

Um to Kuan: Cognitive Load and Verbal Fillers in Multilingual Speech

Melophyl C. Baguio 

Author Information:

Faculty, School of Liberal Arts
Ateneo de Zamboanga University,
Zamboanga City, Philippines

Correspondence:

baquiomelc@adzu.edu.ph

Article History:

Date Received: July 12, 2026

Date Revised: July 24, 2026

Date Accepted: July 24, 2026

Date Published: August 11, 2026

DOI:

<https://doi.org/10.5281/zenodo.21873584>

Document ID: 2026IJROMS0286

Recommended Citation:

Baguio, M. C. (2026). *Um to Kuan: Cognitive Load and Verbal Fillers in Multilingual Speech*. *International Journal of Research on Multidisciplinary Studies*, 1(7), 14–24.

<https://doi.org/10.5281/zenodo.21873584>

Abstract. Contemporary psycholinguistic paradigms recognize verbal fillers as functional byproducts of cognitive processing and speech planning rather than indicators of poor articulation. This study examines the tension between processing utility and communicative competence by analyzing speech disfluencies and cross-linguistic transfer in second-language academic contexts. Utilizing a descriptive mixed-method design with a dual-stage sampling process, seventy (70) first-year undergraduate students at Ateneo de Zamboanga University were selected from five sections of a Purposive Communication course during classroom oral recitations in the 2025–2026 academic year. Spontaneous classroom speech was recorded and transcribed, while Focus Group Discussions provided qualitative insights into subjective cognitive triggers. Quantitative findings revealed that unlexicalized, non-word verbal fillers predominated in student discourse, with *um* (frequency of 69) and *uh* (frequency of 68) emerging as the most frequent disfluencies, whereas lexicalized phrase fillers occurred minimally (frequency of 3). A distinct pattern of cross-linguistic transference under stress was observed through the systematic insertion of localized Filipino and Zamboangueno placeholder tokens, including *Ano o* (frequency of 9), *Ano* (frequency of 4), and *Kuan o* (frequency of 2), into English speech. Thematic analysis of qualitative data revealed that nervousness and the explicit need for processing time were universally identified by 100 percent of participants as the primary catalysts for these disfluencies. These findings demonstrate that verbal fillers and localized placeholder tokens do not reflect linguistic incompetence or communicative deficits. Instead, they function as essential, adaptive cognitive strategies for floor-holding, lexical retrieval management, and coping with high affective filters during evaluative speech tasks.

Keywords: Affective filter; Cognitive load; Cross-linguistic transference; Speech disfluency; Verbal filler

Introduction

In daily human communication, spontaneous speech is rarely characterized by flawless fluency. Instead, casual conversation is systematically punctuated by hesitations, pauses, self-corrections, and non-lexical vocalizations that operate beneath speakers' conscious awareness. Within contemporary psycholinguistics, these structural interruptions are formally recognized as speech disfluencies or verbal fillers. Early foundational work by researchers such as Laserna, Pennebaker, and Seih (2014) categorized these phenomena into two primary manifestations: filled pauses (e.g., *uh* and *um*) and discourse markers (e.g., *I mean*, *you know*, and *like*). While traditional prescriptive linguistic frameworks often dismiss these insertions as signs of poor articulation or communicative deficit, modern psycholinguistic paradigms treat them as crucial, functional byproducts of cognitive processing and speech production.

Fundamentally, verbal fillers serve as essential floor-holding and cognitive stalling mechanisms in spoken discourse. When a speaker encounters a lexical retrieval delay or experiences a high cognitive load while planning an utterance, verbal fillers are deployed to preserve their turn in conversation. According to Van Baalen (2001), these sounds, words, or phrases can appear anywhere in an utterance and be deleted without changing the semantic content. By filling a prospective silence, the speaker signals to the listener that the communicative act is



ongoing, thereby avoiding an abrupt, awkward, or ambiguous silent pause. Because these discourse markers carry pragmatic and structural utility rather than semantic propositions, their production requires minimal cognitive overhead, leaving the brain temporarily free to process low-frequency vocabulary or complex ideas. Consequently, they give the speaker a brief processing window to retrieve target words from their mental dictionary or restructure grammatical frames without changing the core propositional content of the message.

The contextual behavior of these items dictates their classification. For instance, high-frequency lexical units such as *like* or *well* can operate concurrently as core syntactic components (e.g., prepositions or adjectives) or as structurally dispensable verbal fillers depending entirely on the immediate discourse environment. Empirical studies demonstrate that nearly all speakers, regardless of demographic variables or language proficiency, rely on these interpolations when responding under time constraints, in rushed situations, or to unscripted prompts. However, as Yule (2006) defines them simply as breaks in the flow of speech, their frequency and typology are not uniform; they vary heavily across situational contexts and are deeply influenced by speaker anxiety, perceived task difficulty, and underlying divided attention. Scholars have narrowed down the immediate production of filler words into three primary categories: divided attention, infrequent words, and situational nervousness (Dockrell et al., 1998; Oomen & Postma, 2001).

