УДК 004.891:81’255:316.7
DOI https://doi.org/10.32782/2410-0927-2025-23-17

Iryna CHARIKOVA
PhD in Philology, Associate Professor at the English Philology Department, Lesya Ukrainka Volyn National University, 13 Voli Ave, Lutsk, Ukraine, 43025
ORCID: *0000-0002-5080-5216*
Scopus Author ID: *60096263000*

Olena HALAPCHUK-TARNAVSKA
PhD in Philology, Associate Professor at the English Philology Department, Lesya Ukrainka Volyn National University, 13 Voli Ave, Lutsk, Ukraine, 43025
ORCID: *0009-0006-0527-9567*

Oksana BIELYKH
PhD in Philology, Associate Professor at the German Philology Department, Lesya Ukrainka Volyn National University, 13 Voli Ave, Lutsk, Ukraine, 43025
ORCID: *0000-0002-1167-0224*

To cite this article: Charikova, I., Halapchuk-Tarnavska, O., Bielykh, O. (2025). Cultural Challenges in AI Interpreting Systems [Культурні виклики в усному перекладі із застосуванням штучного інтелекту]. *Current Issues of Foreign Philology,* 23, 125–132, doi: https://doi.org/10.32782/2410-0927-2025-23-17

CULTURAL CHALLENGES IN AI INTERPRETING

The article addresses the limitations of current AI Speech-To-Speech Interpreting (S2S) systems in capturing cultural and pragmatic nuances It highlights that AI interpreters often fail to recognize subtle emotional cues, social dynamics, and cultural context, leading to miscommunication, alienation, or a perception of insincerity. The methodology involves analyzing existing architectures and proposing an advanced Adaptive Cascade architecture that integrates modules like Pragmatic Correction (PCM) and Tone & Style Control (TCS). These modules are trained via reinforcement learning with human feedback to enable dynamic sociolinguistic adaptation, considering cultural dimensions such as Hofstede’s model and social parameters like Power Distance Index (PDI) and Individualism/Collectivism (IDV). The scientific novelty lies in the systematic incorporation of sociolinguistic metrics – such as pragmatic adequacy and social acceptability – beyond traditional lexical accuracy. The approach emphasizes modeling social parameters and developing datasets annotated with cultural and emotional metadata to improve the interpretive quality. Overall, achieving truly effective cross-cultural AI interpreting requires moving beyond static models and lexical metrics toward dynamic, context-aware, sociolinguistically informed systems. The paper also discusses the potential for these systems to enhance diplomatic communication and international collaboration by reducing misunderstandings. It advocates for interdisciplinary research combining linguistics, AI, and cultural studies to create more nuanced and ethically responsible interpreting tools. The authors suggest that future work should include real-world testing in diverse cultural settings to validate these models’ effectiveness. Such advancements could significantly improve global communication, fostering greater mutual understanding and respect among different cultures. Integrating cultural dimensions into AI Interpreting can help prevent cultural insensitivity and promote respectful intercultural exchanges.

Key words: *AI Interpreting, S2ST (Speech-to-Speech Translation), cultural dimensions, Hofstede’s Model, pragmatic correction, pragmatic equivalence, power distance index (PDIScore), social acceptability, prosodic adaptation, adaptive cascade.*

Ірина ЧАРІКОВА
кандидат філологічних наук, доцент кафедри англійської філології, Волинський національний університет імені Лесі Українки, просп. Волі, 13, м. Луцьк, Україна, 43025
ORCID: *0000-0002-5080-5216*
Scopus Author ID: *60096263000*

Олена ГАЛАПЧУК-ТАРНАВСЬКА
кандидат філологічних наук, доцент кафедри англійської філології, Волинський національний університет імені Лесі Українки, просп. Волі, 13, м. Луцьк, Україна, 43025
ORCID: *0009-0006-0527-9567*

Оксана БЄЛИХ
кандидат філологічних наук, доцент кафедри німецької філології, Волинський національний університет імені Лесі Українки, просп. Волі, 13, м. Луцьк, Україна, 43025
ORCID: *0000-0002-1167-0224*

Бібліографічний опис статті: Чарікова, І., Галапчук-Тарнавська, О., Бєлих, О. (2025). Культурні виклики в усному перекладі із застосуванням штучного інтелекту [Cultural Challenges in AI Interpreting Systems]. *Акутальні питання іноземної філології,* 23, 125–132, doi: https://doi.org/10.32782/2410-0927-2025-23-17

