

Lost in Translation:

The impact of colloquial Spanish on AI-mediated triage
in the emergency department

Bryce A. Kennedy MSN RN SCRNP
Nebraska Methodist College

Any Chavez-Petit BSN RN
CHI Health Midlands



**NEBRASKA
METHODIST
COLLEGE**
The Josie Harper Campus

Background and Significance

- Effective communication in the emergency department (ED) is essential for
 - Accurate triage
 - Timely diagnosis
 - Initiating treatment
- Millions of Spanish-speaking people seek ED care in the US each year
 - Language barriers lead to
 - Delays in care
 - Adverse patient outcomes

Background and Significance

- Trained medical interpreters are the gold standard
 - Urgency and lack of availability mean
 - ED clinicians sometimes use AI for ad-hoc translation
- AI platforms perform well in
 - Structured, low-acuity contexts
- Unknown how AI performs with
 - Colloquial speech common in high-risk ED encounters
- Clinically significant gap
 - Patients describe symptoms with colloquial expressions that could
 - Challenge AI platforms
 - Increase risk of translation errors

Purpose and Aims

- Purpose

- Evaluate translation accuracy of four AI platforms during
 - Simulated ED triage under
 - Two linguistic conditions
 - Neutral (standard Spanish)
 - Colloquial (regionally-influenced Spanish)

- Aims

- Compare error frequency across AI platforms between conditions
- Categorize error types
- Test whether colloquial language increases translation errors

Design, Setting, and Sample

- Design
 - Controlled, comparative simulation
- Setting
 - Simulated ED triage
- Sample
 - Single ED triage vignette
 - English-speaking nurse
 - Spanish-speaking patient

Methods Flowchart

Step 1: Vignette

ED Triage Scenario:

English-Speaking Nurse
Spanish-Speaking Patient

Two Conditions:

Neutral
Colloquial

Step 2: Input Text

4 AI Platforms:

Gemini | ChatGPT |
Google Translate | DeepL

Serial Translation

Step 3: Collect Outputs

Scoring:

Fluency | Omit/Add |
Distortion | Critical Error

Independent Scoring

Step 4: Statistical Analysis

Methods: Vignette and Conditions

- ED triage
 - English-speaking nurse: 14 questions.
 - Spanish-speaking patient: 14 responses.
 - 26 yo F, ruptured ectopic pregnancy.

Steps:

1. Wrote whole vignette in English
2. Translated patient responses to Spanish
3. Made two conditions
 - a. Neutral: standard Spanish with accepted medical terminology
 - i. 285 Spanish words
 - b. Colloquial: non-standard, regionally-influenced Spanish
 - i. 288 Spanish words
 - ii. Caribbean
 - iii. Mexico
 - iv. Central America
 - v. South America

Methods: Vignette Construction

Patient Response (English)	Human Neutral Translation	Human Colloquial Translation
Sometimes I feel like I might be nauseous. But that might just be the pain in my stomach. I'm not sure. But I haven't vomited. And I have constipation normally.	A veces siento que podría tener náuseas. Pero eso podría ser solo el dolor en el estómago. No estoy segura. Pero no he vomitado. Y normalmente tengo estreñimiento.	A ratos siento que me dan ganas de devolver . Pero a lo mejor solo es el dolor de tripa . No estoy segura. Igual no devolví . Y normalmente me cuesta ir de cuerpo .
This morning, I felt dizzy when I got out of bed. But I didn't faint. I do get dizzy sometimes because I apparently have anemia. But this morning was worse than normal. That's actually the reason my husband convinced me to come in.	Esta mañana me mareé al levantarme de la cama. Pero no me desmayé. A veces me mareo porque aparentemente tengo anemia. Pero esta mañana ha sido peor de lo normal. De hecho, esa es la razón por la que mi esposo me ha convencido para que viniera.	Hoy por la mañana me dio vuelta el mundo al pararme de la cama. Pero no me fui a negro . A veces se me va el mundo porque evidentemente me falta hierro . Pero esta mañana fue peor que otras veces. Justo por eso mi esposo me insistió que viniera .