While verbal fillers are natural elements of spontaneous cognitive processing, their overuse frequently invites social and professional penalties. In instructional and evaluative spaces, an excessive density of disfluencies can drastically undermine a speaker's perceived credibility, making them appear unprepared, inexperienced, or lacking domain expertise (Conrad et al., 2013). Conversely, moderate usage has been shown to assist listener comprehension, with Frederick Conrad et al. (2013) finding that those who utilize filler words moderately are perceived as having the same level of eloquence as those who do not. In academic environments, this tension between processing utility and communicative competence becomes particularly pronounced, particularly during formal evaluative tasks, where nervousness and glossophobia, a public-speaking anxiety, can induce dysfunctional voice quivering, vocalized pauses, and frequent repetitions (García-López et al., 2013).

This study investigates this psycholinguistic tension by examining the verbal filler profiles of first-year college students at Ateneo de Zamboanga University. Specifically, the research analyzes the structural types, frequencies, and perceived cognitive catalysts of verbal disfluencies produced during English oral recitations within the first semester of the 2025–2026 academic year. In addition to investigating standard English discourse markers, this paper explores the cross-linguistic intersection in which L2 (English) speech anxiety triggers reliance on localized L1 and regional discourse markers. By analyzing these behaviors within a structured classroom environment, this study aims to contribute to the broader psycholinguistic literature surrounding second-language production, cognitive load management, and the communicative realities of multilingual academic spaces.

Research Questions

This study specifically aims to answer the following research questions:

1. What are the structural typologies and frequencies of English verbal fillers used by first-year college students during formal oral recitations?
2. What localized, non-English discourse tokens are deployed as cross-linguistic verbal fillers?
3. What are the common cognitive, psychological, and affective reasons that drive first-year college students to utilize these verbal fillers during academic speech tasks?

Scope and Delimitation

This study examines the psycholinguistic profile of verbal fillers produced by first-year undergraduate students during structured second-language (L2) academic speech tasks. Specifically, it investigates the structural typologies and frequencies of English verbal fillers used during formal oral recitations, the use of localized non-English discourse tokens as cross-linguistic verbal fillers, and the cognitive, psychological, and affective factors associated with their use during academic speech tasks. To maintain a focused and feasible analytical framework, the study is delimited to first-year undergraduate students at Ateneo de Zamboanga University during the first semester of the 2025–2026 academic year. Of the thirteen available course sections, data collection was limited to five sections of Purposive Communication (PURCOM) handled directly by the researcher, with convenience sampling employed due to institutional access constraints. The participants comprised 70 students, including 40 females and 30 males. The quantitative analysis focused primarily on vocalized unlexicalized fillers, such as *um* and *uh*, and lexicalized phrase fillers, such as *I mean* and *I guess*. Acoustic repetitions and silent pauses were also



documented for contextual purposes; however, detailed phonological analysis of these features and nonverbal communication channels, including kinesics and facial expressions, were beyond the scope of the study. Data collection was further confined to evaluative, high-stakes classroom oral recitations involving ten standardized academic questions on research methodology. Consequently, the findings describe verbal disfluency patterns within the specific context of evaluative L2 academic speech and should not be generalized to low-stakes classroom interactions, informal conversations, or spontaneous discourse conducted in the learners' first language (L1).

Literature Review

The Cognitive Taxonomy vs. Monolingual Bias in Speech Disfluency Models

While early foundational frameworks by Laserna, Pennebaker, and Seih (2014) established the classical taxonomy dividing speech disfluencies into unlexicalized fillers (*uh*, *um*) and lexicalized phrase markers (*you know*, *I mean*), these paradigms were derived almost exclusively from monolingual English corpora. Consequently, established models assume a binary division between non-propositional acoustic stalling devices and fully articulated, syntactically bound discourse markers.

In instructional and evaluative spaces, an excessive density of disfluencies can drastically undermine a speaker's perceived credibility, making them appear unprepared, inexperienced, or lacking domain expertise (Borkowski, 2023; Conrad et al., 2013). Conversely, moderate usage has been shown to assist listener comprehension, with Conrad et al. (2013) finding that those who utilize filler words moderately are perceived as having the same level of eloquence as those who do not. Furthermore, psycholinguistic assessments of listener perception demonstrate that filled pauses directly influence how audiences evaluate a speaker's certainty and underlying domain authority (Brennan & Williams, 1995; Fox Tree, 2001).

However, this rigid categorization fails when applied to multilingual processing environments. Standard cognitive models, such as Levelt's *Blueprint for the Speaker*, treat disfluencies as internal processing disruptions occurring strictly within a single target language grid. While these taxonomies adequately explain why speakers under cognitive load default to low-overhead fillers like *um* or *uh* to bypass complex phrasal assembly (Goldwater et al., 2010; Oomen & Postma, 2001), they remain completely silent on the structural behavior of cross-linguistic discourse tokens. Modern psycholinguistic frameworks do not account for how a speaker's L1 (native language) or regional repertoire intervenes during an L2 (second language) formulation failure, leaving a major theoretical blind spot regarding how multilinguals manage cognitive load under stress.