КУЛЬТУРНІ ВИКЛИКИ В УСНОМУ ПЕРЕКЛАДІ ІЗ ЗАСТОСУВАННЯМ ШТУЧНОГО ІНТЕЛЕКТУ

Основна мета статті – розглянути обмеження сучасних систем штучного інтелекту для усного перекладу (S2S) у відображенні культурних та прагматичних нюансів. У статті підкреслюється, що перекладачі на основі штучного інтелекту часто не в змозі розпізнати тонкі емоційні сигнали, соціальну динаміку та культурний контекст, що призводить до непорозумінь, відчуження або сприйняття нещирості. Методологія передбачає аналіз існуючих архітектур та пропозицію вдосконаленої адаптивної каскадної архітектури, яка інтегрує такі модулі, як прагматична корекція (PCM) та контроль тону і стилю (TCS). Ці модулі тренуються за допомогою підкріплювального навчання з людським зворотним зв'язком, щоб забезпечити динамічну соціолінгвістичну адаптацію з урахуванням культурних вимірів, таких як модель Хофстеде, та соціальних параметрів, таких як індекс дистанції влади (PDI) та індивідуалізм/колективізм (IDV). Наукова новизна полягає в систематичному включенні соціолінгвістичних показників, таких як прагматична адекватність та соціальна прийнятність, що виходять за межі традиційної лексичної точності. Такий підхід акцентує увагу на моделюванні соціальних параметрів та розробці наборів даних з анотаціями культурних та емоційних метаданих для поліпшення якості усного перекладу. Автори наголошують, що для досягнення справді ефективного міжкультурного перекладу за допомогою штучного інтелекту необхідно вийти за межі статичних моделей і лексичних показників і перейти до динамічних, контекстно-орієнтованих систем, що враховують соціолінгвістичні особливості. У статті також висвітлюється потенціал цих систем для поліпшення дипломатичної комунікації та міжнародної співпраці шляхом зменшення непорозумінь. Дослідники акцентують на необхідності проведення міждисциплінарних досліджень, що поєднують лінгвістику, штучний інтелект і культурологію, для створення більш коректних і етично відповідальних інструментів перекладу. Наголошується на важливості реальних тестів в різних культурних середовищах для перевірки ефективності пропонованих моделей. Це може значно поліпшити глобальну комунікацію, сприяючи більшому взаєморозумінню та повазі між різними культурами. Інтеграція культурних аспектів в усний переклад за допомогою штучного інтелекту може допомогти запобігти культурній нечутливості та сприяти міжкультурному обміну.

Ключові слова: *усний переклад штучним інтелектом, S2ST (Speech-to-Speech Translation), культурні виміри, модель Хофстеде, прагматична корекція, прагматична еквівалентність, індекс дистанції влади (PDIScore), соціальна прийнятність, просодична адаптація, адаптивний каскад.*

The relevance of the research. Cultural dimensions in AI interpreting are shaped by two interconnected global developments: the merging of technologies and the increasing cross-cultural exchanges. While current Speech-to-Speech Translation (S2ST) systems excel in lexical and syntactic accuracy, they encounter limitations in pragmatics and sociolinguistics, reaching a productivity plateau. Interpreting involves not only linguistic conversion but also cultural adaptation. AI interpreters struggle to incorporate culturally specific elements such as politeness, indirectness, and prosody, which are influenced by factors like the Power Distance Index (PDI) and Individualism/Collectivism (IDV), resulting in dysfunction at the linguistic-cultural interface. This issue is especially critical in sensitive fields like diplomacy, international negotiations, and healthcare (Chen M., Chen Z., Chen J., 2024, P. 150–165; Watanabe H., Shindo H., 2022, P. 1980–1984).

Most AI training datasets and models are developed within a WEIRD (Western, Educated, Industrialized, Rich, Democratic) framework, which risks unintentionally applying Western communication standards – such as directness and low reliance on context – to other cultures (Ponti E. M., Liska R., Waseem Z., 2023, P. 301–320). This highlights the need for research into methods for adapting ethical guidelines and ensuring that AI technologies are used fairly and reflectively across diverse cultural contexts globally. Additionally, misinterpretation in business settings can result in failed negotiations, damaged trust, and misunderstandings of partners' intentions. Therefore, it is crucial to create tools that can dynamically adjust the style and tone of interpretation in real time to mitigate these risks.

