Context-Adaptive Optimization for Bilinguals in Processing Time-Pressured AI-Mediated Turn-Taking

ANONYMOUS AUTHOR(S)

Real-time turn-taking in AI-mediated communication (AI-MC) intensifies cognitive demands for second-language speakers, especially when unfamiliar and domain-specific terminology must be understood and responded under time pressure. This paper tested the perceived effectiveness of different display format (L1-only, L2-only, bilingual) of proactive clarifications during dyadic virtual medical consultations with Chinese L1 / English L2 speakers. A lab-based Wizard-of-Oz study using qualitative analysis shows a robust comprehension-phase advantage for L1 clarifications, while production-phase preferences for L2 emerge only when knowledge representation is encoded primarily in L2. The observed pattern showed a three-way interaction of processing context, knowledge representation, and prior language experience. Moreover, bilingual format best supports retention and retrieval when concurrent learning goals are salient but overloads lower-proficiency users with mental stress. These findings argue for context-adaptive, representation-aware bilingual support that tailors real-time mediation support to processing phase, knowledge-encoding history, and situational stakes for inclusive AI-MC.

CCS Concepts: • **Human-centered computing** → **User studies**; **HCI theory, concepts and models**.

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

ACM Reference Format:

Anonymous Author(s). 2026. Context-Adaptive Optimization for Bilinguals in Processing Time-Pressured AI-Mediated Turn-Taking. 1, 1 (May 2026), 7 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, domain-specific terminology creates additional lexical and conceptual processing deficits. While studies on L2 has extensively documented processing advantages of first language (L1) , whether and when L1 versus L2 would benefit different phases of processing real-time communication remains under-specified.

Classical theories of language processing lay the theoretical grounding but not empirical guidance on how to effectively scaffold L2 in AI-MC. **Prediction theory** [21] posits that comprehension operates through rapid predictions of upcoming input, where L1 should provide processing efficiency advantages through direct conceptual priming. However, this theory has not been operationalized to AI-MC where interpretation and evaluation delayed such planning. Similarly, **cognitive load theory** (CLT; [24]) explains how extraneous load imposed by design can interfere with learning and performance, yet provides limited guidance on how context-dependent demands create context-dependent load sources. **Language-dependent memory theory** (LDM; [18, 25]) predicts that memory retrieval efficiency

Manuscript submitted to ACM

increases when retrieval-cue language matches with the original encoding. However, the magnitude and persistence of such encoding-specificity effects across individual differences remains unknown in real-time settings.

Technologies supporting L2 communication often overlook the time-sensitive need of turn-taking and treat them as uniform challenges requiring generic solutions [3, 15, 17, 20, 22]. In real-time support, the core design problem is how to schedule bilingual display so that it reduces processing cost without creating persistent redundancy [9]. This study bridges these works by investigating whether and how bilingual support improves comprehension, production, and learning in real-time AI-MC. By systematically examining processing context, knowledge representation, and individual differences, we operationalized psycholinguistic theories to provide insights for bilingual system design.

2 Related Works

In traditional speech models, [1, 2, 11–13] comprehension maps acoustic/orthographic input to conceptual representations, whereas production requires conceptual representations to activate appropriate lexical forms and phonological encodings [7, 23]. Specifically in L2 studies, comprehension recruits automatic priming mechanisms, while production requires active lexical-form retrieval and execution control [10, 14]. **The Complementarity Principle** [5] establishes that bilinguals' languages are functionally distributed across communicative domains rather than uniformly deployed. In the same line, **conceptual transfer theory** [8] elaborates this mechanism: domain knowledge becomes cognitively organized according to the linguistic structures, conceptual categories, and associative patterns of the original encoding language. Crucially, regular bilinguals differ from trained translators—they often cannot retrieve precise translation equivalents for domain-specific concepts learned in only one language [5]. This suggests that specifying the language(s) through which domain knowledge was originally acquired may be a more powerful moderator of format effectiveness than conventional proficiency metrics alone. **Cognitive load theory** (CLT) [24] provides a framework for understanding how design decisions create cognitive load that interferes with comprehension and learning. Bilingual subtitle enabled selective attention filtering [16] and produced superior meaning recognition and recall [26, 27]. 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. This motivates a conversation-centric design question: **during turn-taking, what display format should appear at which moment, and for what content?** This study investigates how these foundational theories map onto the practical design problem of optimizing language displays in AI-mediated turn-taking. Rather than seeking a universal "best format," we hypothesize that optimal format emerges from the interaction among (1) **processing context** (comprehension vs. production phases), (2) **content characteristics** (domain-specificity and prior familiarity), and (3) **individual differences** (proficiency and language of prior education).

