Difference in the Cognitive Mechanism of Predictive Processing in Computer-Mediated Communication: A Comparison Study of L2 Speakers

Wanqing He, Susan R. Fussell

Department of Information Science, Cornell University
{wh385, sfussell}@cornell.edu

Abstract

Previous studies of predictive mechanisms in computer-mediated communication (CMC) suggested that native speakers (L1) rely on auditory cues and emotion in conversation processing. To understand how the prediction mechanism differs for non-native speakers (L2) in CMC, this study assessed how the loss of multi-modal cues affects word predictability in turn-taking, considering language background and social factors. L2 watched videos, listened to audio, or read a transcript of conversations, and predicted the same set of omitted words with different levels of predictability and semantic relatedness in different CMC. Results showed that, similar to the L1 study in He et al. (2025), higher response similarity but longer response time were observed in conditions with richer cues in L2 predictive processing. Semantic relatedness, self-emotion, attention, and language proficiency did not affect predictability. Participants reporting negative emotions and more limited L2 exposure demonstrated reduced prediction accuracy, particularly in cue-rich environments. These findings expand our understanding of L2 predictive processing in CMC by highlighting how multimodal cue integration operates differently for L2. The results have implications for developing communication technologies and language pedagogies tailored to L2 across various mediated communication contexts.

Keywords: content prediction; turn-taking; computer-mediated communication; second language processing

Introduction

Disfluency is often observed in real-time interactions (Brennan & Williams, 1995; Shriberg, 1994; Smith & Clark, 1993) and can be attributed to multiple aspects: age, gender, speed, clarity, and individual differences in language experience, memory, and cognition among other factors (Bortfeld et al., 2001; Li et al., 1995). One prominent yet often overlooked source of the problem is content prediction in conversation. Prediction functions as a central mechanism of comprehension and production in language processing (Friston, 2010; see Ryskin & Nieuwland, 2023 for a review). These predictions rely on conversational cues, especially the preceding sentence context, and semantic information arisen from the context and the global constructive knowledge.

Effective communication depends on the predictability of conversational dynamics, enabling smooth turn-taking and facilitating the overall flow of communication (Torreira et al., 2015). Studies on face-to-face (F2F) communication have highlighted that speakers rely on a range of verbal and non-verbal cues (De Ruiter et al., 2012; Hadley & Culling, 2022; Gergle et al., 2013; Levinson & Torreira, 2015) to make real-time predictions about upcoming words (Corps et al., 2018, Froehlich et al., 2019; Levinson, 2016; Magyari & De Ruiter, 2012). However, computer-mediated conversation (CMC) introduces distinctive challenges stemming from the variations of cue availability and effectiveness across diverse communication channels, encompassing video, audio, and text (Kalman et al., 2006). Cue-impoverished CMC impairs the accuracy and rate of prediction process (He et al., 2025; Trujillo et al., 2021; Levinson, 2016) and frequently engenders various conversation disruptions (Duan et al., 2021; Lim et al. 2022; Walther, 2011).

Prior work has shown that predictive processing in second language speakers (L2) incurs higher cognitive cost than in first language speakers (L1). L2 predictions tend to occur later, requiring more deliberate cognitive effort and top-down processing strategies (Lev-Ari, 2015). They are more constrained by working memory (Hopp, 2014), rely more on L1-mediated access for word retrieval (Wolter, 2001) and surface-level cues, such as lexical co-occurrence patterns, rather on deeper syntactic or semantic regularities, further inhibiting their predictive efficiency (Hopp, 2015). The additional cognitive effort required for lexical retrieval, syntactic parsing, and semantic integration in L2 often results in greater communicative difficulty and pressure during real-time communication. The discrepancy becomes even more pronounced in CMC, where text-based interactions lack multimodal cues (e.g., prosody, gestures, facial expressions) that can aid prediction in F2F settings (Burgoon et al., 2021).

This study explored the prediction mechanism of L2 in line of He et al. (2025) to understand how language background further shapes the dynamic interplay between sequential predictability and cue informativity across CMC. We employed a word prediction task to examine response similarity, response time, and response rate in different CMC. Given that this is a comparison study conducted concurrently with He et al. (2025), our hypotheses were not built off their findings but from previous theories and empirical findings:

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

H1: the progressive loss of visual-prosodic information from multi-modal to unimodal CMC will create a corresponding gradient in response similarity, reflecting decreased access to articulatory and paralinguistic cues that support predictive processing (Corps et al., 2018; Levinson, 2016); Semantic coherence in predictions will remain relatively consistent across modalities, as conceptual relationships and constructional knowledge are primarily mediated through lexical-semantic networks rather than modality-specific cues (Broderick et al., 2018; Frank & Willems, 2017).

H2: prediction will be slower with increasing cue richness as multimodal signals enable earlier and more confident prediction commitment (Torreira et al., 2015).

H3: Individual differences in emotion and attention will modulate prediction, with more positive emotional states and higher attention levels enhancing cue utilization efficiency across all modalities and partially mitigate the temporal costs of cue deprivation in the text-based conditions.