Recent scientific literature reveals a shift from emphasizing lexical accuracy, measured by metrics like BLEU, toward prioritizing the pragmatic and communicative effectiveness of AI systems. While research in Large Language Models (LLMs) and Neural Machine Translation (NMT), such as the work of A. Kripalani and W. Zhang, addresses biases related to gender and race in training data, it often overlooks systemic, culturally rooted biases that shape communicative etiquette – such as how refusals are expressed in high-context cultures (Graham Y. A, 2015, P. 280–288; Kripalani A., Zhang W., 2023, P. 4421–4435).

Notable advances have been made in modeling politeness, particularly in Korean and Japanese, with algorithms designed to determine social status from names and titles to select appropriate grammatical forms (Lee K., Nagao M., 2018, P. 857–883). However, these are typically monocultural and lack universal models grounded in broader cultural dimensions, such as PDI, for cross-cultural adaptation. Additionally, research by P. Rubin and S. Cappe highlights that prosody (intonation, tempo) significantly influences emotional perception, yet most Text-to-Speech (TTS) modules in S2ST systems employ culturally neutral emotion models that fail to account for cultural differences – such as a low tone indicating anger in Eastern cultures versus calmness in Western contexts (Rubin P., Cappe S., 2020, P. 1–11). This underscores the importance of developing a Cultural Emotion Index to enhance AI's emotional and cultural sensitivity.

H. G. Lee and D. Y. Kim proposed integrating frameworks similar to Hofstede's Model into technological design, primarily focusing on user interface (UX) development rather than real-time, dynamic interpreting algorithms (Lee H. G., Kim D. Y., 2000, pp. 459–475).

While current research addresses technical challenges such as politeness and prosody, it lacks a comprehensive architectural model that operationalizes cross-cultural dimensions like PDI, UAI, and IDV for pragmatic correction in Speech-to-Speech Translation (S2ST). This represents a significant gap that this study aims to address.

The primary **goal of the study** is to develop a theoretical and conceptual framework for the sociolinguistic adaptation of AI interpreting systems, ensuring pragmatic equivalence by embedding cultural dimensions into real-time interpreting processes.

The study's objectives include several **key tasks**:

– Identifying and operationalizing essential cultural dimensions such as Power Distance Index (PDI), Individualism (IDV), and Uncertainty Avoidance (UAI) as input parameters for the Pragmatic Correction Module (PCM) within the Speech-to-Speech Translation (S2ST) system;

– Developing a conceptual framework for automatically assessing social distance and selecting the appropriate register (politeness level) of speakers based on acoustic and textual metadata;

– Designing an Adaptive Cascade architecture for S2ST that incorporates the PCM and a Tone and Style Control Module (TCS) to enable real-time correction of pragmatic content and prosodic features.

– Arguing for a shift from conventional linguistic metrics toward sociolinguistic metrics, such as Pragmatic Adequacy and Social Acceptability, to better evaluate the performance and social appropriateness of AI interpreting systems.

Presentation of the main research material. Artificial intelligence has transitioned from an academic idea to a powerful socio-economic and political instrument with global reach. Nonetheless, alongside technological advancements, the issue of cultural uniformity in most AI developments has become more pressing. Current AI models, including transformers and large language models (LLMs), are predominantly shaped by the WEIRD (Western, Educated, Industrialized, Rich, Democratic) context, which creates a cognitive and cultural disconnect between the creators (developers) and the end-users worldwide (Lee H. G., Kim D. Y., 2000, P. 459–475; Rubin P., Cappe S., 2020, P. 1–11; Watanabe H., Shindo H., 2022, P. 1980–1984).

The globalization of communication imposes extraordinary demands on real-time speech interpreting technologies. Speech-to-Speech Translation (S2ST) systems, combining automatic speech recognition (ASR), neural machine translation (NMT), and speech synthesis (TTS), are increasingly seen as alternatives to traditional human interpreting methods, whether simultaneous or consecutive (Watanabe H., Shindo H., 2022, P. 1980–1984).

Despite notable improvements in lexical and syntactic precision, these systems often fail critically in pragmatics and sociolinguistics. Interpreting involves more than linguistic conversion; it is a form of cultural mediation that requires understanding the context, the speaker's intentions (illocutionary acts), and the social hierarchy of the participants.

