ANONYMOUS AUTHOR(S)

Predictive processing is cognitively challenging in rapid turn-taking, which can induce miscomprehension and misalignment, especially for those with more limited domain-specific knowledge, language proficiency, and cultural experience. We designed a proactive system, XPLAIN, that offloads human prediction to assistive tools, thereby mitigating communication gaps in turn-taking. In a lab-based Wizard-of-Oz study, we evaluated its features (i.e., lexical clarifications, idea and content suggestions, and topic summaries) in scaffolding dyadic virtual meetings on attested theoretical grounds. As expected, XPLAIN improved efficiency and participation in virtual meetings, and promoted inclusivity and active participation, subject to individual differences in language proficiency, personality, and prior experience with AI. This study provides insights for design architects on how to boost smoothness and efficiency in real-time cognitive processing with psychological support for confidence, assurance, engagement, and autonomy. It also advocates for customizing proactive systems aside from the 'one-size-fits-all' approach to internalize user characteristics and background.

CCS Concepts: • **Human-centered computing** → **Laboratory experiments**; **Usability testing**.

Additional Key Words and Phrases: Bilingualism; Proactive agent; AI-mediated communication; Conversation scaffold; Knowledge gap

ACM Reference Format:

Anonymous Author(s). 2026. . 1, 1 (May 2026), 13 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Real-time turn-taking imposes simultaneous cognitive demands on second-language (L2) speakers: comprehend incoming speech, predict appropriate responses, access target-language lexical forms, and articulate within conversational turn-taking constraints. AI-mediated turn-taking (AI-MC) further exacerbates this time-pressing situation by the additional steps of interpreting, evaluating, and integrating AI-suggested content to speech output. For L2 speakers in high-stakes settings, these demands become particularly salient when unfamiliar domain-specific terminology creates additional lexical and conceptual processing deficits. While studies on L2 has extensively documented processing advantages of first language (L1) over L2 in isolation through faster comprehension and conceptual access, whether and when L1 and/or L2 would benefit different phases of processing real-time communication remains under-specified.

Classical theories of bilingual language processing provide theoretical grounding but not empirical guidance for interface design to effectively scaffold L2 in AI-MC. Prediction theory posits the framework that language comprehension operates through rapid generation of predictions about upcoming input. It predicts that L1 should provide processing efficiency advantages during comprehension by enabling stronger predictions through direct conceptual priming. However, prediction theory was developed in face-to-face (F2F) contexts and has not been operationalized for AI-MC where interpretation and evaluation delayed speech planning and production. Similarly, cognitive load theory explains how extraneous load (imposed by design) can interfere with learning and task performance, yet it provides limited guidance on how context-dependent demands create context-dependent load sources. Language-dependent memory theory, which holds that memory retrieval efficiency increases when retrieval-cue language matches original encoding language, predicts that the knowledge learners acquired in L1 would show a greater retrieval benefit through L1 cues,

Manuscript submitted to ACM

yet the magnitude and persistence of such encoding-specificity effects across individual differences in proficiency and time spent in L2-speaking environments remains unknown.

This study investigates how these foundational theories map onto the practical design problem of optimizing bilingual clarification formats in AI-mediated turn-taking. Rather than seeking a universal "best format," we hypothesize that optimal format emerges from the interaction between processing context (comprehension vs. production), content characteristics (domain-specificity and learner familiarity), and individual differences (L2 proficiency and language of prior acquisition).

We operationalize these factors through three research questions grounded in psycholinguistic theory: (1) How does bilingual support affect comprehension and participation when content characteristics systematically vary across domain-specific versus domain-general terminology, and when learner familiarity with that content varies based on prior educational exposure? (2) Does processing demand context modulate which language format optimizes cognitive processing, such that comprehension-phase support should prioritize prediction efficiency (L1 advantage) while production-phase support should prioritize lexical-form accessibility? (3) How do individual differences in L2 proficiency and domain expertise systematically moderate the effectiveness of bilingual support across both content types and processing contexts?

This work contributes to the current research by (1) applying psycholinguistic theories to understand practical AI-MC, 2) showing the value of differentiating phase-specific gaps with distinct theoretical grounds in real time turn-taking; and 3) offering empirically grounded design principles for developing inclusive tools that optimize comprehension, production efficiency, and learning in real-time AI-MC.

2 Related Works

2.1 Bilingual Speech Production: Stage-Specific Cognitive Bottlenecks

Levelt's (1989) widely-adopted speaking model identifies three sequential stages in language production: the Conceptualizer (preverbal message generation), the Formulator (lemma selection and phonological encoding), and the Articulator (motor execution). For bilingual model (de Bot (1992)), language choice operates within the Conceptualizer (macroplanning level) through a language-specification component, while lemma selection and phonological encoding in the Formulator may activate both languages simultaneously or in language-specific way depending on proficiency and task demands.