H4: L2 will show greater cognitive effort in prediction during turn-taking than L1, due to increased lexical-semantic demands (Hopp, 2014; Lev-Ari, 2015). Language proficiency and exposure will serve as a compensatory mechanism, with speakers with more limited experience showing greater performance deterioration in cue-impoverished conditions due to reduced automaticity in prediction-triggered retrieval processes. Speakers of languages that are phonetically or grammatically closer to English are expected to show advantage of predictive processing from L1 knowledge transfer (Sabourin & Stowe, 2008; Sabourin et al., 2016).

Method

Participants

L2 English speakers without known visual or auditory impairments (n = 107; age = 18-54; M_{age} = 21; M = 33, F = 73, non-binary = 1) were recruited via SONA for an in-person or online study in exchange for extra course credit or gift card, excluding 16 online participants due to incompleteness.

Experiment materials

The same materials from He et al. (2025) were used, where two different pairs of volunteers discussed "What do you think of opening a new dairy bar somewhere on campus? Where would be a good choice? What are the pros and cons that you think of?". Both conversations were recorded in a testing room with minimal setups to avoid additional surrounding cues. Both recordings were edited into video, audio, and text-only (V, A, T) versions. The design is fully between subjects: participants were randomly assigned to any of the three CMC types of either conversation 1 or 2. This paradigm ensures that the observed results were not due to any specific aspects of the conversation recordings. Detailed study materials can be found on Open Science Framework (https://osf.io/q4bzr/?view_only=80676cc4547442b8bfc476f411697c8b).

Prediction Task Individual words were omitted in the middle or towards the end of the utterances in each conversation for participants to fill out. We assessed every word slot's property in 1) sequential predictability and 2) semantic relatedness. We defined sequential predictability as the likelihood for a word to occur given its context. It was computed as the word predictability using the-state-of-art language model – generative pre-training 2 (GPT-2; Radford et al., 2019; Szewczyk & Federmeier, 2021). We defined semantic relatedness as how closely related the words are in terms of the taxonomic and thematic relation in context. It was computed using GloVe word embeddings (Pennington et al., 2014), where every word in turn N is compared all the preceding words in the same turn N and all the words in turn $N – 1$ and N - 2. This score was adopted as a measure of the contextual fit of any individual word in the conversation (Luke & Christianson, 2018).

For all words in each conversation, both the predictability score and semantic score were divided into three levels (high, mid, and low), each with a 20% quantile. Paired *t*-tests ensured no multicollinearity across three levels. Six words were selected from each of the nine groups (high-high, high-mid, etc.), giving a total of 54 words omitted from each clip. Note that this nine-way grouping is for counterbalancing purpose only. In the data analysis, both scores were coded as continuous variables. The stimuli from these two conversation recordings were counterbalanced in their predictability scores and semantic scores throughout the task.

Sentence Recall Task (SRT) To quickly assess the English proficiency of participants, we adopted SRT from Culbertson et al. (2020), a novel global construct with an ecologically valid task using naturalistic audio clips that provides reliable measure of L2 proficiency. Concerning the study length, we picked 10 sentences with an equal ratio of male-female voices and with a gradual increase in complexity and length.

Post-test Survey This survey collected the summaries of the main points of the conversation as the attention measure with self-ratings of emotion towards the overall conversation (with a 10-point Likert scale for each emotion word from positive to negative categories). Volunteers and participants filled out the same survey for comparison purpose.

Language Background Questionnaire To understand the language background of L2 more specifically, we compiled an aggregated questionnaire with reference to Language Experience and Proficiency Questionnaire (Kaushanskaya et al., 2020), Language History Questionnaire 3.0 (Li et al., 2019), and Language and Social Background Questionnaire (Anderson et al., 2018). These questionnaires were selected given their ease of use, accessibility, and large sample size that provide a comprehensive measure of background information in language learning and use. Selected questions were slightly modified to gauge the past education and current experience in language usage of English and their native language(s), including onset of English learning/exposure, onset of active use in English, types of instruction language(s) at each education level, and any code-switching habits. We also tailored the questions about different situations and activities of daily language use habits to suit college-student life.

Procedure

Participants used personal computers (online) or researchers' computers (in-lab) to access the study link on SONA in full screen mode. For in-lab studies, participants were placed alone in a testing room. This study was hosted on Pavlovia and run online via Qualtrics survey and JsPsych.

Participants clicked on the screen to start the prediction task in JsPsych. They were informed that they need to type down the next upcoming word when a text box pops up immediately after the video or the audio pauses, or when they see an underline in the text. The instruction varied based on the condition they were assigned to. Participants completed a demonstration trial followed by the prediction task. They were then redirected to Qualtrics survey to fill out summaries, perception and experience, and emotion ratings of the conversation. Then they completed a sentence recall task, where they typed down 10 individually played sentences the best they could. Lastly in Qualtrics, they completed a language background and demographic survey.

Analysis

Performance in three CMC groups was compared against each other. We measured 1) *response similarity* – how related the response is to the actual word of the slot. It was computed as the cosine similarity between the embeddings of the actual word and the response; 2) *semantic relatedness* of each response – how related the response is to any possible words for the slot. It was computed as the weighted average of the embeddings of the first 500 possible words where the weight is the normalized sequential predictability; 3) *response time* – the time taken after the video/audio/text paused and participants typed down the response in the textbox and hit "continue"; 4) *response rate* – the time spent in typing per character in the final response to avoid over penalty of long responses. Considering the rapid growth and significant improvement of computational models, we updated the word sequential predictability with Llama 3.0 (Dubey et al., 2024) for more accurate estimates. Although it yielded a skewed distribution of word slots when categorized into three groups compared to the original distribution generated from GPT-2, they were more consistent with how actual responses were processed in the post-hoc models.