Evaluative Penalties vs. Cross-Linguistic Realities in Philippine Academic Discourse

The ongoing debate surrounding the communicative impact of filler usage further highlights the limitations of Western-centric research. While mainstream communication literature warns of severe professional penalties, loss of domain authority, and degraded listener perception caused by disfluent speech (Conrad et al., 2013), these studies assume a sociolinguistic landscape where fluency is judged against monolingual normative standards. Although alternative perspectives suggest that moderate disfluency can enhance perceived spontaneity or signal honesty (Arciuli, Mallard, & Villar, 2014), neither side of this debate addresses how evaluative anxiety (*glossophobia*) operates in second-language academic contexts, such as those found in Philippine higher education. Crucially, existing bilingual disfluency models typically conceptualize cross-linguistic code-switching as a deliberate, pragmatic choice to switch topic, signal group identity, or convey complex semantic nuances. Current frameworks fundamentally fail to account for localized placeholder tokens, such as *Kuan*, *Ano*, *Kuan o*, and *Ano o*, within specialized Philippine academic discourse.

Unlike standard L1 code-switching, which imports semantically rich words from a native vocabulary, regional Philippine placeholder tokens operate in a uniquely hybrid psycholinguistic space:

- **Lexical Nullity with Syntactic Integration:** Tokens like *Kuan* and *Ano* carry no intrinsic semantic proposition, yet they do not behave like standard unlexicalized non-words (*um/uh*).
- **Automated Floor-Holding Devices:** In localized multilingual ecosystems (e.g., Zamboanga City, where Chavacano, Tagalog, English, and Visayan languages intersect), these tokens function as deeply automated, low-overhead structural buffers.

When L2 (English) speech anxiety triggers a high affective filter (García-López et al., 2013) and causes a real-time lexical retrieval breakdown, classical models predict either a silent pause, an acoustic repetition, or an L2 filled pause (*um/uh*). They fail to explain why multilingual speakers routinely bypass the target language's



disfluency system entirely, deploying L1 regional placeholders (*Kuan*, *Ano*) directly into English syntactic frames to hold the conversational floor without surrendering discourse continuity.

Synthesis and Research Gap

While prior literature proves that cognitive load, divided attention, and vocabulary retrieval friction drive verbal disfluencies (Dockrell et al., 1998; Oomen & Postma, 2001), current paradigms treat cross-linguistic transfer and speech disfluency as separate, unrelated phenomena. By failing to examine how regional placeholder tokens serve as adaptive cognitive coping mechanisms during high-stakes L2 evaluation, existing frameworks provide an incomplete picture of speech production in multilingual academic spaces. This study directly addresses this theoretical gap by examining the structural typologies, frequencies, and affective catalysts of both standard English fillers and localized cross-linguistic tokens (*Ano*, *Kuan*) within the Philippine college classroom.

Methodology

Research Design

This study utilizes a descriptive mixed-methods design to capture a comprehensive, multi-dimensional profile of classroom speech disfluencies. While the investigation relies heavily on a qualitative framework to provide "thick descriptions" of spontaneous speech, it also incorporates descriptive quantitative metrics to systematically calculate the structural frequency and distributional variance of specific filler types. Linguistically, this inquiry is situated within applied psycholinguistics, focusing on speech disfluency as an observable window into a speaker's underlying cognitive load and lexical retrieval constraints during high-stakes L2 production. By capturing natural classroom discourse alongside subjective participant reflections, this design avoids artificial laboratory conditions and prioritizes ecological validity.

Sampling Design

To balance logistical feasibility with statistical representation, a dual-stage sampling mechanism was systematically deployed. In the first stage, non-probability convenience sampling was used to select five distinct classroom sections enrolled in courses that required formal, graded oral recitations as part of their academic assessment. In the second stage, a probability-based fishbowl sampling technique was applied across these five sections to randomly select a final analytical sample of 70 students ($n = 70$) for close linguistic tracking and transcription. This dual-stage sampling mechanism yielded a final study sample of seventy ($n = 70$) students, consisting of forty ($n = 40$) female and thirty ($n = 30$) male participants.

Research Locale

The study was conducted at the Ateneo de Zamboanga University, situated in Zamboanga City, a highly multilingual urban ecosystem in the southern Philippines where fluid code-switching and linguistic convergence are defining features of daily communication. The typical undergraduate student in this environment navigates a complex, shifting linguistic repertoire, possessing varying degrees of proficiency in Chavacano (a localized Spanish-based creole), Tagalog/Filipino (the national language), English (the primary medium of academic instruction), and regional languages such as Tausug or Bisaya. Within this rich linguistic matrix, certain discourse tokens that cross language boundaries usually function as universal acoustic placeholders. When these multilingual individuals encounter heightened cognitive load or evaluative anxiety during formal, exclusively English oral recitations, their executive control networks face intense processing friction. As the brain scrambles to access low-frequency academic terms within the second-language (L2) mental dictionary, a localized retrieval breakdown occurs. Rather than surrendering the conversational floor to an awkward silent pause, the speaker's cognitive processing mechanism bypasses the taxing grammatical assembly required by English phrase fillers. Instead, it defaults to these deeply automated, low-overhead native-language (L1) regional tokens, which may naturally manifest as default cognitive stalling devices precisely because they are heavily reinforced by the region's everyday colloquial discourse. They serve as vital, highly accessible psychological buffers, maintaining conversational continuity and holding the floor while the speaker resolves the immediate L2 lexical gap.



