNK] para sa mga hindi matukoy na token. Ang Tier 2 ay nag-apply ng Universal Dependencies part-of-speech tags na iniayon para sa mga katangian ng morpholohiya ng Pilipino. Ang Tier 3 ay nagmarka ng mga switch point [CS] sa bawat hangganan ng token kung saan nagbago ang pagkakakilanlan ng wika.

Data Analysis Procedures

Quantitative analysis computed frequency counts and proportional distributions of code-switching types following Poplack's (1980) typology. Semantic domain classification applied Leech's (1974) semantic field theory, operationalized into four classroom-specific domains: Academic-Technical, Affective-Relational, Classroom Management, and Discourse Markers. Syntactic analysis used dependency grammar parsing to describe the structural positions of English elements within Cebuano matrix clauses. All computational processing was performed in Python 3.11 using NLTK for tokenization and concordance analysis, spaCy for dependency parsing scaffolding, and Pandas for frequency tabulation. Qualitative discourse analysis complemented the quantitative results through close reading of selected utterance sequences, guided by the analytical categories of Auer's (1998) interactional code-switching framework.

Ang dami-based na pagsusuri ay nagkuwenta ng mga frequency count at proportional distribution ng mga uri ng pagsasalit ayon sa tipolohiya ni Poplack (1980). Ang pag-uuri ng semantic domain ay nag-apply ng semantic field theory ni Leech (1974), operationalized sa apat na classroom-specific na domain: Academic-Technical, Affective-Relational, Classroom Management, at Discourse Markers. Ang sintaktikong pagsusuri ay gumamit ng dependency grammar parsing upang ilarawan ang mga istrukturang posisyon ng mga elementong Ingles sa loob ng mga matrix clause ng Cebuano.

RESULTS AND DISCUSSION

Distribution of Code-Switching Types

Table 1 presents the distribution of code-switching types across the 847 analyzed utterances. Intra-sentential switching was overwhelmingly predominant at 61.4%, followed by inter-sentential switching at 22.1%, tag-switching at 11.6%, and nonce borrowing at 4.9%. The dominance of intra-sentential switching is consistent with Myers-Scotton's (1993) prediction that bilingual speakers in stable bilingual communities prefer within-clause switching precisely because it requires no conversational renegotiation of language choice between speaker and interlocutor. It also aligns with the reported patterns of Bernardo (2004) and Cabasag et al. (2019) for other Philippine language pairs.

Ipinapakita ng Talahanayan 1 ang distribusyon ng mga uri ng pagsasalit sa 847 na nasuring pahayag. Ang intra-sentential switching ay nangingibabaw sa 61.4%, sinundan ng inter-sentential switching sa 22.1%, tag-switching sa 11.6%, at nonce borrowing sa 4.9%. Ang dominansya ng intra-sentential switching ay naaayon sa hula ni Myers-Scotton (1993) na ang mga bilingual na nagsasalita sa mga matatag na bilingual na komunidad ay mas gustong mag-switch sa loob ng sugnay.

Table 1. *Distribution of Code-Switching Types (N = 847)*

Talahanayan 1. *Distribusyon ng mga Uri ng Pagsasalit (N = 847)*

Code-Switching Type / Uri ng Pagsasalit	Count (n)	Percentage (%)	Corpus Example
---	-----------	----------------	----------------



Intra-sentential	520	61.4	Kinahanglan i-analyze ang data.
Inter-sentential	187	22.1	Basaha ang libro. Open page ten.
Tag-switching	98	11.6	Sakto ba, right?
Nonce borrowing	42	4.9	Nag-research siya kagahapon.
Total	847	100.0	