All measures of text responses only considered the first full word response, no matter how many words the rest of the response had or how accurate they were. This ensures fairness across trials and across participants per study instruction. Minor errors such as misspellings and typos were corrected by Speller from Autocorrect (Sondej, 2022) and manually revised to avoid over penalty.

Attention Level Participants summarized the pros and cons mentioned by each speaker in the conversation. We computed the cosine similarity between their responses and the conversation transcript using spaCy (Honnibal & Montani, 2017). The averaged similarity formed a reliable scale (Cronback's $\alpha = .77$) as the overall *attention level*.

Self-Emotion Participants also rated how they felt about the conversation on 10-point Likert scale with different emotion words. These words were reliably loaded onto two different factors: positive ($n = 5$; $\alpha = .89$) and negative self-emotions ($n = 3$; $\alpha = .60$). Each of the ratings were averaged as separate predictors, *positive self-emotion* and *negative self-emotion*.

Language Proficiency In SRT, recall accuracy was defined as the weighted word error rate, where the weight is the inverse of the sentence length in characters and the error score is the edit distance between the response and the answer. The higher the error score, the lower the recall accuracy, and the lower the language proficiency in English.

Language Type As part of the exploratory analysis, native languages (NL) were grouped into six types with a more even distribution: Chinese, Japanese and Korean, Germanic, Romance, Indo-Aryan, and Others (Afro-Asiatic, Austronesian, Niger-Congo, Slavic, Turkic).

Language Usage L2 rated their overall frequency in using English and their native language(s) under different situations. Five questions were highly reliable ($\alpha = .89$) and were averaged to generate an overall language usage score.

Results

We manually reviewed and removed trials with randomly typed words. Responses faster than 200ms and slower than 30000ms were removed for skipping behaviors and attention deviation. We analyzed the results using linear mixed-effect (LME) models in R version 3.4.0 (R Core Team, 2017), with packages LME4 version 1.1.19 (Bates et al., 2015) and lmerTest version 2.0.33 (Kuznetsova et al., 2017).

The *lmer* package was used to define the baseline LME model, including *participant, trial*, and *conversation type* as random effects, and *CMC*, *word slot property*, *positive self-emotion, negative self-emotion*, *attention level*, *type of native language*, *proficiency*, and *language usage* as fixed effects.

Using this LME model, we conducted model selection via Least Absolute Shrinkage and Selection Operator (LASSO) to reduce the total number of predictors. This method yields better theoretical interpretability with model fit given many insignificant and potentially correlated predictors from the LME analysis. All predictors were included as potential covariates for a full scope selection with the best fit. LASSO regression was deployed by *cv.glmnet* package version 1.6.1 (Friedman et al., 2010). With post-hoc models, we used the *emmeans* package (Searle et al., 1980) to gauge all the pairwise comparisons of *CMC*. The corrections of *p* values with three-way comparisons offer a better measure for multiway comparison across CMC types with smaller chances of false positives. All significance below 0.1 has been reported.

Response Similarity (RS)

As expected, participants' performance in RS (range = [-.261, 1.000], *M* = .585, *SD* = .283) was affected by word predictability: compared to words with high predictability and high semantic relatedness (the baseline), words with mid or low predictability showed significantly lower RS regardless of their semantic relatedness except for the high-mid group (β = -.007, *SE* = .019, *p* = .695).

Unexpectedly, CMC was removed in post-hoc model. It is worth noted that in the baseline model, participants had significantly worse RS in A compared to V (β = -.0189, *SE* = .00972, *p* = .055) as shown by the decrease in cosine similarity of the response. With a more accurate pairwise comparison in CMC, however, there was no significant change comparing A to V (β = -.0189, *SE* = .00972, *p* = .127), T to A (β = .0120, *SE* = .0107, *p* = .495), or T to V (β = -.00685, *SE* = .0107, *p* = .797).

As predicted, negative self-emotion had a significant main effect on RS: Participants predicted significantly worse (β = -.00757, *SE* = .00199, *p* < .001) if their negative self-emotion ratings were higher. However, neither did positive self-emotion (β = -.00238, *SE* = .00207, *p* = .253) or attention (β = -.00655, *SE* = .00290, *p* = .821) enter the final model.

As expected, language usage has a main effect on RS (β = .0387, *SE* = .0172, *p* = .0264). Only Indo-Aryan speakers had a slightly worse performance than other language speakers (β = -.0293, *SE* = .0144, *p* = .0444).

Response Semantic Relatedness (SR)

As predicted, SR (range = [-.330, 1.000], *M* = .369, *SD* = .335) was not affected by CMC type (A: β = .00459, *SE* =.00676, *p* = .499; T: β = .00529, *SE* = .00742, *p* = .478 compared to V**)**. Overall score of the words showed significant effect on SR: all groups showed different degrees of significant decrease comparing to the baseline high-high group. The lower the predictability of word slot and the lower the original level of SR of word slot, the lower the semantic relatedness of responses. Only negative self-emotion (β = -.00448, *SE* = .155, *p* = .005) showed a significant deteriorating effect on SR.