Research Participants

The target population for this psycholinguistic inquiry comprised first-year undergraduate students enrolled at Ateneo de Zamboanga University during the 2025–2026 academic year. Linguistically, the sample represented a highly heterogeneous, multilingual cohort reflective of the unique urban ecosystem of Zamboanga City. The participants operated within a fluid, multi-tiered linguistic repertoire, possessing varying degrees of baseline proficiency in Chavacano (a localized Spanish-based creole), Tagalog/Filipino (the national language), and regional Visayan languages (primarily Bisaya/Cebuano and Tausug). Importantly, while all 70 participants used English as their secondary language (L2) and the primary medium of formal academic instruction, their naturalistic everyday discourse was characterized by frequent inter- and intra-sentential code-switching between their native tongues (L1). For the subsequent qualitative phase of the study, a purposive subsample of these participants was invited to participate in Focus Group Discussions (FGDs). The selection criteria for the FGD cohorts required that participants exhibit a high frequency of verbal disfluencies during their recorded oral performances. This ensured that the qualitative data gathered regarding subjective cognitive catalysts, such as evaluative anxiety, psychological filters, and real-time processing bottlenecks, directly mirrored the behavioral manifestations observed in the quantitative speech stream analysis. All 70 participants volunteered for the study, met the inclusion criteria, and completed all phases of the evaluative tasks without attrition.

Research Instrument

The study utilized two primary data elicitation instruments to address its descriptive and analytical objectives: a video and audio recording protocol and a semi-structured focus group discussion (FGD) guide. The video and audio recording protocol involved the use of a high-definition smartphone positioned strategically within the classroom to capture students' spontaneous utterances during the structured academic speech tasks. This instrument documented the natural cadence, pauses, verbal fillers, and relevant phonological variations occurring during the students' oral responses. Following the collection of the speech data, a semi-structured FGD guide was administered across the five participating course sections to elicit students' subjective metacognitive explanations regarding their use of lexicalized, unlexicalized, and localized cross-linguistic filler words. To ensure consistency and standardization across the speech tasks, a validated set of ten (10) academic questions was used, with questions arranged progressively from lower to higher levels of cognitive difficulty and focused on fundamental research design and core communication concepts. Together, these instruments enabled the researchers to document observable speech disfluencies while also exploring the cognitive, psychological, and affective factors underlying students' use of verbal fillers during structured L2 academic speech tasks.

Data Gathering Procedure

Prior to data collection, students were informed of the nature of the psycholinguistic observation, and formal informed consent was secured. The data-gathering process was seamlessly embedded in standard class routines to preserve environmental authenticity. During standard class sessions, the video recording apparatus was activated to cover the front of the classroom. The researcher began the oral recitation by drawing names at random from the fishbowl. Upon selection, the student was presented with one of the 10 progressively difficult academic questions from the validated set. To minimize situational anxiety that could skew normal baseline disfluencies, students were not subjected to arbitrary time constraints and were permitted to articulate their thoughts at an individual, natural pace. If a student could not provide a conceptually complete response, successive names were drawn until the question was satisfactorily resolved and all 10 target items were addressed, capping recording intervals between 30 and 50 minutes per section. Once classroom recordings concluded, the researcher convened the randomly selected participants for a series of Focus Group Discussions. These discussions isolated the students' internal perceptions of speech disfluency, focusing on the psychological, cognitive, and affective factors (e.g., glossophobia, vocabulary gaps) experienced during the preceding recitation task.

Results and Discussions

Research Question 1: What are the structural typologies and frequencies of English verbal fillers used by first-year college students during formal oral recitations?

Table 1. Frequencies of Observed English Verbal Filler Typologies

Structural Category	Specific Verbal Filler	Frequency (f)
Non-Word (Unlexicalized) Fillers	<i>um</i>	69
	<i>uh</i>	68
	<i>so</i>	3
	<i>okay</i>	3
	<i>basically</i>	2
	<i>like</i>	2
Phrase (Lexicalized) Fillers	<i>I mean</i>	3
	<i>I guess</i>	1

As detailed in Table 1, non-word placeholders (*um* and *uh*) dominate the participants' disfluency profiles, whereas phrasal discourse markers occur only marginally, and explicit phrasal fillers (*I mean*, *I guess*) are virtually non-existent. The empirical inventory reveals a pronounced structural skew between non-word unlexicalized fillers and phrase-based lexicalized markers produced during formal second-language (L2) academic speech tasks.