Validated by recent studies ((Wolna et al. (2024) Felker et al. (2018)), cognitive bottlenecks differ between comprehension and production: comprehension engages inverse pathways where acoustic/orthographic input must be mapped to conceptual representations (bottom-up), whereas production requires conceptual representations to activate appropriate lexical forms and phonological encodings (top-down). Bilinguals also strategically switch languages depending on processing demands, showing highly adaptive inhibitory control (Li et al. (2024), Lee et al. (2025)). This interaction predicts that production support should prioritize conceptual clarity alongside form availability, creating dual-purpose scaffolding. In contrast, comprehension support can focus on predictive efficiency because there is no downstream formulation or articulation load.

2.2 Domain Complementarity, Conceptual Organization, and Language-of-Instruction Effects

The Complementarity Principle (Grosjean, 2016) establishes that bilinguals' languages are functionally distributed across communicative domains rather than uniformly deployed: medical terminology remains accessible through

Chinese for those educated in Chinese-medium institutions despite decades in English-speaking environments and substantial English medical competence. This persistence reflects language-dependent memory principles (LDM): retrieval efficiency increases when retrieval cues match the language encoding the original experience (Marian & Neisser, 2000; Tulving & Thomson, 1973).

Conceptual transfer theory (Jarvis, 2010) further elaborates this mechanism that domain knowledge becomes cognitively organized according to the linguistic structures, conceptual categories, and associative patterns of the original encoding language. Bilinguals develop entrenched patterns of categorization and construal through repeated language use within specific discourse communities; these patterns become cognitive routines that persist even when individuals become highly proficient in alternative languages. Critically, this entrenchment means that regular bilinguals function differently from trained translators: they often fail to retrieve precise translation equivalents for domain-specific concepts, particularly when only one language served as the medium of initial learning (Grosjean, 2016). In this case, specifying the language(s) through which domain knowledge was originally acquired may constitute a more powerful moderator of format effectiveness than conventional proficiency assessment metrics.

2.3 Cognitive Load Theory and Context-Dependent Load

Cognitive load theory (CLT) provides a framework for understanding how design decisions create cognitive load that interferes with comprehension and learning (Sweller et al., 2011). Recent research on bilingual subtitle design (Liao et al. (2020)) reveals that contrary to predictions assuming additive processing costs, bilingual subtitles enabled selective attention filtering, adopting one language as their primary information source while maintaining secondary engagement with the other. Moreover, bilingual subtitles produced superior meaning recognition and recall: attention to L2 target words predicted vocabulary retention gains, whereas attention to L1 translations did not (Wang and Pellicer-Sanchez (2022a,b)). This suggests that L1 translations can functionally reduce L2 engagement if presented equally saliently, but when positioned as secondary, dual-language presentation preserves L2 engagement while leveraging L1 semantic priming. These findings indicate that the bottleneck of processing dual-language presentation depends on whether L2 remains the primary engagement focus and whether languages provide complementary rather than redundant information.

2.4 Current Study

This study bridges these literatures by investigating whether and under what conditions bilingual support improves comprehension, production fluency, and learning in real-time AI-MC. By systematically examining three interacting factors (processing context, content characteristics, and individual differences), we operationalize prediction theory, encoding-specificity principles, and cognitive load theory into the context of bilingual interface design. Our investigation specifically addresses how proactive bilingual clarifications can optimize comprehension-phase processing (leveraging L1 prediction advantages) while maintaining production-phase fluency (avoiding L1 interference), and whether these benefits are moderated by prior experience and specific content.

3 Method

3.1 Participant

Ten NS of Chinese (M=; F=; age=[19, 26]) were recruited via the university SONA for extra course credits. Some (N=) grew up and received education in an English-speaking country before 12 years old. Most (N=) lived in an English-speaking country between 12 and 18 years old. A few (N=3) just arrived for college education. Participants were denoted by their participant ID and years spent in an English-speaking country (e.g., C3-7 = 3rd participant, 7 years living in English-speaking country).

3.2 Feature design description

We used an existing design framework [] with its clarification feature only to run a WoZ study with Figma with three essential setups: L1-only, L2-only, and bilingual, as in Figure 1. To reduce the adaptation effort, two interactive components were included: 1) a caption box, where the highlighted green word is clickable, and 2) a sidebar, where each clarification tab pops up separately in a temporal order. The tool was designed to be perceived as a fully automated system with all the clarification outputs proactively generated when clicked during the conversation without any explicit prompting.