Response Time (RT)

RT (range = [207.6, 29989.7], M = 4327.9, SD = 3626.4) showed mixed patterns from our hypotheses. As expected, words with low predictability showed significantly longer RT (low-low: β = 416.927, *SE* = 210.846, *p* = .048; low-mid: β = 665.537, *SE* = 262.190, *p* = .011; low-high: β = 668.409, *SE* = 277.712, *p* = .016) compared to the high-high baseline. Participants did not show significant longer RT in A compared to V (β = 442.365, *SE* = 378.398, *p* = .245), but significantly shorter RT in T compared to V (β = -2002.934, *SE* = 422.316, *p* < .001), an opposite pattern as hypothesized. With a pairwise comparison in CMC type, there was no significant difference comparing A to V (β = 442, *SE* = 378, *p* = .048), but again significantly shorter RT in T compared

Table 1: Post-hoc model for RS. Bold lines indicate significance. Baseline: Slot property high-high.

Fixed effects	**Estimate**	***SE***	***p***
(Intercept)	**0.695**	**0.061**	**0.022**
Slot property: high-mid	-0.007	0.019	0.695
Slot property: high-low	**-0.077**	**0.023**	**0.001**
Slot property: mid-high	**-0.182**	**0.019**	**<.001**
Slot property: mid-mid	**-0.129**	**0.022**	**<.001**
Slot property: mid-low	**-0.089**	**0.020**	**<.001**
Slot property: low-high	**-0.134**	**0.025**	**<.001**
Slot property: low-mid	**-0.057**	**0.024**	**0.017**
Slot property: low-low	**-0.189**	**0.019**	**<.001**
Negative self-emotion	**-0.008**	**0.002**	**<.001**
Language proficiency	-0.001	0.000	0.146
NL: Japanese/Korean	0.010	0.011	0.383
NL: Germanic	0.020	0.014	0.174
NL: Indo-Aryan	**-0.029**	**0.014**	**0.044**
Language usage	**0.039**	**0.017**	**0.026**

Table 2: Post-hoc model for SR. Bold lines indicate significance. Baseline: Slot property high-high.

Fixed effects	**Estimate**	***SE***	***p***
(Intercept)	0.592	0.159	0.156
Slot property: high-mid	**-0.026**	**0.014**	**0.060**
Slot property: high-low	**-0.120**	**0.017**	**<.001**
Slot property: mid-high	**-0.176**	**0.014**	**<.001**
Slot property: mid-mid	**-0.175**	**0.016**	**<.001**
Slot property: mid-low	**-0.246**	**0.014**	**<.001**
Slot property: low-high	**-0.395**	**0.019**	**<.001**
Slot property: low-mid	**-0.352**	**0.018**	**<.001**
Slot property: low-low	**-0.345**	**0.014**	**<.001**
CMC type: audio	0.005	0.007	0.499
CMC type: text	0.005	0.007	0.478
Language proficiency	0.000	0.000	0.981
Positive self-emotion	0.000	0.001	0.885
Negative self-emotion	**-0.004**	**0.002**	**0.005**
Attention level	-0.013	0.020	0.515
NL: Chinese	-0.003	0.014	0.820
NL: Japanese/Korean	-0.009	0.015	0.554
NL: Others	-0.002	0.014	0.884
NL: Germanic	0.008	0.012	0.494
NL: Romance	-0.002	0.014	0.913
NL: Indo-Aryan	-0.021	0.017	0.213
Language usage	0.006	0.014	0.669

to V (β = -2003, *SE* = 422, $p < .0001$) and in T compared to A (β = -2445, *SE* = 419, $p < .0001$).

Participants with more negative self-ratings of emotion spent significantly longer in typing responses (β = 153.391, *SE* = 86.545, *p* = .080). Attention level (β = -922.966, *SE* = 1141.597, *p* = .421) did not show any significant main effect. Language usage had a marginal faciliatory effect of RT (β = -1336.176, *SE* = 783.604, *p* = .091), but not language proficiency (removed in post-hoc model). Speakers of other languages showed significantly longer RT (β = 1082.363, *SE* = 531.190, *p* = .044).

Response Rate (RR)

Given the varying lengths of multi-word response while we only considered the first full-word entry, we performed the same analysis with RR to avoid over-penalty of long responses in RT (range = [88.32, 21564.20], M = 1116.80, SD = 1271.94). First, word slot property showed a more random pattern: only the high-mid (β = 130.161, *SE* = 74.939, *p* = .083) and mid-mid (β = 224.877, *SE* = 85.879, *p* = .091) groups showed marginally significantly slower RR, while the mid-low (β = -151.817, *SE* = 78.948, *p* = .055) group showed significantly faster RR compared to baseline high-high slots.