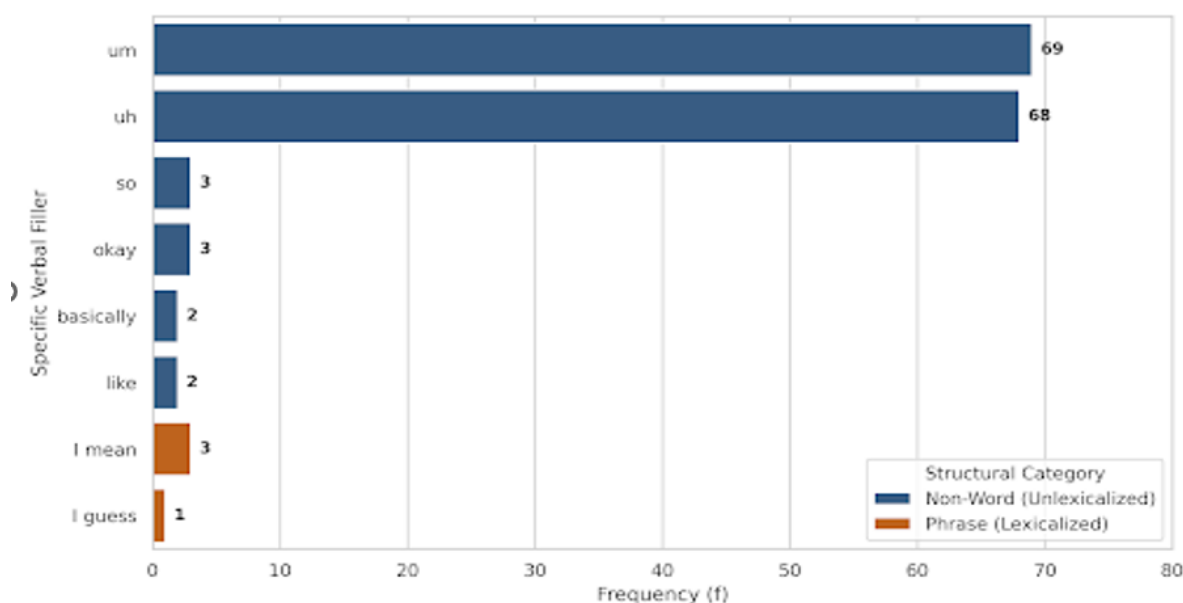


Figure 1. Frequency Distribution of English Filler Typologies (n=70)

From a psycholinguistic standpoint, this distribution provides strong empirical support for Levelt's *Blueprint for the Speaker* and cognitive load paradigms. Under high evaluative anxiety, the massive processing demands of real-time semantic formulation tax working memory capacity. Because lexicalized phrase fillers require full syntactic insertion and phonetic assembly within the Formulator, deploying them imposes an additional cognitive processing tax that an overloaded cognitive architecture actively avoids. Conversely, *um* and *uh* function as low-overhead, non-propositional acoustic placeholders requiring negligible lexical or grammatical encoding. Automatically routing these vocalizations through the Articulator allows speakers to signal a localized delay in the formulation stage without surrendering the conversational floor. In addition to vocalized fillers, mechanical disfluencies, such as structural repetitions and silent pauses, were documented as alternative floor-holding strategies.

Table 2. Frequency of Structural Repetitions and Silent Pauses

Structural Disfluency Type	Frequency (f)
Repetition (Acoustic/Syllabic)	19
Silent Pause	12

As shown in Table 2, mechanical disruptions manifested through both structural repetitions and silent pauses. Within cognitive disfluency literature, these structural interruptions represent distinct processing behaviors:

- Structural Repetitions ($f = 19$):** Syllabic and acoustic repetitions serve as active, vocalized backtracking mechanisms. In line with Oomen and Postma (2001), who demonstrated that task-based activities under divided attention trigger marked increases in repetitions, these occurrences indicate moments where an unlexicalized filler alone failed to buffer cognitive latency. Instead of yielding the turn, the speaker re-executes an already formulated syllable to buy immediate execution time while the Formulator re-plans the subsequent lexical choice.
- Silent Pauses ($f = 12$):** While unlexicalized fillers are deployed to preserve floor-holding control, unvocalized silent gaps signal transient breakdowns in both macro-planning and real-time execution. The presence of silent pauses aligns with the disfluency taxonomy established by Goldwater et al. (2010), which demonstrates that when lexical retrieval delays exceed a critical processing window, cognitive stalling mechanisms transition from vocalized buffers (*um/uh*) to temporary silent blockages.

The core finding of the paper is that speakers default to low-overhead fillers (*um, uh, Ano, Kuan*) rather than phrase fillers (*I mean, I guess*) to minimize cognitive execution cost.

Let the cognitive cost of producing a structural disfluency type i be K_i :

$$(1) K_i = \text{Cost}_{\text{syntactic}}(i) + \text{Cost}_{\text{phonetic}}(i)$$

The probability $P(D_i)$ of selecting a specific filler type D_i when cognitive load is near capacity ($\mathcal{L} \approx C_{\text{total}}$) is inversely proportional to its processing overhead:

$$(2) P(D_i) \propto \exp\left(-\frac{K_i}{\tau}\right)$$

Where τ represents the speaker's real-time cognitive stress factor.