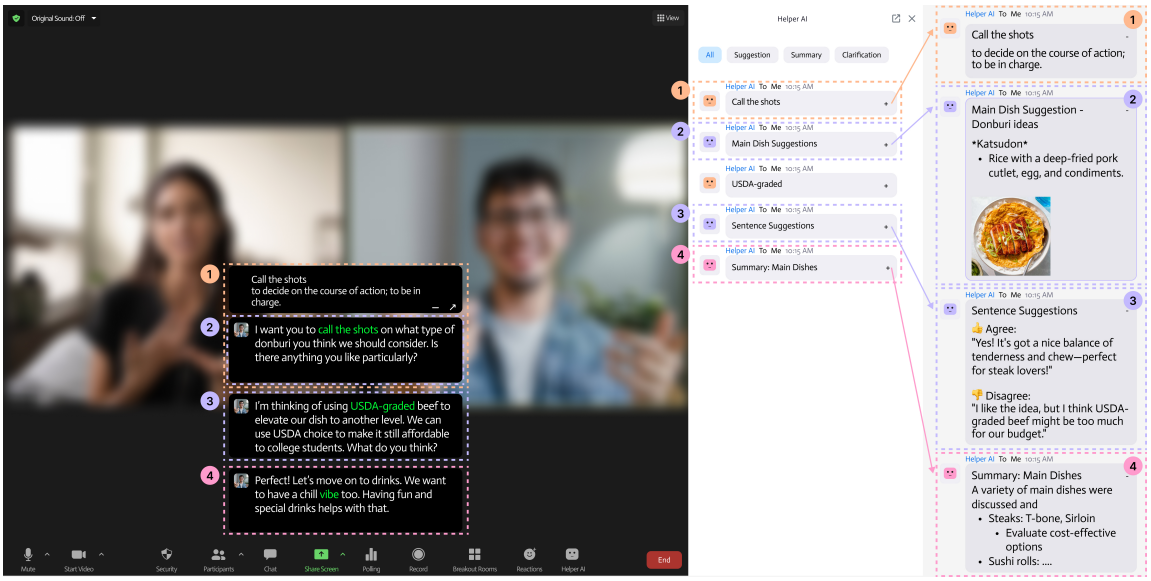

Fig. 1. Zoom interface with XPLAIN; 1: clarification box popped up above caption once clicked with minimal eye movement track, saved as minimized in the side bar for future reference; 2: idea suggestion; 3: full-sentence suggestion; 4: summary. 2 3 4 all popped up as expanded windows. Only one feature expands at a time to avoid visual overload. (N.B. the caption stacks were for illustration purposes: only one caption box stayed at the bottom position as the conversation progressed.)

3.3 Materials

The conversation topic was to have a virtual medical visit with a dietitian at university health center. The procedure was standardized by making the confederate part of the conversation fully pre-scripted to consistently guide the

conversation flow. Validated via pre-testing (N=3), this topic ensures medium-high L1 familiarity but low L2 familiarity for target participants growing up in a foreign country whilst gives naturalistic exchanges via open-ended questions.

We selected twelve terms commonly used by NS in everyday context related to diet and medical conversations. They were counterbalanced into four categories based on (1) processing context (comprehension vs. production) in the conversation, and (2) medical domain specificity (high vs. low). In the comprehension context, words were placed in the middle of the confederate's speech where there was no time pressure of response, whereas in the production context, words were placed in the middle of the question where the participant was required to take the turn over quickly. Each of the three words in each category were assigned to be (1) L1-only (Figure **??**), (2) L2-only (Figure **??**), or (3) bilingual (Figure **??**) formats. The order of words was counterbalanced in the conversation to avoid any ordering effect. The maximal length of clarifications was restricted to twenty words. All features were only applied to the zoom account of the participants and distributed evenly in the conversation to balance the cognitive load. More detailed materials can be found on Open Science Framework (link removed for anonymity).

We put every line of caption in the conversation as one page in Figma to avoid any extended reading of the confederate's speech. If the highlighted target term was the last few word of the sentence, we displayed them with the previous sentence to keep the processing window of similar length to other terms for participants to interact instead of skipping it due to the short window.

A semi-structured interview protocol was developed, asking participants to reflect on their communication experience: overall comprehensibility of the conversation, perceived effectiveness of each feature, change in behavioral engagement, change in communication effort, comfort with the interface, alongside their background and prior experience with AI.