It is essential to differentiate between Machine Translation (MT), which primarily handles written text and emphasizes static accuracy, and AI Interpreting (AI I), which processes dynamic, multimodal data such as voice, intonation, and pauses under real-time constraints. The current inability of AI to incorporate cultural norms results in dysfunction at the linguistic-cultural interface, where technically accurate interpretation may nonetheless lead to socially inappropriate communication (Gao S., Li Y., Liu Z., 2023, P. 7821–7835).

The foundational theories guiding this research are Speech Act Theory (J. Austin, J. Searle) and Sense Theory (Paris School), which emphasize that an interpreter's role is to transmit meaning and intent (illocutionary act), rather than merely translating words (Austin J. L., 1975; Halle E. T., 1976; Searle J. R., 1979; Seleskovitch D., 1984). The innovative aspect of this study is the development of a systematic framework that operationalizes cultural dimensions as parameters for algorithmically adjusting the input data in Speech-to-Speech Translation (S2ST) systems.

Within AI Interpreting, attaining pragmatic equivalence – where the interpreted message evokes the same intent or reaction as the original – is of utmost importance. For example, in cultures with a high Power Distance Index (PDI), a direct command like *Do this now* from a manager is seen as a definitive order. When interpreting this into a low PDI culture, it might be rendered as *You must do this immediately*, which could be interpreted as aggressive or disrespectful. An optimal AI interpreter for low PDI contexts might instead soften it to *It is highly recommended that this be prioritized*, maintaining the command's purpose while aligning with culturally appropriate language norms.

Hofstede's model serves as a valuable tool for anticipating communication preferences and styles across different cultures (Al-Rubaie A., Ahmad S., 2024; Hofstede G., 2001).

Based on Hofstede's framework, an effective AI interpreting system from English to Korean must recognize the social hierarchy between speakers. For instance, when a junior manager communicates with a senior colleague, the neural machine translation (NMT) model should select the most respectful form – Jondaenmal – for verbs and nouns, even if the original English sentence lacks formal markers. This is essential to prevent social offense and ensure culturally appropriate communication.

To thoroughly evaluate the shortcomings of current AI interpreting systems, it is important to categorize errors beyond conventional linguistic metrics such as Word Error Rate (WER) and BLEU scores (Ponti E. M., Liska R., Waseem Z., 2023, P. 301–320). We propose a typology of pragmatic failures arising from algorithms' cultural insensitivity. These failures stem from the lack of a Pragmatic Correction Module (PCM) within the standard Speech-to-Speech Translation (S2ST)

Table 1

Cultural Dimensions and Their Impact on AI Interpreting System Design

Hofstede's Dmension	Impact on S2ST System Design	Examples of Technical Requirements
Power Distance Index (PDI)	Identifies the requirement to represent social hierarchy and incorporate honorifics appropriately in communication	AI should adaptively choose the appropriate grammatical style (formal or informal) based on the context and the listener's social setting
Individualism vs. Collectivism (IDV)	Identifies whether the emphasis is on personal achievement or collective well-being, which influences pronoun selection and usage	AI should have the capacity to mitigate confrontational remarks in collectivist cultures or emphasize direct accountability in individualist cultures
Uncertainty Avoidance (UAI)	Establishes the need for clear, detailed, and transparent interpretation to ensure understanding and accuracy	AI must steer clear of slang and idiomatic expressions in high Uncertainty Avoidance (UAI) contexts and deliver interpretations that are thorough and free of ambiguity
Long-Term Orientation (LTO)	Influences how urgency and the clarity of plans are perceived.	In cultures characterized by high Long-Term Orientation (LTO), interpreting should be careful when using absolute terms such as ‘always’ or ‘never’ to avoid misrepresentation or misunderstanding

Compiled by the authors based on Hofstede's cultural dimensions model.

architecture, illustrating that technically accurate interpreting can still be socially inappropriate or diplomatically risky.

The following key scenarios exemplify violations of communicative adequacy across three primary domains: misrepresentation of politeness levels (PDI failure), improper pragmatic transformations (UAI failure), and insufficient prosodic expression (emotional failure). These issues are directly linked to the Power Distance Index (PDI), which measures how much a society accepts unequal power and status distribution (Table 2).

These instances illustrate that AI struggles to modify the way it expresses ideas while maintaining the original meaning, particularly when it must prevent ambiguity in communication (UAI) (Table 3).

These instances demonstrate that AI makes mistakes in prosodic synthesis by misunderstanding cultural standards for expressing emotions (Table 4).