Rather than functioning as isolated errors, these structural repetitions and silent gaps demonstrate that when non-word vocalizations fail to resolve an immediate processing delay, speakers shift across a continuum of mechanical disfluencies to manage ongoing speech architecture under stress.

Research Question 2: What localized, non-English discourse tokens are deployed as cross-linguistic verbal fillers?

Table 3. Distribution of Localized Filipino/Zamboangueño Verbal Fillers

Localized Discourse Token	Frequency (f)
<i>Ano o</i>	9
<i>Ano</i>	4
<i>Kuan o</i>	2



As cataloged in Table 3, participants routinely injected localized placeholder tokens into their English discourse. Structurally, these tokens mirrored the behavior of *um* and *uh*: they were devoid of propositional semantic content, added no informational value, and were deployed strictly to stall for processing time without introducing awkward silence.

This behavior aligns directly with the Language Transference Paradigm under Cognitive Load. When multilingual speakers undergo evaluative L2 speech tasks, high affective filters trigger lexical retrieval friction within the L2 mental dictionary. When executive controls fail to rapidly locate an L2 target term, the cognitive processing system defaults to deeply automated native-language (L1) regional tokens. In this context, *Kuan* and *Ano* function as highly accessible, syntactic placeholders—signaling an intact conceptual goal alongside a temporary breakdown in L2 lexical translation.

Research Question 3: What are the common cognitive, psychological, and affective reasons that drive first-year college students to utilize these verbal fillers during academic speech tasks?

Table 4. Synthesized Perceptions of the Causes and Uses of Verbal Fillers

Perceived Cognitive/Affective Trigger	Frequency (n=70)	Percentage (%)
Nervousness	70	100%
Time to think and process ideas	70	100%
Anxiety	60	86%
Hesitations	60	86%
Occupied/Overloaded thoughts	50	71%
Doubtfulness/Uncertainty	50	71%
Structural pausing	40	57%
Self-correction/Repair mechanisms	40	57%
Lexical gaps/Loss for words	40	57%
Deciding on thematic options	40	57%

As synthesized in Table 4, affective distress and processing constraints dominate participant explanations. The universal identification of evaluation anxiety and processing latency confirms Krashen's Affective Filter Hypothesis and García-López et al.'s (2013) findings on glossophobia. Evaluative anxiety creates a temporary cognitive barrier that restricts rapid access to L2 mental stores, forcing the speech mechanism to insert acoustic buffers (*um*, *uh*, *Ano*, *Kuan*) to maintain turn control while navigating lexical retrieval friction. Consequently, verbal disfluencies in this environment reflect active, adaptive cognitive strategies rather than linguistic deficits.

The paper notes that 100% of participants experienced nervousness and an explicit need for processing time, which acts as a barrier to rapid L2 mental dictionary access. We can define effective L2 retrieval rate (R_{L2}) as a function of L2 proficiency ($\mathcal{P}$) gated by the Affective Filter ($\mathcal{A} \in [0,1]$):

$$(3) R_{L2} = \mathcal{P} \cdot (1 - \mathcal{A})$$

Where:

$\mathcal{A} \rightarrow 1$ under high evaluative anxiety (evaluative stress), causing $R_{L2} \rightarrow 0$ (a complete L2 lexical retrieval breakdown).

Ethical Considerations

In accordance with institutional guidelines and international research ethics standards for human subjects, strict ethical safeguards were implemented throughout this study. Prior to the commencement of data collection, the study protocol, informed consent procedures, and data management plans were formally reviewed and granted full ethical approval by the Ateneo de Zamboanga University Research Ethics Committee (AdZU-REC). Given that data collection involved video recording classroom performances and exploring subjective anxieties during Focus Group Discussions (FGDs), the protocol prioritized participant autonomy, psychological safety, and absolute data confidentiality. Because data collection coincided with formal oral recitations, environments inherently linked to heightened evaluative anxiety, special care was taken to minimize psychological distress. To eliminate the compounding effect of an invasive research presence, high-definition video recording