3.4 Procedure

Two laptops joined over Zoom and connected in TeamViewer. Zoom windows were pinned on top of the Figma page's placeholder to mimic an online meeting platform integrated with XPLAIN. The confederate monitored and progressed the conversation via TeamViewer. Participants were instructed to have an unconstrained, naturalistic conversation, where they would be asked questions about their daily diet and exercise routine. They were advised not to share any personal health details if concerned. After consent, participants joined a demo session to familiarize and interact with the features and the interface (3 mins) before the official study. In the study session (12 mins), the confederate acted as a real dietitian and conducted an informal virtual consulting session with individual users. Users then shared their feedback via a semi-structured interview (15-30 mins).

4 Results

All interviews were recorded and transcribed by Otter.ai and manually reviewed for errors. Using an inductive and interpretive approach [1], general themes were first identified from the transcripts, with high-level themes and their relationships further constructed. Codes were refined through an iterative process of creating and combining themes.

4.1 Domain-Specificity and L1 Familiarity Modulate Clarification Format Effectiveness

Participants (8/10) reported substantially larger processing benefits with L1-only (Chinese) format for domain-specific terminology, especially in medical and technical contexts. They consistently attributed this advantage to their familiarity of processing L1 characters, more importantly, morphological transparency, the capacity of Chinese compound words to convey semantic meaning through character components. When terminology was familiar from past education and life experience in China, participants achieved rapid comprehension (5 seconds) with Chinese character-based

partial meaning inference, whereby individual character components conveyed semantic content even for unfamiliar compounds. One participant noted, "In Chinese, even if they don't know the whole word, the word is made up of a few characters, which helps as they recognize some of the characters" (C-participant).

Participants reported difficulty in processing English technical terms. Random switching of clarification languages disrupted comprehension as the format didn't match the language they were thinking in

So when I see the Chinese version, it feels more like familiar, and I can quickly understand the meaning of the words. But for the English version, I feel like, instead of maintain the conversation, I feel like, Oh, I do. I needed to learn new English word for me, it's like a learning interface. And for the bilingual, bilingual is also the same thing, because usually my daily life when I like writing or reading English paper or not reading, sometimes I will use the bilingual tools. I mean one size Chinese, another size English, so that interface for me is like learning something, and I feel I don't need it out during a conversation. (11111)

However, 2 of 10 participants with US-based technical education or high medical literacy preferred L2-only (English) format, suggesting that technical familiarity, rather than morphological transparency alone, moderates the effect. L1 advantage diminishes for domain-general vocabulary: half of the participants found non-technical terms equally understandable across formats, with 3 participants expressing sufficient contextual inference to eliminate need for explicit clarification. This pattern aligns with our H1a prediction (domain-specificity × L1 familiarity interaction).

However, H1c (low L1 familiarity eliminates L1 benefit) was only partially supported: when a domain-specific terminology was unfamiliar in both languages (e.g., "fennel"), participants (4/10) reported minimal L1 advantage, yet some (6/10) still benefited from rapid L1 scanning before reading L2 clarification under time pressure.

Notably, bilingual format (English + Chinese) served as a universal fallback (4 of 10 rated as most helpful overall) when terminology was unfamiliar in both languages. Participants (4/10) who rated it as the most helpful overall valued cross-language validation: "Bilingual helps more than Chinese only because some terms are unfamiliar even in Chinese; bilingual allows me to check both languages to confirm understanding" (35611). The benefit to facilitate learning with bilingual clarifications was also exclusively highlighted: 8 of 10 participants reported that cross-language semantic mapping enhanced their episodic encoding and terminology retention compared to single-language formats, such that when they could confidently skip the clarifications if they encountered the word for the second time in the same conversation.

4.2 Processing Context Determines Optimal Language Display Format

4.2.1 Comprehension contexts (partner speech, no immediate production demand): strong L1 advantage. Participants (8/10) demonstrated faster and more fluent comprehension with Chinese-only format during comprehension. Rapid processing (5 seconds for Chinese vs. 15+ seconds for English) enabled participants to complete clarification reading within natural conversational pauses, maintaining flow without articulation interruption. Participant estimated that they could read and understand Chinese in a second (35611, 34408, 94228), but needed at least 5 seconds for English (35611, 34408). The reduced cognitive load of L1 processing allowed concurrent attention to partner speech: "it might takes a few seconds to answer the question, but in general, it wouldn't disrupt a conversation...[it] give[s] me a moment to read this" (34561), in support of H2a (L1 advantage in comprehension).

4.2.2 Production contexts (response planning with immediate output requirement): conditional L2 advantage. Participants (5/10) exhibited preference for English-only clarifications during response formulation. One participant explained: "I used Chinese to understand the context and used English to respond... English also helped integrate into the conversation

directly." (29257) This reflects lexical-form alignment: English clarifications provided direct L2 lexical forms, preventing translation lag during speech output, such as any tip-of-the-tongue phenomena that frequently occur when semantic comprehension (via L1) requires remapping to L2 production forms.