3 Method

3.1 Participant

Ten L1 Chinese speakers (M=2; F=8; age=[19, 26]) were recruited via university system for extra course credits. Participants were denoted by ID and years spent in an English-speaking country (e.g., C3-3 = 3rd participant, 3 years).

3.2 Materials

We used an existing design framework [6] with its clarification feature to run a Wizard-of-Oz (WoZ) study in Figma with L1-only, L2-only, and bilingual displays (Figure 1). 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.

The topic was a virtual medical visit with a dietitian at University Health. Confederate speech was standardized as fully pre-scripted to consistently guide the conversation. Validated via three pre-testings, this topic ensures medium-high L1 but low L2 familiarity for target users from a foreign country with naturalistic exchanges via open-ended questions.

We selected twelve terms commonly used in diet and medical conversations. They were counterbalanced into four categories based on processing context and medical domain specificity. For comprehension context, words appeared in confederate's speech without response demand; for production context, words appeared in question with response pressure. Each of the three words in each category were assigned to be L1-only, L2-only, or bilingual formats (Fig. 1). The order of words was counterbalanced in the conversation to avoid any ordering effect. The maximal length of clarifications was restricted to twenty words. Clarifications were only applied to participant accounts and distributed evenly in the conversation to balance the cognitive load. Every line of caption in the conversation was one page in Figma to avoid any extended reading of the confederate's speech. If the target term was the last few word of the sentence, we displayed them with the previous part to keep a similar reading window to other terms to avoid unintended skipping.

A semi-structured interview protocol was developed, asking participants to reflect on their communication experience, such as perceived effectiveness of each format, change in behavioral engagement, change in communication effort, comfort with the interface, alongside their background and prior language experience.

3.3 Procedure

Two laptops joined over Zoom and connected in TeamViewer. Zoom windows were pinned on top of the Figma page's placeholder to mimic that the tool was integrated in Zoom. A confederate monitored and progressed the conversation via TeamViewer. Participants were instructed to have an unconstrained, naturalistic conversation with questions about their daily diet and routine. They were advised not to share any personal 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

Inductive interview analysis [4] revealed consistent, **phase-specific preferences** for clarification format, moderated by proficiency and knowledge-encoding history. Participants (8/10) reported substantially larger processing benefits with L1-only (Chinese) format for domain-specific terminology during comprehension, consistently attributing this advantage to faster reading and morphological transparency that enabled partial inference even when the term was unfamiliar. For more domain-general vocabulary, they reported minimal differences across formats or relied on conversational context rather than clarification. When a term was unfamiliar in both languages, bilingual display functioned as a reliable fallback because it allowed cross-language checking and semantic confirmation with psychological safety (2/10).