devices were positioned discretely within the classroom to preserve a naturalistic setting. During the qualitative FGD phases, where students discussed their experiences with nervousness and linguistic performance, the researchers maintained a non-judgmental and supportive environment. Students were reassured that speech disfluencies, such as um, uh, or localized placeholders such as Kuan and Ano, represent a universal and normative cognitive process rather than a measure of intellectual or linguistic incompetence. To safeguard the identities of the seventy ($n = 70$) participants, strict data de-identification protocols were enforced immediately following data collection. All raw orthographic transcripts and qualitative themes were stripped of personally identifiable information (PII), and participants were assigned alphanumeric codes (e.g., Participant 01 and Student A), which were used exclusively in the final manuscript and corresponding data tables. All raw video files, audio recordings, and verbatim transcripts were stored in secure, encrypted cloud repositories and password-protected local drives accessible solely to the primary research team. In compliance with data privacy regulations, all primary audiovisual assets will be permanently deleted and purged from all storage media five years after the publication of the study, ensuring that no recognizable records remain. All potential participants were provided with a comprehensive briefing regarding the study's objectives, the nature of the data to be collected, including video and audio recording, and the intended dissemination of the findings. Participants provided voluntary written informed consent prior to any data collection and were explicitly informed that their participation or refusal to participate would have no bearing on their academic standing, grades, or relationship with the university. Furthermore, participants were explicitly granted the right to withdraw from the study at any point without penalty. These ethical safeguards were implemented to ensure that the study upheld the principles of informed consent, participant autonomy, privacy, confidentiality, psychological safety, and responsible management of research data throughout the research process.

Conclusions

Based on the empirical evidence, thematic analysis, and psycholinguistic synthesis, this study establishes three primary conclusions regarding the nature of disfluency in multilingual academic environments. First, under heightened cognitive load and evaluative anxiety, human working memory systematically prioritizes low-overhead unlexicalized non-word fillers, particularly um and uh, over syntactically complex phrase markers such as I mean and I guess. Because lexicalized phrase fillers require full syntactic and phonetic assembly within the Formulator, an already strained cognitive architecture actively avoids them to conserve processing capacity for real-time speech macro-planning. Second, cross-linguistic transference functions as a highly automated and adaptive strategy for multilingual speakers. When a temporary vocabulary breakdown occurs within the L2 (English) mental dictionary during high-stakes tasks, the executive control network defaults to deeply familiar L1 and regional placeholder tokens, such as Ano and Kuan, enabling speakers to maintain the conversational floor and preserve discourse continuity without imposing substantial grammatical processing demands. Third, the findings support a shift from a deficit-based to a strategy-oriented understanding of verbal disfluencies, indicating that verbal fillers and localized placeholder tokens do not necessarily represent linguistic incompetence, poor articulation, or systemic communicative deficits but may instead function as cognitive tools that enable second-language learners to navigate intense affective filters, manage cognitive capacity thresholds, and negotiate real-time lexical retrieval difficulties in demanding academic contexts. Beyond the localized context of the study, these findings have important implications for global multilingual education, particularly in higher education institutions across non-native English-speaking regions that may operate within prescriptive or monolingual-biased evaluation frameworks that penalize disfluency and enforce rigid language standards. By demonstrating that code-switched placeholder tokens and filled pauses may function as cognitive buffering strategies rather than straightforward indicators of poor proficiency, the study challenges deficit-based language ideologies and provides a basis for educational policymakers and instructional designers to consider more linguistically inclusive assessment methods that recognize spontaneous speech disfluencies as potentially natural and adaptive mechanisms of multilingual cognitive processing rather than automatically treating them as targets for punitive remediation. To expand upon these empirical insights and address the boundaries of the present inquiry, future psycholinguistic research should pursue several complementary directions, including multimodal disfluency tracking that integrates vocalized hesitations with nonverbal communication channels such as kinesics and micro-facial expressions; proficiency-tier comparative analysis examining the structural frequency and typology of localized placeholder tokens such as Kuan and Ano across varying levels of multilingual proficiency; and acoustic precision measurement using specialized phonetic analysis software such as Praat to determine the precise duration, in



milliseconds, of silent pauses and filled hesitations. These directions may provide high-precision quantitative evidence for cross-linguistic comparisons of disfluency profiles and contribute to a more comprehensive understanding of the cognitive, linguistic, and affective mechanisms underlying speech production among multilingual learners.