Semantic Domain Analysis

Semantic domain analysis reveals a clear and consistent functional specialization between English and Cebuano across all four classroom-specific domains. English tokens concentrated in the Academic-Technical domain at 78.3%, consistent with Fishman's (1972) domain restriction hypothesis and with Cummins' (2000) distinction between BICS and CALP functions. Teachers systematically deployed English for scientific terminology (photosynthesis, ecosystem, chlorophyll), mathematical vocabulary (denominator, equation, perimeter), and metacognitive academic directives (analyze, compare, predict, evaluate). Cebuano tokens, by contrast, concentrated in Affective-Relational (52.1%) and Classroom Management (30.6%) domains, deploying encouragement expressions (Maayo kaayo! — Very good!), expressions of urgency (Dali na — Hurry up), and solidarity markers (Okay ba tanan? — Is everyone okay?). This complementarity mirrors the findings of Zentella (1997) in the Puerto Rican context and Martin-Jones (1995) in South Asian bilingual classrooms.

Ang pagsusuri ng semantic domain ay nagpapakita ng malinaw at consistent na functional specialization sa pagitan ng Ingles at Cebuano sa lahat ng apat na classroom-specific na domain. Ang mga token sa Ingles ay nakakonsentra sa Academic-Technical na domain sa 78.3%, naaayon sa domain restriction hypothesis ni Fishman (1972) at sa pagkakaiba ni Cummins (2000) sa pagitan ng mga function ng BICS at CALP. Ang mga token sa Cebuano naman ay nakakonsentra sa Affective-Relational (52.1%) at Classroom Management (30.6%) na mga domain.

Table 2. Semantic Domain Distribution by Language

Talahanayan 2. Distribusyon ng Semantic Domain ayon sa Wika

Semantic Domain / Domain ng Kahulugan	English (%)	Cebuano (%)	Sample Tokens
Academic / Technical	78.3	9.2	photosynthesis, fraction, analyze, predict
Affective / Relational	8.4	52.1	Maayo kaayo! Ayaw kulba. Dili ko masuko.
Classroom Management	6.9	30.6	Hunong na. Open your books. Paminaw.
Discourse Markers	6.4	8.1	okay, so, right?, uy, ha, man gyud

Syntactic Patterns in Code-Switching



Three recurrent syntactic configurations account for 89.2% of intra-sentential switching instances. Pattern A — English Noun within Cebuano Determiner Phrase — involves the insertion of English nominal heads within Cebuano DP frames, as in *ang cells sa body*, *ang fraction nga gamay*, and *ang equation nga complex*. In all instances, the Cebuano topic marker *ang* precedes the English noun, and the Cebuano linker *nga* introduces post-nominal modification, confirming Cebuano as the matrix language in the sense of Myers-Scotton (1993). Pattern B — Cebuano Modal plus English Verb Root with Cebuano Affix — involves English verbal roots embedded within Cebuano modal constructions with Cebuano voice morphology attached: *kinahanglan i-analyze*, *pwede mag-compare*, *angay i-submit*. This pattern constitutes classic nonce borrowing as theorized by Poplack (1980) and parallels the Ilocano patterns documented by Rubino (2005). Pattern C — Reported Speech in English within Cebuano Complement Clause — involves full English clauses embedded as complements of Cebuano reporting verbs: *Giingnan ko kamo nga open your books*, *Nakaingon ang teacher nga the answer is ten*.

Tatlong recurrent na sintaktikong configuration ang sumasaklaw sa 89.2% ng mga intra-sentential switching instance. Ang Pattern A — Ingles na Pangngalan sa loob ng Cebuano Determiner Phrase — ay kinabibilangan ng pagpasok ng mga Ingles na nominal head sa loob ng mga Cebuano DP frame, tulad ng *ang cells sa body*, *ang fraction nga gamay*, at *ang equation nga complex*. Ang Pattern B — Cebuano Modal kasama ang Ingles na Verb Root na may Cebuano Affix — ay kinabibilangan ng mga Ingles na verbal root na naka-embed sa loob ng mga Cebuano modal construction na may nakakapit na Cebuano voice morphology: *kinahanglan i-analyze*, *pwede mag-compare*. Ang Pattern C — Direktang Pananalita sa Ingles sa loob ng Cebuano Complement Clause — ay kinabibilangan ng mga buong Ingles na sugnay na naka-embed bilang mga complement ng mga Cebuano na pandiwa ng pag-uulat.