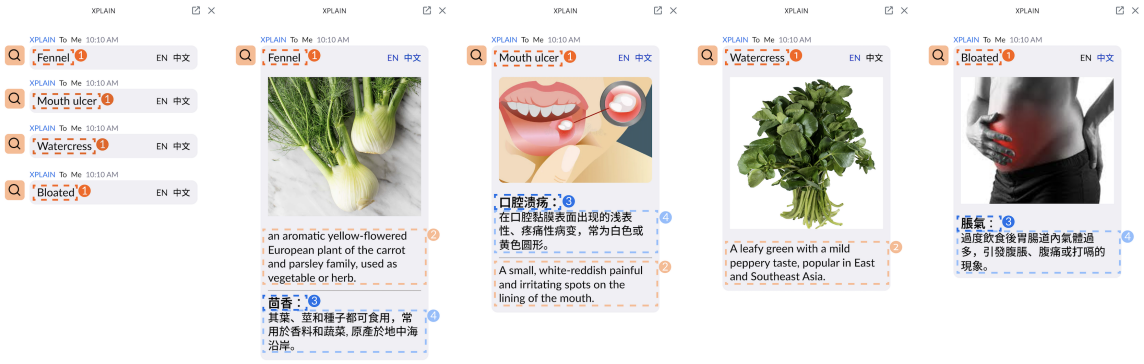

Fig. 1. Clarificationss in Zoom sidebar: the first column shows a sequence of minimized history of terms clarified. Each of the formats is illustrated—bilinguals, L2-only and L1-only, where 1: English term (L2); 2: English clarifications; 3: Chinese term translation (L1); 4: Chinese clarification. 2 3 4 all popped up as expanded tabs. Only one tab expands at a time to avoid visual overload. Participants can click the icon on the top right corner to view the other language when provided with a single-language clarification.

Processing context strongly shaped perceived optimal format. During comprehension, most participants (8/10) described faster uptake and smoother flow with L1 format because they could read within natural pauses and keep attention on partner speech. In production contexts with immediate output pressure, participants (5/10) preferred L2 format because it offered direct access to lexical forms and reduced mental translation cost. However, production preferences were not uniform: participants frequently selected formats based on which language the concept was originally learned in, "if you have college here and so [for] some words...the first time you know is in English, and you basically never know what is in Chinese." (C2-6). This pattern complicates a simple “L2 for production” rule and points to educational history/encoding language as a key moderator in production-phase optimization.

L2 proficiency, education background, and living experience 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 (C3-3). 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: "because we’re talking English, so I feel like switching is...not destructive, but just unnecessary" (C1-7). This proficiency-moderation effect emerged consistently across both technical and non-technical vocabulary, suggesting general cognitive-processing differences rather than domain-specific effects. Beyond proficiency, participants’ language of prior education and living experience shaped stable preferences, consistent with the idea that domain knowledge organization influences which cues facilitate rapid retrieval under time pressure.

Clarifications did not significantly disrupt conversation flow when **positioned at appropriate timing**. Most participants (9/10) reported no disruption 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...because I honestly didn’t really understand most of the terms that were highlighted green...And I actually appreciate [if there could be] more terms [clarified]" (C10-2). 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. " (C8-5). Some

participants (3/10) also noted that mouse movement and clicking could induce attentional shifts and brief insecurity, suggesting that reducing interaction cost matters for preserving speech flow.

Perceived learning and memory benefits varied by format and by whether learning was an active goal during the conversation. Bilingual format was most often associated with stronger retention because cross-language mapping supported encoding and later recall, particularly for users who could allocate attention to both languages without overload (8/10). For example, C8-5 noted that "there is not really any tension between what makes you feel comfortable and what helps to learn, because bilingual can resolve everything". Meanwhile, lower-proficiency participants (5/10) described bilingual display as feeling like “learning mode” that was too effortful during live interaction, indicating a tradeoff between immediate fluency and long-term acquisition. Single-language formats supported learning when the target concept was already well-grounded in that language (L1 for familiar concepts) or when participants wanted English-only exposure for first-time learning and future use in English contexts.

Finally, **scenario appropriateness** shaped motivation to use the tool. Medical and other high-stakes contexts elicited the strongest desire (8/10) for clarification due to unfamiliar terminology, perceived power imbalance, and fear of misunderstanding. Participants (4/10) also suggested that cultural grounding of terms matters: some concepts were seen as more interpretable in English due to U.S.-specific framing, while other technical translations were perceived as harder to understand literally across languages. Together, these reports indicate that format choice should adapt not only to processing phase and user characteristics but also to domain stakes and term origin.

5 Discussion

Overall, the strong 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.

In comprehension, L1 format leverages rapid character-based processing via phono-morphological transparency, partial meaning inference enabling contextual-based comprehension, and direct conceptual priming with lower retrieval effort. This operationalizes prediction theory: L1 triggers existing conceptual representations more directly than L2 translation without explicit decoding effort, thereby reducing cue-integration latency. Whereas in production, the switch to L2 showed user’s spontaneous reduction of extraneous cognitive load incurred by the mismatch between lexical forms and mental representation. This finding explains phases-specific preference where users optimize the trade-off between latency and language consistency.