Methods: AI Platforms

Gemini Flash 2.5 (gemini.google.com)	Low-latency, native multimodal LLM
Chat GPT-5.1 (chatgpt.com)	Multimodal generative pre-trained transformer LLM
Google Translate (translate.google.com)	Neural machine translation platform
DeepL Pro (deepl.com) Latin American	Convolutional neural network platform

- Serial translation
 - Question (English-Spanish)
 - Response (Spanish-English)
- LLMs required specific instructions

Methods: Output Scoring

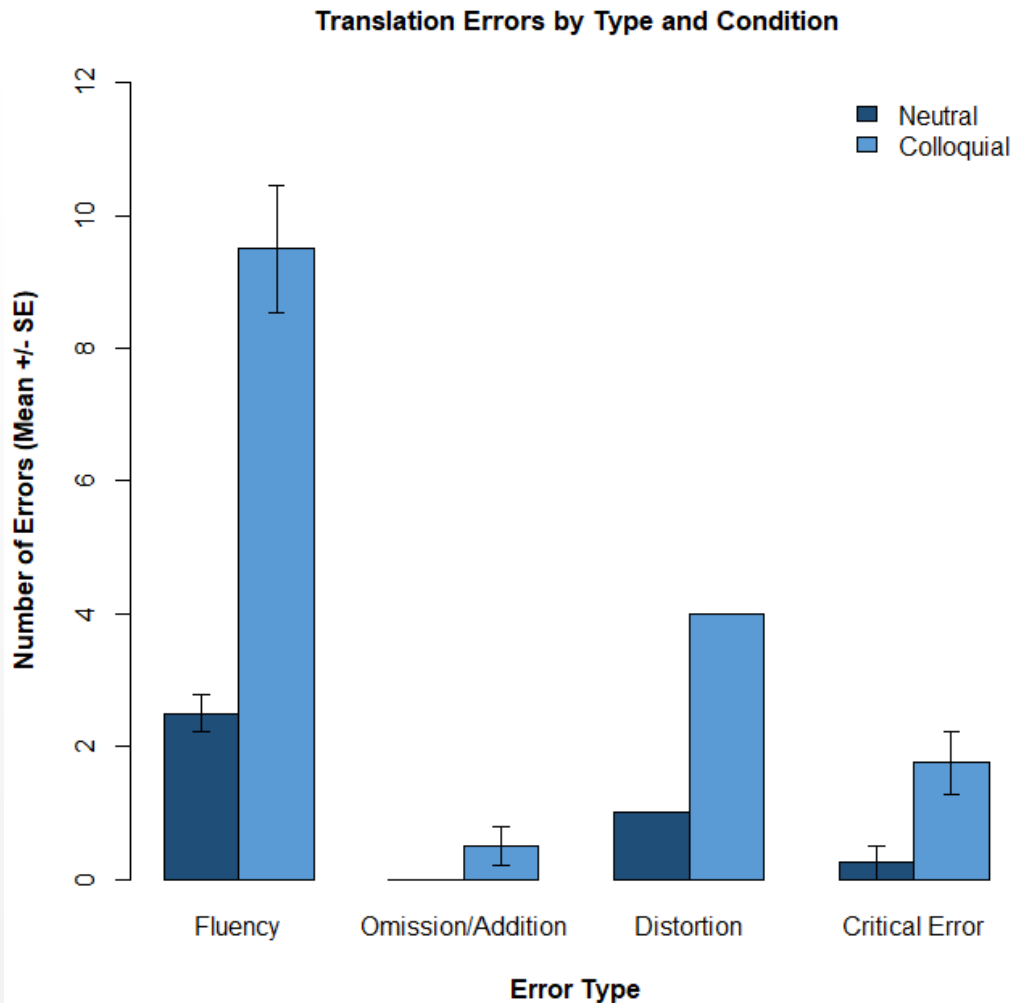
4 Error Categories:

1. Fluency: how natural the produced language sounds
2. Omission and addition: information left out or added
3. Distortion: incorrect or ambiguous meaning
4. Critical error: distortion that changes clinical meaning

- Errors counted only once
- Independent scoring
 - Discrepancies resolved via consensus
- No back-translation

Results

- 78 total errors
 - Most (65/78, 83%) were in Spanish-English translation
- Mean all-type error per platform
 - Neutral
 - Mean = 3.75 (SD = 0.96)
 - Colloquial
 - Mean = 15.75 (SD = 2.22)
 - $t(3) = 11.11$, $p = 0.002$, 95% CI [8.56, 15.44]
- Translation errors by type and condition
 - Colloquial produced more errors of each error type