However, participants (5/10) also showed flexible format selection dependent on in which language the term was originally learned. As participant 29257 also noted, "[Bilingual works better for production] because I used Chinese to understand the context and used English to respond, English also helped integrate into the conversation directly." This dual-processing pattern occurred when terminology was bilingual-accessible but stored preferentially in the L1 lexicon. This preference is especially prominent for participants educated in China who learned professional or domain-specific terms in Chinese: "[for] an international student just got here for college, from China, they will probably have a easier time understanding the technical terms in Chinese." (34495) while "if you have college here and so [for] some some words...the first time you know is in English, and you basically never know what is in Chinese." (35824). The difference in lexicon storage dependent on learning and living experience confounded pure processing-context predictions, suggesting H2b (L2 advantage in production) requires qualification by individual lexical storage patterns and educational background.

4.3 L2 Proficiency as Primary Individual-Difference Moderator

L2 proficiency strongly moderated format effectiveness across all contexts. Lower-proficiency participants (TOEFL equivalent 80-90) exhibited greater benefits from bilingual and L1-only formats: reduced anxiety, faster comprehension, higher confidence in response formulation. One lower-proficiency participant stated, "the one that I felt most anxious about, it's probably just like English description...due to lack of additional support", highlighting perception of L2-only as cognitively demanding (35644). In contrast, three higher-proficiency participants (TOEFL 95-100) with early English exposure expressed comfort with English-only format and sometimes skepticism toward L1 support as unnecessary redundancy: "because we're talking English, so I feel like switching is, I guess not destructive, but just unnecessary" (35824). This proficiency-moderation effect emerged consistently across both technical and non-technical vocabulary, suggesting general cognitive-processing differences rather than domain-specific effects.

Educational background and living experience reinforced proficiency effects. Participants (4/10) from immersive L1 environment demonstrated stronger preference for L1 terminology regardless of current L2 proficiency, indicating that the effect of instruction language persist despite years in English-speaking environments. One participant noted, "In contrast, two participants (34495, 35611) educated primarily in US-based schools found English equally transparent or preferred it for authenticity in academic and professional contexts, "It's also because of learning and living experience. Like, because you have college here and so some some words you the first time you know is in English, and you basically never know what is in Chinese" (34495). The proficiency-moderation effect aligns with expertise-reversal theory: scaffolds that benefit lower-proficiency learners can burden higher-proficiency users, in support of H3a.

Two participants explicitly stated that they separated lexicons based on when and how the words were learned. "when I'm learning English, I also memorize, like, most of my vocab in English with translation in Chinese. And that way, like, when I was, like, young, study English. I just know the Chinese word for every English word I know. But after knowing more words, I think I learned most of the words in English context, especially some of the words I don't really know how to explain in Chinese. So it really depends on the word, whether it's [what] I've memorized in English, or I just know in the context where I previously memorized with translation and Chinese" (34561). The effectiveness of clarification is dependent on acquisition history, semantic association and lexicon construction in each language domain.

4.4 Clarification Timing Impacted Conversation Flow

Clarifications did not significantly disrupt conversation flow or speech fluency when positioned appropriately. Most participants (9/10) reported no disruption to flow when clarifications appeared end-of-turn or during response-planning windows. Half of the participants explicitly noted non-disruptive experience: "I don't think it was disruptive at all, to be honest, because I honestly didn't really understand most of the terms that were highlighted green. I'm not sure that was intentional, but there [were] other terms that [I] also don't understand, but the ones that were highlighted were really helpful...And I actually appreciate [if there could be] more terms [clarified]" (35611). Reading clarifications of the optimal format can occur concurrently with response preparation within the time-frame of natural conversational pauses: "when I read really quickly, it would not, definitely not, hurt or disrupt the conversation. I could just make the conversation continue. " (34408).

However, one participant experienced notable disruption not before production, but during comprehension when clarifications appeared mid-sentence: "Production is at the end...and then for comprehension in the middle of the sentence, and you have to stop and think about it. And then sometimes you miss...what you're actually talking about" (35824). Mid-sentence clarifications interrupted the natural speech rhythm by forcing explicit processing interruptions during partner utterances, suggesting that temporal positioning critically impacts fluency.