Knowledge representation location (i.e., language of domain education) emerged as the second primary predictor of format preference and more stable than proficiency level, which operationalizes LDM principles in AI-MC context. L1-educated participants maintained consistent L1 preference regardless of proficiency level and living experience in L2 environment, showing a deeper level of specificity in knowledge encoding. However, L2-dominants showed greater format flexibility during comprehension and preferentially switched to L2 during production. This lexical-organization principle suggests that production-phase optimization require alignment between context and knowledge representation within domain and languages, a novel specificity in time-sensitive AI-MC contexts.

The interaction between processing context and knowledge representation reveals a novel theoretical finding. General L1-preference in comprehension remains robust because morphological transparency and direct priming operate independently of the original encoding language. The heterogeneity in production-phase preferences reflects educational-history contingency: users with L1-encoded knowledge maintained L1 for production, where lexical-storage

accessibility outweighed output-language-alignment benefits, while users with L2-encoded knowledge switched to English for production, where output-language-alignment outweighed translation-effort costs.

When explicit learning was a goal alongside conversation, higher-proficiency users leveraged bilingual format to establish dual memory traces, consistent with Paivio's dual-coding theory [19]. 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.

5.1 Design Implications: Context-Adaptive Bilingual Clarification Systems

First, our findings support implementing **L1 as default** for comprehension and low-stress conversation with universal benefit. Proficiency assessment and education history documentation can adapt the metrics from defaults for high-proficiency users or those with English-based education to reduce extraneous load of mental translation.

A **dual-level clarification support** can reduce reported anxiety of misunderstanding and provide desired confidence confirmation: (1) lightweight confirmation mode for partially-familiar terms , and (2) full explanation mode for unfamiliar terms. **smart highlighting**: low-frequency domain-specific terms should receive automatic clarification, whereas high-frequency domain-general terms can require user action to reduce information overload.

The observed L2 disruption of L2 format highlights the need of **time-pressure detection via interaction latency**. In high-pressure contexts, the system can modulate the availability and frequency of clarifications accordingly to real-time feedback. In low-pressure contexts, it can provide more details or switch to **bilingual mode for dual encoding benefit** when learning goals are present alongside comprehension demands.

6 Conclusion, Limitations and Future Directions

This study shows that processing context and knowledge representation independently modulate optimal clarification format in bilingual AI-MC. 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.

This study could incorporate quantitative insights to validate participants' self-estimated claims and the magnitude of the effect observed. Upcoming analyses of speech output, reading time, and eye gaze would substantiate our findings with empirical precisions. Moreover, knowledge representation and processing context were both operationalized dichotomously. More fine-grained representation mapping would provide continuous measures for quantifiable results. Lastly, current sample limits generalization beyond bilingual undergraduates at US universities. Future studies should test diverse age groups, professions, and language pairs to assess generalizability.

These findings bridge established psycholinguistic principles, where instead of conventional proficiency-centered personalization, profiling mechanisms that track knowledge representation structures improve design robustness for high-precision personalization. The study contributes empirically-grounded design principles for developing inclusive, context-adaptive mediation *in the moment* of AI-MC without compromising fluency, autonomy, or engagement.

References

[1] Kees De Bot. 2020. A bilingual production model: Levelt's' speaking'model adapted. In *The bilingualism reader*. Routledge, 384–404.

[2] Kees De Bot and Robert Schreuder. [n. d.]. Word production and the bilingual lexicon. *The bilingual lexicon* 191 ([n. d.]), 214.

[3] Wen Duan, Naomi Yamashita, Yoshinari Shirai, and Susan R. Fussell. 2021. Bridging Fluency Disparity between Native and Nonnative Speakers in Multilingual Multiparty Collaboration Using a Clarification Agent. In *Proceedings of the ACM on Human-Computer Interaction*, Vol. 5. 1–31.

doi:10.1145/3479579

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

[5] François Grosjean. 2016. The Complementarity Principle and its impact on processing, acquisition, and dominance. *Language dominance in bilinguals: Issues of measurement and operationalization* (2016), 66–84.

[6] Wanqing Psyche He and Susan R. Fussell. 2025. Proactivity in Scaffolding Comprehension and Production in Real-Time Turn-Taking: A Case Study of Bridging Communication Gaps for Non-Native Speakers. In *Companion Publication of the 2025 Conference on Computer-Supported Cooperative Work and Social Computing (CSCW Companion '25)*. Association for Computing Machinery, New York, NY, USA, 477–482. doi:10.1145/3715070.3749273

[7] Gregory Hickok and David Poeppel. 2007. The cortical organization of speech processing. *Nature reviews neuroscience* 8, 5 (2007), 393–402.