Analysis of the above scenarios (Table 1, 2, 3) clearly shows that the linguistic and cultural interface dysfunction of AI interpreting systems is not the result of random errors, but is rooted in systematic methodological gaps. Neural models, trained primarily on text data and optimised for lexical accuracy, have proven blind to the multilayered nature of spoken language.

In fact, AI interpreters act like ‘cultural idiots’ – technically competent but socially

Table 2

Incorrect modelling of politeness (PDI failure)

Cultural Dimension	Scenario	AI Interpreting Failure (AI I)	Consequence
High PDI (Korea, Japan)	Business meeting: A subordinate Korean staff member (A) communicates with a senior supervisor (B), asking, Do you understand the new proposal?	The AI interprets the message into Korean for the senior manager using casual speech (Banmal) and unsuitable verb endings, which can be perceived as disrespectful or rude	Social Offence. The manager interprets this as an act of impoliteness and lack of respect from the subordinate. The communication fails because the AI overlooks the social hierarchy between A and B, which is reflected in the PDI score
Low PDI (Germany, Netherlands)	Overly formal interpreting: The AI renders a highly formal expression from a high Power Distance Index (PDI) culture, such as China, into German by directly maintaining the excessive politeness and honorifics, which may result in an unnatural or awkward tone	The German audience perceives the interpreting as excessively elaborate, insincere, or unnecessarily verbose, which can undermine trust and professionalism	Loss of trust. In cultures that prioritize straightforwardness, such behavior by AI is seen as unprofessional and lacking openness

Incorrect pragmatic transformation (UAI failure).

Table 3

Cultural Context Mismatch and AI Interpreting Failures

Cultural Dimension	Scenario	AI Interpreting Failure (AI I)	Consequence
High-context culture (China) vs. low-context culture (United States)	**Negotiating an agreement:** A Chinese negotiator (A) subtly conveys uncertainty by saying, *This might be challenging to carry out considering our current circumstances*	The AI interpreter provides a direct interpreting to the American counterpart: *We have a small challenge*, which may misrepresent the true intent or significance of the message.	**Misinformation.** In high-context cultures, this phrase functions as a definitive yet courteous rejection. Conversely, in low-context cultures, it is seen as a small hurdle that can be easily addressed. As a result, the American negotiators keep pushing, unaware that the agreement has already collapsed
High UAI (France) vs. Low UAI (United Kingdom)	**Question about the plan:** The AI renders the concise, straightforward question from the British individual, *What's the plan for next week?* into French, potentially without considering cultural nuances	The AI interprets directly without including the polite expressions or formal introductions necessary for French business interactions, which can lead to perceptions of rudeness or aggression	**Perception of aggression.** The French partner views the concise and straightforward question as too commanding, impatient, or hostile, because it breaches the customary standards of formal communication and politeness

Inadequate prosodic synthesis (emotional failure).

Table 4

Culture-Specific Emotional and Prosodic Failures in AI Interpreting

Cultural Dimension	Scenario	AI Interpreting Failure (AI I)	Consequence
Collectivism and Harmony Preservation (Vietnam)	**Expressing Anger:** A Vietnamese user conveys discontent through a calm, restrained tone combined with courteous language, while speaking slightly faster than usual.	A TTS system trained on Western datasets fails to detect these nuanced acoustic signals as indicators of anger, resulting in voice synthesis that sounds neutral and monotonic.	**Escalation of Conflict.** The service representative (listener) underestimates the severity of the issue and reacts negligently because the AI has misrepresented the level of customer dissatisfaction.
Masculinity (MAS) and Success (the USA)	**Motivational speech:** An American speaker employs a lively, loud, and rapid tone to convey confidence and authority.	AI renders this speech for a feminine culture such as Scandinavia by preserving an excessively loud and rapid delivery, which may not align with cultural norms.	**Alienation.** In cultures that emphasize modesty and harmony, such as feminine cultures, this AI tone is seen as invasive, confrontational, and lacking authenticity, which diminishes the effectiveness of the communication.

incompetent – because the current S2ST architecture ignores sociolinguistic and pragmatic context. The main methodological challenges that need to be overcome to create culturally sensitive AI Interpreting lie in the non-lexical components of speech, namely prosody and communicative intent (Lee K., Nagao M., 2018, P. 857–883).