Participants (3/10) also reported that moving mouse around to click on the highlighted word deviated their attention and caused psychological insecurity: "the only time I was like slight panic was when one of the words didn't pop up on the side, so I had to click the green thing which was moving the mouse more than I thought I would...because I thought all I had to do is click on the tab when it popped up." (94228), suggesting automatic pop-up design is preferred over spontaneous clicks to minimize interactions. (11111, 94228, 96904)

4.5 Learning and Memory: Bilingual Advantage Over Single Formats

Learning benefits varied systematically by format. Bilingual format showed strongest learning support (8/10), with cross-language semantic mapping enhancing episodic encoding. Participants reported, "[Bilingual format] is really helpful for helping people learn and memorize the concept, having value what it really helps memorize the terms. Because if you just have Chinese. I would just like, learn the term in Chinese. Like, I remember, at the end of a survey, I was asked to recall, like, what those terms mean. Like, I will appreciate if I actually learn this in English that I know, like, what exactly, too much. But if I learn this in Chinese, I like, okay, yeah, I understand what it says. But if you put up like, different affiliations in English, I'm like, okay, like, what? Like, I actually don't know. So that's why I say like, if you have like, two different things. It helps with it helps you with your learning, go into different format and like, actually talking to a doctor when there's, like, no transcription whatsoever, I think that's the best way for me to learn. Like, if I were to, like, just learn this in Chinese, I would process everything in Chinese and not, like, really leaning on the English part. So that's why I said like will be more helpful, even though it took extra efforts to process everything" (35611). However, for low proficiency speakers, they were unable to handle simultaneous learning during the conversation, "I mean one Chinese, another English, so that interface for me is like learning something, and I feel I don't need it out during a conversation" (11111).

Chinese-only format particularly supported memory retrieval and retention for terminology or concept already familiar in L1 with reduced mental effort (6/10). One paricipant noted that "I still like to learn some of the words in a Chinese context, because you don't have to memorize more English words when you memorize the meaning of it, which, if I don't know the English word, I would say I probably know what it is in Chinese, because I think my

Chinese vocab is better than my English vocab.. in that way, it helps me learn", due to existing conceptual depth in L1 that enables elaborative encoding (34561). English-only format was preferred for learning unfamiliar terminology (2/10), possibly due to explicit structural clarity and lower translation cost of English definitions that support first-time learning: "I think it's much more simpler [in] English. Like, that's that much like, it's not my like, my generic stuff and anything like using as I have like, a systems to talk about is so actually, I want to speak English, I would say is easier to for me to remember something." (C-participant).

For participants with mixed Chinese-English education (5/10), bilingual format aligned naturally with mental translation and code-switching patterns in their daily life: "there is not really any tension between what makes you feel comfortable and what helps to learn, because bilingual can resolve everything" (94228). Participant 29257 explained: "I went to school in China where they would have Chinese teachers teaching in Chinese and Canadian teachers teaching in English so Chinese/bilingual fit the best and was most effective for learning." This finding suggests format selection should consider both immediate comprehension demands (Chinese advantages for speed) and longer-term learning goals (bilingual optimal for retention).

4.6 Context Appropriateness: Differential Motivation Across Settings

Medical contexts showed highest tool motivation (8/10), driven by high stakes, unfamiliarity, knowledge gap, and power imbalance between patient and doctor. Participants explicitly valued clarification support for medical contexts: "I also go to a doctor most of time... Like, completely really confused... I think medical is the most applicable scenario" (35611). One participant noted, "when it comes to conversations where the topic you're not really familiar with, like going to doctors or having a medical conferences would be really helpful, because even I go to University health, I still need to look up some words before I go over a lot of clarifications with the nurse or doctors. So having those tools for those topics will be really helpful", reflecting real-world comprehension challenges in high-stakes encounters (34561).

Participants (2/10) also showed format preferences depending on their cultural understanding in social contexts. Participant explained clearly,

> "the formats really were a little bit dependent on cultural context of the term then. So for instance, some of the terms were more specific to American culture...you might need more English versus something straight up [that is] confusing to both native speakers and non native speakers. So then that might be something that you might just want to get in Chinese...Other things is that some medical or scientific things from English that directly translate to Chinese, it's even harder for for people to understand." (34495)

The clarification format should take into account of the origin and background of the term, modulated by the user's knowledge and cultural background, where untranslated terms may show greater semantic transparency for understanding in domain-specific contexts.

The preference of how detailed the clarification should be is also subject to the scenario. As participants (2/10) (11111, 34495) noted, "if...in the class, I have to learn something, then the clarification part definitely helps me a lot. But for this case...it's a doctor patient conversation. So as long as I can respond to the doctor correctly, appropriately, that's enough. I don't have to understand, really understand what the word means." (11111) Other participants (4/10) showed lower motivation unless material was novel or unfamiliar.

L2 proficiency (7/10) aligns with expertise-reversal theory: scaffolds for lower-proficiency learners can burden higher-proficiency users. Participant 34561 (high) expressed preference for independence: "it's really helpful to have

like what it means in Chinese" for confirmation rather than primary support. Participant 35611 (lower) relied on L1 primarily: "the one that I felt most anxious about, it's probably just like English description."