[8] Scott Jarvis. 2010. Comparison-based and detection-based approaches to transfer research. *Eurosla yearbook* 10, 1 (2010), 169–192.

[9] Geza Kovacs. 2013. Smart subtitles for language learning. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. 2719–2724.

[10] Yun-Wei Lee, Patrick Rebuschat, and Aina Casaponsa. 2025. Top-down and bottom-up bilingual speech production: The effects of language context on inhibitory control. *Bilingualism: Language and Cognition* (08 2025), 1–13. doi:10.1017/S1366728925100357

[11] Willem JM Levelt. 1993. *Speaking: From intention to articulation.*

[12] Willem JM Levelt. 1995. The ability to speak: From intentions to spoken words. *European Review* 3, 1 (1995), 13–23.

[13] Willem JM Levelt, Ardi Roelofs, and Antje S Meyer. 1999. A theory of lexical access in speech production. *Behavioral and brain sciences* 22, 1 (1999), 1–38.

[14] Chuchu Li, Katherine J Midgley, Victor S Ferreira, Phillip J Holcomb, and Tamar H Gollan. 2024. Different language control mechanisms in comprehension and production: Evidence from paragraph reading. *Brain and language* 248 (2024), 105367.

[15] Lizi Liao, Grace Hui Yang, and Chirag Shah. 2023. Proactive Conversational Agents in the Post-ChatGPT World. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, Taipei Taiwan, 3452–3455. doi:10.1145/3539618.3594250

[16] Sixin Liao, Jan-Louis Kruger, and Stephen Doherty. 2020. The impact of monolingual and bilingual subtitles on visual attention, cognitive load, and comprehension. *The Journal of Specialised Translation* 33 (2020), 70–98.

[17] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 5286–5297. doi:10.1145/2858036.2858288

[18] Viorica Marian and Ulric Neisser. 2000. Language-dependent recall of autobiographical memories. *Journal of Experimental Psychology: General* 129, 3 (2000), 361.

[19] Allan Paivio. 1990. *Mental representations: A dual coding approach.* Oxford university press.

[20] Zhenhui Peng, Yunhwan Kwon, Jiaan Lu, Ziming Wu, and Xiaojuan Ma. 2019. Design and Evaluation of Service Robot's Proactivity in Decision-Making Support Process. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–13. doi:10.1145/3290605.3300328

[21] Martin J. Pickering and Simon Garrod. 2007. Do people use language production to make predictions during comprehension? *Trends in Cognitive Sciences* 11, 3 (2007), 105–110. doi:10.1016/j.tics.2006.12.002

[22] Evan F. Risko and Sam J. Gilbert. 2016. Cognitive Offloading. *Trends in Cognitive Sciences* 20, 9 (2016), 676–688. doi:10.1016/j.tics.2016.07.002

[23] Dorothee Saur, Björn W Kreher, Susanne Schnell, Dorothee Kümmerer, Philipp Kellmeyer, Magnus-Sebastian Vry, Roza Umarova, Mariacristina Musso, Volkmar Glauche, Stefanie Abel, et al. 2008. Ventral and dorsal pathways for language. *Proceedings of the national academy of Sciences* 105, 46 (2008), 18035–18040.

[24] John Sweller, Paul Ayres, and Slava Kalyuga. 2011. The expertise reversal effect. In *Cognitive load theory*. Springer, 155–170.

[25] Endel Tulving and Donald M Thomson. 1973. Encoding specificity and retrieval processes in episodic memory. *Psychological review* 80, 5 (1973), 352.

[26] Andi Wang and Ana Pellicer-Sánchez. 2022. Incidental vocabulary learning from bilingual subtitled viewing: An eye-tracking study. *Language Learning* 72, 3 (2022), 765–805.

[27] Andi Wang and Ana Pellicer-Sánchez. 2023. Examining the effectiveness of bilingual subtitles for comprehension: An eye-tracking study. *Studies in Second Language Acquisition* 45, 4 (2023), 882–905.