Discussion

The findings collectively confirm the Matrix Language Frame model's applicability to Cebuano-English classroom code-switching: across all three dominant syntactic patterns, Cebuano provides the morphosyntactic skeleton — topic markers, ligature particles, aspect affixes, and modal verbs — while English supplies lexical content in academically specialized positions. This has direct implications for NLP pipeline design. A tokenizer that treats Cebuano-English text as English — as most publicly available tokenizers would — will fail to parse the morphological structure of forms like *nag-analyze*, *i-compute*, or *mag-research*, because the affixes belong to a different morphological system than the root. A Cebuano-first tokenization strategy, followed by an embedded English lexical lookup, would more accurately reflect the grammatical organization of this data.

Pinagsama-sama, kumpirma ng mga natuklasan ang pagiging naaangkop ng Matrix Language Frame model sa Cebuano-Ingles na pagsasalit sa silid-aralan: sa lahat ng tatlong dominanteng sintaktikong pattern, ang Cebuano ay nagbibigay ng morphosyntactic skeleton — mga topic marker, ligature particle, aspect affix, at modal verb — habang ang Ingles ay nagbibigay ng lexical content sa mga akademikong espesyalisadong posisyon. Mayroon itong direktang implikasyon para sa disenyo ng NLP pipeline.

From a pedagogical perspective, the semantic domain complementarity documented here suggests that teachers have developed an implicit sociolinguistic competence in domain assignment that aligns with Fishman's (1972) theoretical predictions and Cummins' (2000) BICS-CALP distinction. While this implicit competence is pedagogically productive — Cebuano serves affective scaffolding while English anchors academic vocabulary — it also risks reinforcing students' association of Cebuano with informal, non-academic registers, potentially undermining the long-term goals of MTB-MLE to establish Cebuano as a fully capable academic language. Tupas (2015) raised precisely this concern, warning that unreflective code-switching in MTB-MLE classrooms could inadvertently perpetuate the symbolic subordination of regional languages that the policy sought to reverse.

Mula sa pedagogikal na pananaw, ang semantic domain complementarity na naitala dito ay nagmumungkahi na ang mga guro ay nagbuo ng isang implicit sociolinguistic na kakayahan sa domain assignment na naaayon sa mga teoretikal na hula ni Fishman (1972) at ng BICS-CALP distinction ni Cummins (2000). Habang ang implicit na kakayahang ito ay pedagogikally productive, mayroon din itong panganib na palakasin ang asosasyon ng mga mag-



aaral ng Cebuano sa mga impormal, hindi akademikong rehistro, na potensyal na napipinsala ang mga pangmatagalang layunin ng MTB-MLE.

Conclusion and Recommendations

This study has demonstrated that Cebuano-English code-switching in elementary classroom instruction in Carmen District, Cebu, is a systematic, structured, and functionally differentiated phenomenon consistent with the predictions of the Matrix Language Frame model and the domain restriction hypothesis. Cebuano consistently functions as the matrix language providing grammatical architecture, while English is embedded primarily as technical and academic lexical content. Three dominant syntactic integration patterns account for nearly 90% of intra-sentential switches, and semantic analysis reveals complementary functional specialization between the two languages across four classroom discourse domains.

Ipinakita ng pag-aaral na ito na ang Cebuano-Ingles na pagsasalit sa elementaryang pagtuturo sa silid-aralan sa Carmen District, Cebu ay isang sistematiko, istrukturado, at functionally differentiated na penomeno na naaayon sa mga hula ng Matrix Language Frame model at ng domain restriction hypothesis. Ang Cebuano ay patuloy na gumaganap bilang matrix language na nagbibigay ng gramatikong arkitektura, habang ang Ingles ay naka-embed pangunahin bilang teknikal at akademikong lexical content.