The result for CMC type was similar: there was no significant difference between V and A (β = 107.661, *SE* = 108.031, *p* = .321), but significantly shorter RT in T compared to V (β = -753.896, *SE* = 115.248, $p < .001$), an opposite pattern as hypothesized. With a pairwise comparison in CMC, there was no significant difference between A and V (β = 108, *SE* = 108, *p* = .579), but significantly shorter RT in both T (β = -754, *SE* = 115, $p < .001$) and A (β = -862, *SE* = 114, $p < .001$) comparing to V.

Unlike in RT, neither did negative self-emotion (β = -7.177, *SE* = 24.515, *p* = .770) show any significant effect on RR, nor did language proficiency or usage enter the post-hoc model. However, Japanese, Korean (β = -302.474, *SE* = 137.737, *p* = .031) and Indo-Aryan speakers (β = -308.815, *SE* = 156.664, *p* = .052) showed significantly shorter RR.

Table 3: Post-hoc model for RT. Bold lines indicate significance. Baseline: Slot property high-high.

Fixed effects	**Estimate**	***SE***	***p***
(Intercept)	**5164.825**	**1104.748**	**<.001**
Slot property: high-mid	273.946	205.859	0.183
Slot property: high-low	-139.042	249.122	0.577
Slot property: mid-high	278.281	214.354	0.194
Slot property: mid-mid	149.069	243.160	0.540
Slot property: mid-low	281.202	218.386	0.198
Slot property: low-high	**668.409**	**277.712**	**0.016**
Slot property: low-mid	**665.537**	**262.190**	**0.011**
Slot property: low-low	**416.927**	**210.846**	**0.048**
CMC type: audio	442.365	378.398	0.245
CMC type: text	**-2002.934**	**422.316**	**<.001**
Language Proficiency	-6.391	18.557	0.731
Positive self-emotion	-36.681	82.386	0.657
Negative self-emotion	**153.391**	**86.545**	**0.080**
Attention level	-922.966	1141.597	0.421
NL: Chinese	608.818	402.827	0.134
NL: Japanese/Korean	-247.500	538.501	0.647
NL: Others	**1082.363**	**531.190**	**0.044**
NL: Germanic	-515.164	616.126	0.405
Language usage	**-1336.176**	**783.604**	**0.091**

Table 4: Post-hoc model for RR. Baseline: Slot property high-high.

Fixed effects	**Estimate**	***SE***	***p***
(Intercept)	**1773.170**	**285.175**	**<.001**
Slot property: high-mid	**130.161**	**74.939**	**0.083**
Slot property: high-low	-101.174	89.491	0.259
Slot property: mid-high	117.031	77.906	0.133
Slot property: mid-mid	**224.877**	**85.879**	**0.009**
Slot property: mid-low	**-151.817**	**78.948**	**0.055**
Slot property: low-high	26.959	97.371	0.782
Slot property: low-mid	43.232	91.935	0.638
Slot property: low-low	-47.691	75.541	0.528
CMC type: audio	107.661	108.031	0.321
CMC type: text	**-753.896**	**115.248**	**<.001**
Positive self-emotion	-23.880	22.855	0.299
Negative self-emotion	-7.177	24.515	0.770
Attention level	-350.274	304.040	0.252
NL: Japanese/Korean	**-302.474**	**137.737**	**0.031**
NL: Germanic	-189.998	158.037	0.232
NL: Romance	-127.104	125.192	0.313
NL: Indo-Aryan	**-308.815**	**156.664**	**0.052**

Discussions

In this study, we extended predictive processing in CMC to L2 populations. Our findings broadly support the hypothesis that predictive mechanisms are fundamentally altered in CMC compared to F2F interactions, with particularly pronounced effects for L2. We interpreted the results within contemporary frameworks of prediction and addressed the implications for both theoretical understanding and practical applications in communication technology.

Word Property Differences

Our results suggested that prediction was sensitive to both sequential predictability and semantic relatedness in the context. As predicted, the progressive loss of visual-prosodic information created a corresponding gradient in prediction similarity and efficiency, reflecting decreased access to articulatory and paralinguistic cues that support predictive processing. Compared to L1, L2 were more sensitive to the drop in word predictability and semantic relatedness. Unlike L1 in He et al. (2025), where the effect only surfaced when the change was salient enough, the deteriorating effect of lower sequential predictability on L2 was more pronounced in both RS and RT. The heavier reliance on sequential probability in L2 may have resulted in a lower degree of automaticity and less effective integration of predictive cues in relation to language experience, which necessitated more deliberate, delayed, and resource-intensive approaches to prediction in line with Hopp's claim (2015).

The results of RR were not as straightforward: it was significantly lower in the mid-low slots, but significantly longer in high-mid and mid-mid compared to high-high slots. Releveling of baseline to mid-levels revealed a lack of significance, while releveling to low-low restored part of the significance. This exploratory evaluation showed that the high-high and low-low word slots demonstrated a fundamental difference in predictive mechanism from their inherent distributional statistics. This unexpected observation may also be due to the update from GPT-2 to Llama 3.0 model where some words reassigned to the mid and high levels. Although predictability was still a continuous measure, slot properties was no longer evenly distributed. Future studies could reevaluate this finding by updating the chosen slots and imposing a stricter limit of the input range.