The first, and critically important, methodological challenge concerns prosodic processing. Interpreting requires not only the transformation of words but also the adequate reproduction of the speaker's emotions, intentions, and social status. These elements are encoded in prosody – intonation, tempo, rhythm, and pauses. However, speech synthesizers (TTS) in AI Interpreting systems typically use universal, culturally neutral models of emotions, which leads to systematic mismatches (Rubin P., Cappe S., 2020, P. 1–11).

Another problem is the cultural variability of emotional expression. What is acoustically encoded as 'calm' in one culture (e.g., low tone) may be interpreted as 'disinterest' or even 'hidden aggression' in another. This requires the development of a Cultural Emotion Index and an adaptive TTS module. Prosody (intonation, tempo, pitch) is a key carrier of culturally coded information. The AI-TTS module should not just synthesise words, but reproduce a culturally appropriate tone (Kripalani A., Zhang W., 2023, P. 4421–4435).

For example, in Finnish culture (low UAI, individualism), long pauses in conversation are normal and acceptable, indicating deliberation. If an AI system interpreting for an American (low UAI, high individualism) listener does not reproduce this pause and synthesises the voice at a fast pace, this may lead to an underestimation of the Scandinavian speaker, who will appear ‘uncertain’ or ‘hasty’. A successful AI interpreting system must maintain the functional equivalence of pauses.

Another illustration is the emotional index, which differs across cultures, with emotions being conveyed in unique ways. In some high-context Eastern societies (collectivist), dissatisfaction may be expressed through subtle tonal shifts or overly polite, artificial language. If a TTS system trained on Western data interprets this as a neutral tone, it will fundamentally misunderstand the speaker’s true intent.

The methodological issue is the challenge of encoding social dynamics into the interpreting algorithm. In high-context cultures like China and Japan, refusals are seldom explicit; for example, the phrase *It might be somewhat difficult* functions as a definitive rejection.

AI mistakes occur when such expressions are interpreted literally for low-context audiences (e.g., Americans), resulting in *There is a minor challenge*, which can be misleading. The AI must incorporate a Pragmatic Transformation Module that rephrases this as *We cannot proceed with this option at this time*, while preserving the original intent.

Modeling social distance involves creating a Social Communication Graph that calculates D (A, B) – the social distance between speaker A and listener B – based on factors such as titles, age, and nationality. This D value then guides the NMT model in choosing appropriate politeness forms.

To ensure cultural appropriateness, S2ST systems should be designed according to the principles of the Adaptive Cascade Model, integrating correction modules that dynamically adapt interpreting.

The traditional cascade model, which involves Automated Speech Recognition (ASR) → Neural Machine Translation (NMT) → Text-to-Speech (TTS), is inadequate on its own. Instead, we propose a more comprehensive architecture that incorporates additional components to improve performance:

$$\text{S2STAdaptive} = \text{ASR} \rightarrow \text{NMT} \rightarrow \text{PCM} \rightarrow \text{TCS} \rightarrow \text{TTS}$$

The Pragmatic Correction Module (PCM) processes the output from the NMT system and applies culturally specific rules, derived from parameters such as PDI and IDV, to modify the text. These adjustments may include adding or removing respectful forms or altering the strength of categorical statements, reflecting the cultural traits of the respective language and society.

The Tone and Style Control Module (TCS) then receives the corrected text from the PCM and communicates with the TTS system to implement the required prosodic modifications, such as tempo, pitch, and emotional tone adjustments.

Integrating the PCM and TCS modules solves architectural limits but creates a deeper challenge: how to train these modules. Unlike traditional NMT, which uses objective metrics such as BLEU scores for accuracy, pragmatic and sociolinguistic corrections require other forms of validation, focusing on social and cultural appropriateness rather than lexical similarity.

Cultural adequacy is a subjective, context-sensitive quality. It requires judgments about how acceptable the interpreted content is within a social setting. To help PCM make immediate decisions – such as softening categorical statements, enhancing politeness, or modifying illocutionary acts – a mechanism must directly integrate human assessments of communicative effectiveness into training. Therefore, to optimize PCM performance, use Reinforcement Learning for Pragmatics (RLfP). RLfP lets the system learn from rewards for culturally appropriate behavior rather than only from lexical accuracy (Gao S., Li Y., Liu Z., 2023. P. 7821–7835).