Based on these findings, four recommendations are advanced. First, NLP developers targeting Philippine classroom language processing should adopt Cebuano-centric morphological tokenization pipelines with embedded English lexical layers rather than English-first or Filipino-first approaches. Second, DepEd curriculum designers should treat the documented code-switching patterns as empirical evidence for developing explicit bridging activities that leverage the observed domain specialization, helping students consciously transfer academic vocabulary between Cebuano and English rather than treating them as unrelated systems. Third, teacher education programs in the Visayas should incorporate applied sociolinguistics content that equips teachers to make their code-switching choices pedagogically deliberate and metacognitively transparent to students. Fourth, future research should extend this corpus longitudinally across academic years to track whether MTB-MLE implementation progressively shifts the balance of classroom language use toward Cebuano as intended by policy.

Batay sa mga natuklasan na ito, apat na rekomendasyon ang isinulong. Una, ang mga developer ng NLP na nagtarget ng Philippine classroom language processing ay dapat mag-adopt ng Cebuano-centric morphological tokenization pipeline na may mga embedded na Ingles na lexical layer. Pangalawa, ang mga designer ng kurikulum ng DepEd ay dapat tratuhin ang mga naitala na pattern ng pagsasalit bilang empiryal na katibayan para sa pagbuo ng mga tahasang bridging activity. Pangatlo, ang mga programa sa edukasyon ng guro sa Visayas ay dapat mag-incorporate ng applied sociolinguistics content. Pang-apat, ang susunod na pananaliksik ay dapat palawakin ang corpus na ito nang longitudinal sa iba't ibang taon ng pag-aaral.

References

- Aguilar, G., Kar, S., & Solorio, T. (2020). Named entity recognition for code-switched data: Building resources and training models from scratch. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing* (pp. 5419–5434). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.emnlp-main.437>
- Auer, P. (Ed.). (1998). *Code-switching in conversation: Language, interaction and identity*. Routledge.
- Barman, U., Das, A., Wagner, J., & Foster, J. (2014). Code mixing: A challenge for language identification in the language of social media. In *Proceedings of the First Workshop on Computational Approaches to Code Switching* (pp. 13–23). Association for Computational Linguistics. <https://doi.org/10.3115/v1/W14-3902>
- Bautista, M. L. S. (2004). Tagalog-English code switching as a mode of discourse. *Asia Pacific Education Review*, 5(2), 226–233. <https://doi.org/10.1007/BF03026282>