Educational history emerged equally powerful as proficiency. Participants educated in China (29257, 35611) showed persistent L1 preference regardless of years in English-speaking country, whereas those educated in English (34561) showed flexibility. This suggests language-of-instruction creates durable effects on knowledge organization.

Domain expertise showed weak effects (3/10): confounded with lexical-familiarity location rather than true expertise advantage. Participant 35611 (medical background) still preferred L1 Chinese for medical terms, indicating familiarity location in L1 (not expertise) predicts preference.

5 Discussion

5.1 Context-Dependent Format Optimization

This study instantiates language processing theories in bilingual real-time AI-MC, showing that optimal clarification format depends critically on the interaction among processing context, individual knowledge representation, and working memory constraints. The strength of comprehension-context L1 advantage versus moderate production-context L2 advantage suggests that comprehension-phase optimization is more robust across individuals, while production-phase optimization is more individual-specific, contingent on educational history and lexical storage patterns.

During comprehension, L1 (Chinese) format leverages three mechanisms: 1) **rapid character-based processing** via phono-morphological transparency (0.5-2s); 2) **partial meaning inference** enabling contextual-based comprehension before full lexical retrieval, and 3) **direct conceptual priming** reducing decoding effort versus L2 lexical retrieval (>5s). This operationalizes prediction theory's priming mechanism: L1 triggers existing conceptual representations more directly than L2 translation without explicit decoding effort, thereby improving processing efficiency by reducing cue-integration latency with lower extraneous cognitive load to transform visual input into conceptual representation.

In contrast, direct L2 form access in production prevents tip-of-the-tongue phenomena by maintaining L2 lexical retrieval pathways active during speech formulation. The intuitive switch from L1 for comprehension to L2 for production showed a spontaneous reduction of extraneous cognitive load incurred by the mismatch between perceived lexical forms and mental representation within each processing phase. This finding explains phases-specific preference where users optimize the trade-off between comprehension-latency load and production-form-maintenance.

However, individual lexical-storage effects create heterogeneity, where those who learned terms in L1 maintained L1 preference even for production, creating persistent retrieval-path advantages based on acquisition history independent of their current L2 proficiency. This lexical-organization principle suggests that production-phase optimization require alignment between conversational context and knowledge representation within different domains and languages, demonstrating a novel specificity in time-sensitive AI-MC contexts. Moreover, this study reveals extraneous load has context-dependent sources (retrieval-time cost for comprehension; form-maintenance conflict for production). Optimal extraneous-load reduction thus requires context-specific intervention.

5.2 Knowledge Representation Architecture and Language-Dependent Memory

Beyond processing context, knowledge representation location—operationalized as language of formal domain education—emerged as the second primary predictor of format preference and more stable than current proficiency level. This finding operationalizes LDM principles in AI-MC context.

China-educated participants maintained consistent L1 preference regardless of proficiency level and living experience in L2 environment. This pattern contradicts a surface interpretation that higher proficiency should reduce L1 dependence. Instead, it supports encoding-specificity: knowledge initially encoded through L1 linguistic concepts creates durable retrieval pathways specifically linked to L1 cues. Later exposure to L2 equivalents creates parallel but weaker representations because encoding conditions were fundamentally different. However, US-educated participants showed substantially greater format flexibility during comprehension and preferentially switched to English during production, suggesting that knowledge encoded through L2-medium instruction activates more efficiently via L2 cues. This contrast indicates that language-of-acquisition is a strong predictor of domain-specific format preferences.

5.3 The Interaction: Processing Context × Knowledge Representation

The most theoretically novel finding emerges from examining the interaction between processing context and knowledge representation, where the optimal format emerges from their joint optimization.

5.3.1 Comprehension-phase L1 advantage is robust to knowledge representation variation. Comprehension-phase L1-preference remains robust regardless of educational history because L1 processing advantages (i.e., morphological transparency, direct priming) operate independent of the original encoding language. The observed L2 preference for domain-specific terms suggests that educational history modulates long-term preference, but processing efficiency still drives short-term comprehension choices when both languages were available.

5.3.2 Production-phase is contingent on knowledge representation. The heterogeneity in production-phase preferences reflects educational-history contingency: participants with L1-encoded knowledge maintained L1 for production, where lexical-storage accessibility outweighed output-language-alignment benefits, while participants with L2-encoded knowledge switched to English for production, where output-language-alignment outweighed translation-effort costs.