Training the PCM requires a Reward Model that assesses not just the correctness of words but the social acceptability and effectiveness of the interpreting. This model must be trained on localized evaluations provided by native speakers of the target culture. Such an approach enables AI to transcend mere grammatical reproduction and develop the ability to emulate the cultural competence of an experienced human interpreter.

The effective deployment of the Pragmatic Correction Module (PCM) and its training through Reinforcement Learning for Pragmatics (RLfP) depend on the development of a new, high-quality class of training resources. Without such resources, even the most advanced architecture capable of sociolinguistic adaptation will remain theoretical if trained solely on outdated corpora designed only for lexical accuracy.

For an AI interpreter to transcend being merely a 'cultural idiot' and develop the capacity for culturally sensitive behavior, it must be trained on data that explicitly captures social and pragmatic contexts. This necessitates a fundamental shift in how S2ST corpora are collected and annotated – from basic transcription to multi-parameter modeling of social dynamics.

Corpora used for S2ST should be annotated beyond simple transcription, incorporating metadata that captures the social context, such as PDIScore (Power Distance Index), IDVScore (Individualism vs. Collectivism Index), Social Distance, and Emotional Index. This enables the AI model to learn how to infer these parameters from acoustic and lexical cues. For instance, PDIScore indicates how much less influential societal members – such as institutions and organizations – are in accepting and anticipating unequal power distribution. This index reflects societal attitudes toward inequality: high PDIScore values are typical in countries like Malaysia and Mexico, where society is hierarchical, and communication tends to be formal, emphasizing status and respect. AI interpreting in such contexts must employ the highest honorifics and formal language (e.g., Korean Jondaenmal). Conversely, countries like Austria and Denmark, with low PDIScore, favor equality, resulting in direct, informal communication that relies less on titles. AI interpreters should adopt a neutral, straightforward style, avoiding excessive formality or modesty, which could seem insincere.

These numerical indicators are used in our research to dynamically adjust interpreting style within the Pragmatic Correction Module (PCM). For example, when interpreting from a language with low IDVScore (high collectivism) to one with high IDVScore (individualism), the PCM should convert an indirect refusal into a direct, explicit statement, aligning with the target culture's preference for directness.

To assess the effectiveness of the Adaptive Cascade Model and ensure the proper functioning of the Pragmatic Correction Module (PCM), it is essential to move beyond traditional validation techniques. Metrics like BLEU (Bilingual Evaluation Understudy) or WER (Word Error Rate) are inadequate because they only measure lexical similarity to a reference text and fail to account for pragmatic failures or sociocultural nuances.

Therefore, the evaluation of AI interpreting should focus on communicative effectiveness, which requires the development of new, human-centred metrics (Cui L., Hu J., 2024, pp. 1200–1215). One such metric is Pragmatic Adequacy (PA), which assesses how accurately the interpreting maintains the illocutionary intent (the speaker's purpose or goal). For instance, it examines whether a polite question is interpreted as such, rather than as a directive or a mere statement, ensuring that the intended effect on the listener within the target culture is achieved.

Social Acceptability (SA): This metric is more comprehensive and evaluates how suitable the generated voice and speech style are within a specific social environment. The SA assessment involves examining prosodic features such as pace, intonation, and emotional tone, as well as register, which pertains to the level of formality and respectfulness. Achieving a high SA score indicates that the AI respects cultural norms and avoids causing social discomfort or breaches of etiquette.

To enhance the PCM module continuously and address hidden cultural biases, implementing a Cultural Red Teaming process is essential. This involves engaging cross-cultural teams Cultural Red Teams) – native-speaking experts who deliberately test the AI by attempting to induce etiquette violations, use taboo language, or adopt inappropriate tones in sensitive situations (e.g., conveying sympathy, criticism, or humour). This empirical approach helps identify flaws in pragmatic correction algorithms, thereby strengthening the system`s ability to handle cultural differences effectively.

Conclusions and prospects for further research. This research illustrates that the effectiveness and societal acceptance of AI interpreting systems hinge more on their sociolinguistic flexibility than on lexical accuracy. Specifically, cultural dimensions such as the Power Distance Index (PDI) and Individualism (IDV) play a crucial role by providing guidelines for pragmatic adjustments and prosodic modifications during interpretation. Incorporating Hofstede's cultural framework alongside relevant sociolinguistic metrics (like PDIScore and IDVScore) into the speech-to-speech interpreting architecture is essential to address and resolve issues at the linguistic-cultural interface. Without a Pragmatic Correction Module (PCM), an AI interpreter would be merely a 'cultural novice' – technically proficient but lacking the capacity for meaningful cross-cultural communication.