CMC Differences

Unexpectedly, L2 did not show difference of RS in word prediction in different CMC types. Unlike L1 who were sensitive to auditory cue loss (He et al., 2025), L2's performance was not critically different with the loss of visual or auditory cues. A potential floor effect of cue integration may be observed: although L2 were sensitive to word predictability to facilitate processing, prediction in any mediated channel was challenging enough such that L2 could not effectively employ visual and auditory cues in parallel with lexical cues to aid content prediction in cue-rich CMC.

Regarding RT and RR, the temporal cost of prediction increased when more multimodal cues were available. This consistent pattern in both L2 and L1 from He et al. (2025) showed that regardless of language experience, participants may have experienced heavier cognitive load in audio processing than reading. However, this facilitation effect may also arise from the nature of reading: our slow measure could not avoid text skimming behaviors or balance individual variations in reading speed or button pressing behaviors (Hopp, 2014; Mani & Huettig, 2014; Martin et al., 2013). Future studies should employ fast measures (e.g., sliding window paradigm) to provide a better measure of reading behaviors to validate the empirical findings in this study.

Language Background

More daily language usage showed a facilitatory effect for both content and temporal predictions, but not prediction rate. It aligns with usage-based models (e.g., Bybee, 2006; Tomasello, 2009) emphasizing experience-driven mechanism. The absence of proficiency effect is in line with some prior work showing a lack of correlation between L2 proficiency and prediction (Kaan & Grüter, 2021), but against the mainstream framework where L2 would experience heightened cognitive load during prediction processes moderated by proficiency level. The relatively homogenous language proficiency in our participants may have weakened the proficiency-related variance: they were late bilinguals who recently moved to the US for college education (a critical change in L2 exposure). Future studies could compare diverse proficiency levels (e.g., cumulative immersion vs. instructional) to generalize the findings.

Notably, a mixed group of speakers in Afro-Asiatic, Austronesian, Niger-Congo, Slavic, Turkic languages showed longer RT, while Japanese, Korean, and Indo-Aryan speakers showed shorter RR, implicating L1-L2 distance in CMC prediction efficiency. More controlled investigations of typological priming (Ito & Pickering, 2021; Ito et al., 2018; Schlenter, 2023) could disentangle L1 transfer effects.

Emotion and Perception

Consistent findings surfaced in RS, SR, and RT: negative self-emotion had a consistent deteriorating effect. Especially for RS, the potential auditory floor effect (A vs. T) showed that the availability of auditory cues could not compensate for the exacerbation from negative self-emotions.

Unlike L1 in He et al. (2025), L2 with more negative self-emotion showed significantly longer RT but not significantly shorter RR. This indicates that negative self-emotion leads to slower RT regardless of response length. However, attention did not show any effect, indicating that L2 participants might not be disadvantaged in attention deviation, but potentially processed more slowly to predict with better accuracy, typed responses at a consistent rate throughout, or struggled longer to maintain a good level of comprehension. Crucially, preserved RR amid slower RT suggests accuracy-preservation strategies–L2 prioritized response similarity over speed to deliberately mitigate potential prediction errors.

The lack of attention effect contradicts attentional control theory but may reflect task constraints: as passive predictors rather than active interlocutors, participants lacked the joint attentional scaffolds that modulate real conversation. Given that observer may have elicited different prediction mechanisms compared to interlocutor, future research could elucidate the attention-emotion interaction via EEG studies.

Implications

The findings underscore the critical roles of multimodal cues, linguistic exposure, and social factors in affecting predictive processing strategies in CMC. These findings necessitate refinements to adaptive predictive coding models particularly regarding L2 speakers' reliance on sequential statistics over conceptual networks. This study also elucidates key parameters in constructing theoretical backgrounds of L2 predictive processing models, informing tool design in different CMC contexts, highlighting critical metrics to sustain smooth prediction and integration, and calling for linguistically grounded scaffolding of real-time prediction.

Acknowledgements

We would like to thank our anonymous reviewers for their feedback and Morten H. Christiansen for the study conceptualization. We also would like to thank research assistants Harim Hahn and Jay Huang for their effort.

References

Anderson, J. A. E., Mak, L., Chahi, A. K., & Bialystok, E. (2018). *Language and Social Background Questionnaire--Revised (LSBQ)* [Database record]. APA PsycTests. https://doi.org/10.1037/t66278-000.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*. https://doi.org/10.18637/jss.v067.i01.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender. *Language and Speech, 44*(2), 123-147. https://doi.org/10.1177/00238309010440020101.

Brennan, S. E., & Williams, M. (1995). The feeling of another's knowing: Prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language, 34*(3), 383–398. https://doi.org/10.1006/jmla.1995.1017.

Broderick, M. P., Anderson, A. J., Di Liberto, G. M., Crosse, M. J., & Lalor, E. C. (2018). Electrophysiological Correlates of Semantic Dissimilarity Reflect the Comprehension of Natural, Narrative Speech. *Current biology: CB*, *28*(5), 803–809.e3. https://doi.org/10.1016/j.cub.2018.01.080

Burgoon, J.K., Manusov, V., & Guerrero, L.K. (2021). Nonverbal Communication (2nd ed.). Routledge. https://doi.org/10.4324/9781003095552.

Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, *82*, 711–733.