- Bernardo, A. B. I. (2004). McKinley's questionable gift: Over a century of English in Philippine education. In J. W. Tollefson & A. B. M. Tsui (Eds.), *Medium of instruction policies: Which agenda? Whose agenda?* (pp. 155–173). Lawrence Erlbaum Associates.
- Bialystok, E. (2001). *Bilingualism in development: Language, literacy, and cognition*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511605963>
- Cabasag, N., Barrios, A., Galvante, P., Andrada, M., & Morales, J. (2019). Hate speech detection in Filipino code-switched social media text. In *Proceedings of the 3rd International Conference on Natural Language Processing and Information Retrieval* (pp. 76–80). ACM. <https://doi.org/10.1145/3342827.3342846>
- Canagarajah, A. S. (1995). Functions of codeswitching in ESL classrooms: Socialising bilingualism in Jaffna. *Journal of Multilingual and Multicultural Development*, 16(3), 173–195. <https://doi.org/10.1080/01434632.1995.9994599>
- Canagarajah, A. S. (2011). Codemeshing in academic writing: Identifying teachable strategies of translanguaging. *The Modern Language Journal*, 95(3), 401–417. <https://doi.org/10.1111/j.1540-4781.2011.01207.x>
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
- Cruz, J. C. B., & Cheng, C. (2021). Establishing baselines for text classification in low-resource languages. *International Journal on Natural Language Computing*, 10(4), 1–15. <https://doi.org/10.5121/ijnlc.2021.10401>
- Cummins, J. (2000). Language, power and pedagogy: Bilingual children in the crossfire. *Multilingual Matters*. <https://doi.org/10.21832/9781853596773>
- Dekker, D., & Young, C. (2005). Languages of instruction for whom? Language use in multilingual education programs in the Philippines. *Journal of Southeast Asian Education*, 6(1), 1–24.
- Department of Education (DepEd). (2012). DepEd Order No. 16, s. 2012: Policy guidelines on the implementation of grades 1 to 10 of the K to 12 basic education curriculum effective school year 2012–2013. Department of Education, Philippines.
- Dumanig, F. P., David, M. K., & Symaco, L. P. (2012). Language management in multilingual families in the Philippines and Malaysia. *Journal of Multilingual and Multicultural Development*, 33(6), 527–540. <https://doi.org/10.1080/01434632.2012.706301>
- Fishman, J. A. (1972). *The sociology of language: An interdisciplinary social science approach to language in society*. Newbury House.
- Gollan, T. H., & Ferreira, V. S. (2009). Should I stay or should I switch? A cost-benefit analysis of voluntary language switching in young and aging bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(3), 640–665. <https://doi.org/10.1037/a0014981>
- Gonzalez, A. (1998). The language planning situation in the Philippines. *Journal of Multilingual and Multicultural Development*, 19(5–6), 487–525. <https://doi.org/10.1080/01434639808666365>
- Grosjean, F. (2010). *Bilingual: Life and reality*. Harvard University Press.
- Gumperz, J. J. (1982). *Discourse strategies*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511611834>
- Hornberger, N. H. (2003). Continua of biliteracy: An ecological framework for educational policy, research, and practice in multilingual settings. *Multilingual Matters*. <https://doi.org/10.21832/9781853596568>
- Jefferson, G. (2004). Glossary of transcript symbols with an introduction. In G. H. Lerner (Ed.), *Conversation analysis: Studies from the first generation* (pp. 13–31). John Benjamins. <https://doi.org/10.1075/pbns.125.02jef>
- Khanuja, S., Dandapat, S., Srinivasan, A., Sitaram, S., & Choudhury, M. (2020). GLUECoS: An evaluation benchmark for code-switched NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 3575–3585). ACL. <https://doi.org/10.18653/v1/2020.acl-main.329>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Leech, G. N. (1974). *Semantics: The study of meaning*. Penguin Books.
- Li Wei. (2011). Moment analysis and translanguaging space: Discursive construction of identities by multilingual Chinese youth in Britain. *Journal of Pragmatics*, 43(5), 1222–1235. <https://doi.org/10.1016/j.pragma.2010.07.035>