Although a conceptual framework has been proposed, it is important to acknowledge its key limitations, which pave the way for future research:

– Addressing the static nature of current cultural models, such as Hofstede's, which do not account for intracultural variations (regional differences, subcultures) or the rapid evolution of communication norms. Developing dynamic sociolinguistic models capable of learning social status shifts during conversations – such as automatically recognizing transitions from formal to informal speech – is necessary;

– Empirical validation of communicative effectiveness, which involves systematically comparing AI-mediated negotiations with those conducted by human interpreters. Such studies should evaluate not only lexical accuracy but also factors like participant satisfaction, trust, and the achievement of negotiation goals (e.g., successful agreement) across different cultural contexts;

– Expanding the resource base by creating an open, multi-parameter audio corpus for training PCM and reinforcement learning frameworks, especially for low-resource languages. These datasets should include not only transcriptions but also cultural annotations (such as PDI and IDV scores) and honorific systems, ensuring an ethical and inclusive future for AI interpreting.

REFERENCES:

1. Al-Rubaie A., Ahmad S. The Impact of Hofstede's Cultural Dimensions on UX/UI Design for Global AI Applications: A Review of Emerging Trends. International Journal of Human-Computer Studies. 2024. Vol. 181. Article 103045.
2. Austin J. L. How to Do Things with Words. 2nd ed. Harvard University Press, 1975. 168 p.
3. Chen M., Chen Z., Chen J. Scaling Speech-to-Speech Translation to Large Language Models: A Unified Framework for Multimodal Pragmatic Adaptation. Transactions on Audio, Speech, and Language Processing. 2024. Vol. 33. P. 150–165.
4. Cui L., Hu J. Reinforcement Learning for Dialogue Management: A Reward Model Based on User Satisfaction and Trust. IEEE Transactions on Audio, Speech, and Language Processing. 2024. Vol. 32. P. 1200–1215.
5. Gao S., Li Y., Liu Z. Addressing Social Bias and Stereotypes in Neural Machine Translation via Reinforcement Learning with Human Feedback (RLHF) for Politeness. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL). Toronto: ACL, 2023. P. 7821–7835.
6. Graham Y. A Cursory Analysis of the Curation of the BLEU Metric. Proceedings of the 10th Workshop on Statistical Machine Translation (WMT). Lisbon: ACL, 2015. P. 280–288.
7. Hall E. T. Beyond Culture. Anchor Books, 1976. 298 p.
8. Hofstede G. Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations Across Nations. 2nd ed. SAGE Publications, 2001. 616 p.
9. Kripalani A., Zhang W. Mitigating Sociocultural Bias in Large Language Models through Cross-Cultural Prompting. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL). Toronto: ACL, 2023. P. 4421–4435.
10. Lee H. G., Kim D. Y. A model for the cultural adaptation of interface design: The case of the Korean Web. Interacting with Computers. 2000. Vol. 12, no. 5. P. 459–475.
11. Lee K., Nagao M. Modeling Politeness and Social Status in Neural Machine Translation: A Korean-Japanese Case Study. Computational Linguistics. 2018. Vol. 44, no. 4. P. 857–883.
12. Ponti E. M., Liska R., Waseem Z. On the Inadequacy of Current Evaluation Metrics for Cross-Lingual Pragmatic Tasks. Computational Linguistics. 2023. Vol. 49, no. 2. P. 301–320.
13. Rubin P., Cappe S. Prosody and Perception: Cultural Variation in the Expression and Interpretation of Emotion in Speech Synthesis. Speech Communication. 2020. Vol. 125. P. 1–11.
14. Seleskovitch D. Interpréter pour traduire. Didier Érudition, 1984. 308 p.
15. Searle J. R. Expression and Meaning: Studies in the Theory of Speech Acts. Cambridge University Press, 1979. 187 p.
16. Watanabe H., Shindo H. Cross-Cultural Communication in S2ST: Bridging the Gap between Prosody and Illocutionary Act. Interspeech 2022, Proceedings of the 23rd Annual Conference of the International Speech Communication Association. Interspeech, 2022. P. 1980–1984.

Дата першого надходження рукопису до видання: 14.11.2025
Дата прийнятого до друку рукопису після рецензування: 10.12.2025
Дата публікації: 30.12.2025