Recommendations

Based on the empirical findings, quantitative distributions, and qualitative insights of this study, several actionable recommendations are offered for language educators, instructional designers, multilingual undergraduate students, and future psycholinguistic researchers. For language educators and instructional designers, faculty should shift toward supportive and non-punitive evaluation frameworks by reducing rigid, deficit-based language grading criteria that penalize or seek the absolute eradication of verbal fillers, as excessive monitoring of filled pauses may heighten students' evaluation anxiety, exacerbate cognitive overload, and potentially induce the very disfluencies instructors aim to reduce. The implementation of low-stakes and ungraded oral recitations is therefore recommended to lower the affective filter and minimize performance-driven glossophobia. In addition, Purposive Communication (PURCOM) curricula should incorporate structured "pause training" modules to help students replace vocalized disfluencies with professional floor-holding strategies. This training may follow a three-step pedagogical routine: first, an awareness stage in which short recorded student speech samples are used to help learners identify individual trigger points for filled pauses, such as *um* and *uh*, and localized placeholders, such as *Ano* and *Kuan*; second, a two-second silent buffer in which students participate in structured prompt drills and are explicitly instructed to pause silently for approximately two seconds before responding to complex, high-cognitive-demand questions, thereby allowing additional time for L2 macro-planning; and third, floor-holding gestures, in which students practice pairing brief silent pauses with intentional open-hand gestures or micro-pauses that signal continued control of the conversational turn and reduce the perceived need to use *um* or *Ano* as acoustic space-fillers. For multilingual undergraduate students, active management of L2 lexical access gaps is recommended to reduce over-reliance on automated L1 or regional placeholders, such as *Ano* and *Kuan*, during L2 retrieval difficulties under stress. Students may memorize and practice low-overhead L2 stalling or bridging phrases, such as "To phrase this accurately..." or "Looking at the core concept...", which can serve a similar floor-holding function while maintaining L2 speech continuity. They should also develop strategic synonymous substitution through circumlocution, using broader and more frequently encountered English vocabulary to describe a concept when a specific low-frequency academic term is temporarily inaccessible, thereby reducing the likelihood of executive control breakdown. Students should further be encouraged to dismantle the "illusion of transparency" by recognizing that audiences rarely perceive internal evaluative nervousness as intensely as the speaker experiences it; understanding that momentary hesitations can represent normal and universal cognitive buffering mechanisms rather than signs of failure may help reduce affective distress during high-stakes recitations. Moreover, collaborative domain-specific verbalization should be encouraged through regular peer-led speaking activities using specialized course terminology to automate vocabulary retrieval pathways and reduce real-time cognitive friction during formal classroom tasks. For future psycholinguistic researchers, subsequent investigations should build upon the empirical foundations of this study without duplicating its objectives by pursuing three distinct directions: multimodal disfluency modeling, which expands analysis to nonverbal communication channels, particularly kinesics such as hand gestures and micro-facial expressions, to examine how physical body language interacts with vocalized fillers in managing cognitive load during L2 production; proficiency-tiered transference analysis, which examines the structural frequency and selection of localized code-switched placeholders, such as *Kuan* and *Ano*, across distinct levels of multilingual proficiency, including novice, intermediate, and advanced L2 speakers, to model how language mastery influences reliance on L1 disfluency buffers under stress; and high-precision acoustic phonetics, which employs specialized acoustic measurement software, such as Praat, to calculate the precise millisecond duration of silent pauses and vocalized hesitations, thereby providing fine-grained quantitative metrics for cross-linguistic disfluency profiles in bilingual speech.



References

- Almansa-Moreno, J. M., Díez-Bedmar, M. B., & García-López, L. (2013). From being a trainee to being a trainer: Helping peers improve their public speaking skills. *Revista de Psicodidáctica / Journal of Psychodidactics*, 18(2), 331–342. <https://doi.org/10.1387/RevPsicodidact.6419>
- Arciuli, J., Mallard, D., & Villar, G. (2010). "Um, I can tell you're lying": Linguistic markers of deception versus truth-telling in speech. *Applied Psycholinguistics*, 31(3), 397–411. <https://doi.org/10.1017/S0142716410000044>
- Borkowski, A. (2023). Evaluating the impact of disfluencies on the perception of speaker competence using neural speech synthesis. In *Proceedings of the International Congress of Phonetic Sciences (ICPhS)* (pp. 1789–1793). International Congress of Phonetic Sciences.
- Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to speakers' knowledge. *Journal of Memory and Language*, 34(3), 383–398. <https://doi.org/10.1006/jmla.1995.1017>
- Conrad, F. G., Jans, M., Orlowski, R. A., Levenstein, R., Nielsen, D., & Schober, M. F. (2013). Interviewer speech and the success of survey invitations. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 176(1), 191–210. <https://doi.org/10.1111/j.1467-985X.2012.01056.x>
- Dockrell, J. E., Messer, D., George, R., & Wilson, G. (1998). Children with word-finding difficulties—Prevalence, presentation, and naming problems. *International Journal of Language & Communication Disorders*, 33(4), 445–454. <https://doi.org/10.1080/136828298247721>
- Fox Tree, J. E. (2001). Listeners' uses of um and uh in speech comprehension. *Memory & Cognition*, 29(2), 320–326. <https://doi.org/10.3758/BF03195357>
- Goldwater, S., Jurafsky, D., & Manning, C. D. (2010). Which words are hard to recognize? Prosodic, lexical, and disfluency factors that increase speech recognition error rates. *Speech Communication*, 52(3), 181–200. <https://doi.org/10.1016/j.specom.2009.10.001>
- Laserna, C. M., Pennebaker, J. W., & Seih, Y. T. (2014). Um... who likes to say you know: Filler words are used as a function of age, gender, and personality. *Journal of Language and Social Psychology*, 33(3), 328–338. <https://doi.org/10.1177/0261927X14526993>
- Oomen, C. C. E., & Postma, A. (2001). Effects of divided attention on the production of filled pauses and repetitions. *Journal of Speech, Language, and Hearing Research*, 44(5), 997–1004. [https://doi.org/10.1044/1092-4388\(2001/078\)](https://doi.org/10.1044/1092-4388(2001/078))
- Van Baalen, M. (2001). *Fillers in spoken interaction* [Doctoral dissertation, Utrecht University]. Utrecht University Repository.
- Yule, G. (2006). *The study of language* (3rd ed.). Cambridge University Press.