- Lorente, B. P., & Tupas, T. R. F. (2014). Demythologizing English as a language of development. In R. Rubdy & L. Alsagoff (Eds.), *The global-local interface and hybridity: Exploring language and identity* (pp. 69–85). Multilingual Matters.
- MacSwan, J. (2000). The architecture of the bilingual language faculty: Evidence from intrasentential code switching. *Bilingualism: Language and Cognition*, 3(1), 37–54. <https://doi.org/10.1017/S1366728900000134>
- Macnamara, J. (1967). The bilingual's linguistic performance: A psychological overview. *Journal of Social Issues*, 23(2), 58–77. <https://doi.org/10.1111/j.1540-4560.1967.tb00576.x>
- Mager, M., Gutierrez-Vasques, X., Sierra, G., & Meza-Ruiz, I. (2019). Challenges of language technologies for the indigenous languages of the Americas. In *Proceedings of the 27th International Conference on Computational Linguistics* (pp. 55–69). Association for Computational Linguistics.
- Martin-Jones, M. (1995). Code-switching in the classroom: Two decades of research. In L. Milroy & P. Muysken (Eds.), *One speaker, two languages: Cross-disciplinary perspectives on code-switching* (pp. 90–111). Cambridge University Press. <https://doi.org/10.1017/CBO9780511620867.005>
- McEnery, T., & Hardie, A. (2012). *Corpus linguistics: Method, theory and practice*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511981395>
- Myers-Scotton, C. (1993). *Duelling languages: Grammatical structure in codeswitching*. Clarendon Press.
- Myers-Scotton, C., & Jake, J. L. (2001). Explaining aspects of code-switching and their implications. In J. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 84–116). Blackwell.
- Parafita Couto, M. C., Munarriz, A., Epelde, I., Deuchar, M., & Oyharçabal, B. (2015). Gender conflict resolution in Spanish-Basque mixed DPs. *Bilingualism: Language and Cognition*, 18(2), 304–323. <https://doi.org/10.1017/S1366728913000473>
- 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>
- Probyn, M. (2009). 'Smuggling the vernacular into the classroom': Conflicts and recursions in underprepared school language environments. *International Journal of Bilingual Education and Bilingualism*, 12(2), 193–209. <https://doi.org/10.1080/13670050802153137>
- Rubino, C. (2005). Ilocano. In A. Adelaar & N. P. Himmelmann (Eds.), *The Austronesian languages of Asia and Madagascar* (pp. 326–372). Routledge.
- Samson, J. N., Tino, J. J., & Tumanda, A. R. (2018). Toward a Universal Dependencies treebank for Tagalog. In *Proceedings of the 16th International Workshop on Treebanks and Linguistic Theories* (pp. 198–208). Association for Computational Linguistics.
- Skutnabb-Kangas, T. (2000). Linguistic genocide in education — or worldwide diversity and human rights? Lawrence Erlbaum Associates.
- Solorio, T., & Liu, Y. (2008). Learning to predict code-switching points. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing* (pp. 973–981). Association for Computational Linguistics.
- Tupas, R. (2015). Inequalities of multilingualism: Challenges to mother tongue-based multilingual education. *Language and Education*, 29(2), 112–124. <https://doi.org/10.1080/09500782.2014.977295>
- Tupas, R., & Lorente, B. P. (2014). A 'new' politics of language in the Philippines: Bilingual education and the new challenge of the mother tongues. In P. Siemund, I. Gogolin, M. E. Schulz, & J. Davydova (Eds.), *Multilingualism and language diversity in urban areas* (pp. 317–338). John Benjamins. <https://doi.org/10.1075/hsm.7.15tup>
- UNESCO. (2003). *Education in a multilingual world: UNESCO education position paper*. UNESCO Publishing.
- Valdez, P. N. M., Caña, R. J., & Minton, T. C. W. (2016). Results of Mother Tongue-Based Multilingual Education (MTB-MLE) in reading and writing across grade levels. *The Asia-Pacific Education Researcher*, 25(5–6), 789–798. <https://doi.org/10.1007/s40299-016-0309-x>
- Winata, G. I., Madotto, A., Wu, C., Shin, J., Xu, P., Xu, H., & Fung, P. (2019). Learning multilingual meta-embeddings for code-switching and adversarial examples. In *Findings of the Association for Computational Linguistics: EMNLP 2019* (pp. 3switching–3523). ACL. <https://doi.org/10.18653/v1/D19-6004>
- Zentella, A. C. (1997). *Growing up bilingual: Puerto Rican children in New York*. Blackwell.



Note on Filipino Translations / Tala sa mga Pagsasalin sa Filipino

All Filipino translations appearing in this document were produced for the purpose of accessibility within multilingual Philippine academic contexts. The translations represent the authors' own rendering of the source English text and are not drawn from any external translated source. All references cited in this paper are listed in their original English-language publication form in accordance with APA 7th edition citation standards.