Results: Neutral Condition Error Examples

Error Type	Input	Output
Fluency	Aparte de la anemia, que yo sepa, no.	As far as I know, apart from anemia, no. (DeepL)
Distortion	Are you having any unusual vaginal bleeding or discharge?	¿Está teniendo algún sangrado vaginal inusual o flujo ? (ChatGPT)
Critical Error	Are you having any urinary symptoms? For example are you going more than normal, less than normal, hurts to go, blood in the urine, urgency?	¿Tiene algún síntoma urinario? Por ejemplo, ¿ defeca con más frecuencia o menos frecuencia, le duele orinar, tiene sangre en la orina o urgencia? (Google Translate)

Results: Colloquial Condition Error Examples

Error Type	Input	Output
Fluency	Era a veces sí, a veces no, pero ahora llevo dos o tres días que no me da tregua.	It used to be sometimes yes, sometimes no, but now I've had two or three days that it doesn't give me a break. (Gemini)
Distortion	Me duele la guata hace como dos semanas. Lo que más me molesta es acá abajo, a la derecha. Además me tira un poquito la panza y el hombro derecho. Tengo las mamas molestas, pero chance es por amamantar.	My stomach has been hurting for about two weeks. What bothers me most is down here, on the right side. My belly and right shoulder are also pulling a little . My breasts are sore, but that's probably from breastfeeding. (Google Translate)
Omission/Addition	Sí. Estoy juntada y tengo vida sexual.	Yes. I'm living with my partner and I have a sex life. (Gemini)
Critical Error	Sip. Ando marcando hace unos días. Pero no es tan raro porque tengo un dispositivo.	Yep. I've been dialing for a few days. But it's not that unusual because I have a device . (DeepL)

Implications

- AI platforms performed
 - Well with neutral, standardized Spanish
 - Poorly with colloquial Spanish
- AI platforms could compromise ED triage accuracy and patient safety
- ED staff should exercise extreme caution with AI platforms during patient interactions

Conclusions

- Limitations

- External validity
 - Single high-acuity clinical vignette
 - Pan-colloquial Spanish blend
 - Text-only translation instead of speech-to-text
- Reliability
 - Bias in coding and error evaluation (specifically “Fluency”)
 - Scorers not blinded to condition
 - Authors not credentialed in medical interpretation or linguistics
 - Rely on self-reported linguistic proficiency

- Future Directions

- Expanded patient scenarios
- Additional AI platforms
- Real-world ED interactions
- Low-resource languages

Acknowledgement

Special thanks to Prof. Julieta Michelutti (Spanish language and literature, Córdoba, Argentina) for reviewing and revising our English-Spanish translation, and for providing suggestions for formal and colloquial words and phrases.

References

Ali, A. F., Alhebshi, S. H. S., & Alaqad, M. H. (2025). Evaluation of AI medical interpretation. *LinguAmerta*, 2(1), 1-8.

<https://jurnal.pnj.ac.id/index.php/linguamerta/article/download/7821/3657>

Brandenberger, J., Stedman, I., Stancati, N., Sappleton, K., Kanathasan, S., Fayyaz, J., & Singh, D. (2025). Using artificial intelligence based language interpretation in non-urgent paediatric emergency consultations: A clinical performance test and legal evaluation. *BMC Health Services Research*, 25(1), 138. <https://doi.org/10.1186/s12913-025-12263-1>

Cela Gutiérrez, C. (2025). AI-assisted interpreting: Valuable tool for professional interpreters or job displacement?. *Traduction et Langues*, 24(1), 263-281. <https://accedacris.ulpgc.es/bitstream/10553/143146/1/Ai%20assisted%20Interpreting.pdf>

Elderkin-Thompson, V., Silver, R. C., & Waitzkin, H. (2001). When nurses double as interpreters: A study of Spanish-speaking patients in a US primary care setting. *Social Science & Medicine*, 52(9), 1343-1358. [https://doi.org/10.1016/S0277-9536\(00\)00234-3](https://doi.org/10.1016/S0277-9536(00)00234-3)

Flores, G., Abreu, M., Barone, C. P., Bachur, R., & Lin, H. (2012). Errors of medical interpretation and their potential clinical consequences: A comparison of professional versus ad hoc versus no interpreters. *Annals of Emergency Medicine*, 60(5), 545-553. <https://doi.org/10.1016/j.annemergmed.2012.01.025>

Heath, M., Hvass, A. M. F., & Wejse, C. M. (2023). Interpreter services and effect on healthcare-a systematic review of the impact of different types of interpreters on patient outcome. *Journal of Migration and Health*, 7, 100162. <https://doi.org/10.1016/j.jmh.2023.100162>