5.3.3 Proficiency modulates interaction magnitude but not direction. Lower-proficiency participants experienced the interaction more acutely: their working memory constraints and fragile L2 representations made L1-L2 mapping more effortful. Higher-proficiency participants showed more flexibility but maintained the same directional interaction: comprehension still preferred L1 for efficiency, and production still followed lexical-storage pattern.

5.3.4 Learning objectives alter interaction. When explicit learning was a goal (on top of or instead of real-time comprehension/production), higher-proficiency participants leveraged bilingual format to establish dual memory traces (consistent with Paivio's dual-coding theory; Paivio, 1986). This suggests a three-way interaction: processing context × knowledge representation × learning objectives. When time pressure is relieved or when cognitive load is manageable, bilingual format enabled dual-encoding benefits precisely because existing knowledge representation is robust enough to support selective attention to both languages without extensive mental effort.

5.4 Design Implications: Context-Adaptive Bilingual Clarification Systems

Our findings support design principles for context-adaptive bilingual clarification systems. First, the dominant L1-advantage in comprehension suggests that implementing **L1 as default clarification language** for comprehension phases and low-stress conversations provides universal benefit across proficiency levels, educational backgrounds, and lexical familiarity. Given the difference in knowledge representation structure, high-proficiency users or those with English-based education benefit more from **bilingual format** with reduced extraneous load of mental translation. Proficiency assessment and education history documentation during onboarding can adapt the metrics from defaults.

Participants reported that when they "roughly knew" a term, clarifications provided confidence confirmation rather than enabling comprehension. A **dual-level clarification support** can reduce anxiety about potential misunderstanding: (1) lightweight confirmation mode for partially-familiar terms (1–2 word definition, no elaboration), and (2) full explanation mode for unfamiliar terms (definition, context). This addresses both confidence-building and learning functions without imposing equal cognitive load on both. Moreover, the higher engagement rate of technical and medical terms suggests **smart highlighting** that prioritizes domain-specific terms to domain-general ones. Low-frequency domain-specific terms should receive automatic clarification, whereas high-frequency domain-general terms can require user action to reduce information overload.

Participants in fast-paced conversational contexts reported that even L2 clarifications (which matched conversation language) felt disruptive due to the reading time incompatible with rapid turn-taking. This highlights the need of **time-pressure detection via interaction latency** (e.g., response time between speaker turn-ending and listener turn-starting, disfluency rate of the speaker, etc.). In high-pressure contexts, the system can modulate the availability and frequency of clarifications accordingly to real-time feedback. In low-pressure contexts, the system can provide more details or switch to **bilingual mode for dual encoding benefit** when learning goals are present alongside comprehension demands (e.g., asynchronous lecture, professional colloquium).

6 Limitations and Future Directions

First, this qualitative analysis lacks quantitative insights to validate the claims and the magnitude of the effect observed. Upcoming analyses of speech output, reading time, and eye gaze data from video recordings would quantify processing difference of L1-L2 format beyond self-report to substantiate our findings with precisions.

Current sample (N=10 for qualitative interviews) enables hypothesis-grounded findings but limits generalization beyond bilingual undergraduates at US universities. Future studies should test diverse age groups, professions (medical personnel, financial professionals), and language pairs to assess generalizability.

In the current design, knowledge representation and processing context were both operationalized dichotomously. More fine-grained representation mapping via recognition tasks or word-association tasks would provide continuous measures that lead to quantifiable findings.

The confound between processing context and lexical-storage location (both correlate with optimal format) necessitates future designs that experimentally manipulate these factors independently, perhaps by testing participants on terms learned in different languages versus production contexts fixed experimentally.

7 Conclusion

This study demonstrates that processing context and knowledge representation independently modulate optimal clarification format in bilingual CMC. Comprehension-phase L1-preference is robust and universal, reflecting efficiency advantages in L1 morphological access and automatic priming. Production-phase preferences are individual-contingent, reflecting whether domain-specific terminology was encoded in L1 or L2 education. These findings operationalize established psycholinguistic principles (i.e., language-dependent memory, encoding specificity, comprehension-production asymmetries) to the CMC/AI-MC system design context. The key design implication is that conventional proficiency-only personalization is insufficient; profiling mechanisms that track knowledge representation structures improve design robustness for high-precision personalization. Implementation requires moving beyond universal design toward evidence-based personalization accounting for both processing context and educational history. These findings contribute empirically-grounded design principles for developing inclusive, context-adaptive AI systems that enhance

non-native speakers' participation in real-time multilingual communication without compromising fluency, autonomy, or engagement.

References

[1] Barney G. Glaser and Anselm L. Strauss. 2017. *The discovery of grounded theory: strategies for qualitative research.* Routledge, London New York.