Corps, R. E., Crossley A., Gambi, C., & Pickering, M. J. (2018). Early preparation during turn-taking: Listeners use content predictions to determine what to say but not when to say it. *Cognition, 175*, 77–95. https://doi.org/10.1016/j.cognition.2018.01.015.

Culbertson, G., Andersen, E., & Christiansen, M. H. (2020). Using utterance recall to assess second language proficiency. *Language Learning, 70*(Suppl 2), 104–132. https://doi.org/10.1111/lang.12399.

De Ruiter, J. P., Bangerter, A., & Dings, P. (2012). The interplay between gesture and speech in the production of referring expressions: Investigating the tradeoff hypothesis. *Topics in cognitive science*, *4*(2), 232-248. https://doi.org/10.1111/j.1756-8765.2012.01183.x.

Duan, W., Yamashita, N., Shirai, Y., & Fussell, S. R. (2021). Bridging Fluency Disparity between Native and Nonnative Speakers in Multilingual Multiparty Collaboration Using a Clarification Agent. *Proceedings of the ACM on Human-Computer Interaction, 5(CSCW2)*, 1-31. https://doi.org/10.1145/3479579.

Dubey, A., Jauhri, A., Pandey, A., & et al. (2024). The Llama 3 Herd of Models. arXiv:2407.21783v2. Version 2. https://arxiv.org/abs/2407.21783.

Frank, S. L., & Willems, R. M. (2017). Word predictability and semantic similarity show distinct patterns of brain activity during language comprehension. *Language, Cognition and Neuroscience*, *32*(9), 1192–1203. https://doi-org/10.1080/23273798.2017.1323109.

Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*. https://doi.org/10.18637/jss.v033.i01.

Friston K. (2010). The free-energy principle: a unified brain theory?. *Nature reviews. Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Gergle, D., Kraut, R. E, & Fussell, S. R. (2013). Using Visual Information for Grounding and Awareness in Collaborative Tasks. *Human–Computer Interaction*, *28*(1), 1-39.

Hadley, L. V., & Culling, J. F. (2022). Timing of head turns to upcoming talkers in triadic conversation: Evidence for prediction of turn ends and interruptions. *Front Psychol, 13*, 1061582. https://doi.org/10.3389/fpsyg.2022.1061582.

Han, L., Kashyap, A. L., Finin, T., Mayfield, J., & Weese, J. (2013). *UMBC_EBIQUITY-CORE: Semantic textual similarity systems. Second Joint Conference on Lexical and Computational Semantics (*SEM), Volume 1: Proceedings of the Main Conference and the Shared Task: Semantic Textual Similarity*, *1*, 44–52

He, W., MacDonald, C. C., Yoo, Y., Eizayaga, M., Shim, R., Katreczko, L. D., & Fussell, S. R. (2025). Exploring Content Predictability in Turn-Taking Through Different Computer-Mediated Communications. In *Proceedings of the 31st International Conference on Computational Linguistics*, 7949–7962, Abu Dhabi, UAE. Association for Computational Linguistics.

Hinojosa, J. A, Moreno, E. M., & Ferré, P. (2019). Affective neurolinguistics: towards a framework for reconciling language and emotion. *Language, Cognition and Neuroscience, 35*, 813 - 839. https://doi.org/10.1080/23273798.2019.1620957.

Honnibal, M. (2017). spaCy 2: Natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. *(No Title)*.

Hopp, H. (2014). Working Memory Effects in the L2 Processing of Ambiguous Relative Clauses. *Language Acquisition, 21*, 250 - 278. https://doi.org/10.1080/10489223.2014.892943.

Hopp, H. (2015). Semantics and morphosyntax in predictive L2 sentence processing. *International Review of Applied Linguistics in Language Teaching, 53*(3), 277-306. https://doi.org/10.1515/iral-2015-0014.

Ito, A., & Pickering, M.J. (2021). Chapter 2. Automaticity and prediction in non-native language comprehension.

Ito, A., Pickering, M.J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual world eye-tracking study. *Journal of Memory and Language, 98*, 1-11.

Kaan, E., & Grüter, T. (2021). Prediction in second language processing and learning. *Prediction in second language processing and learning*, 1-24.

Kalman, Y. M., Ravid, G., Raban, D. R., & Rafaeli, S. (2006). Speak *now* or forever hold your peace: Power law chronemics of turn-taking and response in asynchronous CMC. *Vortrag auf der*, *56*.

Kaushanskaya, M., Blumenfeld, H. K., & Marian, V. (2020). The Language Experience and Proficiency Questionnaire (LEAP-Q): Ten years later. *Bilingualism (Cambridge, England)*, *23*(5), 945–950. https://doi.org/10.1017/s1366728919000038.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*. https://doi.org/10.18637/jss.v082.i13.

Lev-Ari, S. (2015). Comprehending non-native speakers: theory and evidence for adjustment in manner of processing. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.01546.

Levinson, S. C. (2016). Turn-taking in human communication–origins and implications for language processing. *Trends in cognitive sciences*, *20*(1), 6-14. https://doi.org/10.1016/j.tics.2015.10.010.

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in psychology*, *6*, 731. https://doi.org/10.3389/fpsyg.2015.00731.

Li, E. C., Williams, S. E., & Volpe, A. D. (1995). The effects of topic and listener familiarity on discourse variables in procedural and narrative discourse tasks. *Journal of communication disorders, 28*, 39-55. https://doi.org/10.1016/0021-9924(95)91023-Z .

Li, P., Zhang, F., Yu, A., & Zhao, X. (2020). Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*, *23*(5), 938–944. doi:10.1017/S1366728918001153

Lim, H., Cosley, D., & Fussell, S. R. (2022). Understanding Cross-lingual Pragmatic Misunderstandings in Email Communication. *Proceedings of the ACM on Human-Computer Interaction, 6(CSCW1),* 1-32. https://doi.org/10.1145/3512976

Luke, S. G., & Christianson, K. (2018). The Provo Corpus: A large eye-tracking corpus with predictability norms. *Behavior research methods*, *50*(2), 826-833. https://doi.org/10.3758/s13428-017-0908-4.

Mani, N., & Huettig, F. (2014). Word reading skill predicts anticipation of upcoming spoken language input: a study of children developing proficiency in reading. *Journal of experimental child psychology*, *126*, 264–279. https://doi.org/10.1016/j.jecp.2014.05.004.

Martin, C. D., Thierry, G., Kuipers, J. R., Boutonnet, B., Foucart, A., & Costa, A. (2013). Bilinguals reading in their second language do not predict upcoming words as native readers do. *Journal of memory and language*, *69*(4), 574-588. http://dx.doi.org/10.1016/j.jml.2013.08.001.

Pennington, J., Socher, R., & Manning, C. (2014). Glove: Global vectors for word representation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. Association for Computational Linguistics. https://doi.org/10.3115/v1/d14-1162.

R Core Team. (2017). R: A language and environment for statistical computing. *Foundation for Statistical Computing, Vienna, Austria.*

Radford, A., Wu., J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models Are Unsupervised Multitask Learners. *OpenAI Blog*. https://insightcivic.s3.us-east-1.amazonaws.com/language-models.pdf.

Ryskin, R., & Nieuwland, M. S. (2023). Prediction during language comprehension: what is next?. *Trends in Cognitive Sciences, 27*(11), 1032–1052. https://doi.org/10.1016/j.tics.2023.08.003.

Sabourin, L., & Stowe, L. A. (2008). Second language processing: When are first and second languages processed similarly?. *Second Language Research*, *24*(3), 397-430.

Sabourin, L., Stowe, L. A., & De Haan, G. J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research*, *22*(1), 1-29.

Schlenter, J. (2023). Prediction in bilingual sentence processing: How prediction differs in a later learned language from a first language. *Bilingualism: Language and Cognition*, *26*(2), 253–267. doi:10.1017/S1366728922000736

Searle, S. R., Speed, F. M., & Milliken, G. A. (1980). Population marginal means in the linear model: An alternative to least squares means. *The American Statistician, 34*, 216-221. https://doi.org/10.1080/00031305.1980.10483031

Shriberg, E. (1994). *Preliminaries to a theory of speech disfluencies.* PhD thesis, University of California, Berkeley.

Smith, V. L., & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and Language, 32*(1), 25–38. https://doi.org/10.1006/jmla.1993.1002.

Sondej, F. (2022). Filyp/autocorrect: Spelling Corrector in python. https://github.com/filyp/autocorrect

Szewczyk, J. M., & Federmeier, K. D. (2022). Context-based facilitation of semantic access follows both logarithmic and linear functions of stimulus probability. *Journal of memory and language*, *123*, 104311. https://doi.org/10.1016/j.jml.2021.104311

Tomasello, M. (2009). The usage based theory of language acquisition. In E. L. Bavin (Ed.), *The Cambridge handbook of child language*. Cambridge, UK: Cambridge University Press.

Torreira, F., Bögels, S., & Levinson, S. C. (2015). Breathing for answering: The time course of Response Planning in conversation. *Frontiers in Psychology, 6*, 284. https://doi.org/10.3389/fpsyg.2015.00284

Trujillo, J. P., Levinson, S. C., & Holler, J. (2021). Visual information in computer-mediated interaction matters: Investigating the association between the availability of gesture and turn transition timing in conversation. In *Human-Computer Interaction. Design and User Experience Case Studies: Thematic Area, HCI 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings, Part III 23* (pp. 643-657). Springer International Publishing. https://doi.org/10.1007/978-3-030-78468-3_44.

Verhees, M. W., Chwilla, D. J., Tromp, J., & Vissers, C. T. (2015). Contributions of emotional state and attention to the processing of syntactic agreement errors: evidence from P600. *Frontiers in psychology*, *6*, 388. https://doi.org/10.3389/fpsyg.2015.00388.

Walther, J. B. (2011). Theories of computer-mediated communication and interpersonal relations. *The handbook of interpersonal communication*, *4*, 443-479.

Walther, J. B., & Parks, M. R. (2002). Cues filtered out, cues filtered in. *Handbook of interpersonal communication*, *3*, 529.

Wolter, B. (2001). Comparing the L1 and L2 mental lexicon. *Studies in Second Language Acquisition, 23*, 41-69. https://doi.org/10.1017/S0272263101001024.