

Evaluating the Effectiveness and Accuracy of Instant Translator Technologies versus Human Translators

Hung Quach, Alex Gordon PI: Dr. Shantanu Lal
Columbia University College of Dental Medicine



INTRODUCTION

In 2019, it was recorded that 70.9 million individuals, over the age of 5 years old, were Non-English speakers (US Census, 2019). One study in particular found that patients with limited English proficiency were more likely to experience adverse events in the healthcare setting due to communication errors in comparison to English speaking patients (Divi et al., 2007). Thus, as the population of Non-English speakers rises, language barriers will continue to contribute to the healthcare disparities across the US. Current solutions to these disparities includes third-party interpreter services or help from a bilingual non-healthcare individual. This study evaluates the effectiveness of instant translator technologies, ("Pocketalk" device) in communicating important dental healthcare information in contrast to third-party interpreters. These technologies may prove to be an important step in alleviating the care gaps that Non-English speaking patients experience in the dental healthcare setting.

DENTAL PHRASE EXAMPLES

Oral Health Instruction (Source: ADA and CDC):

1. Brush teeth twice a day with fluoride toothpaste
2. To brush properly angle the toothbrush the gum at a 45-degree angle and perform gentle circular stroke

Dental/Medical History Questions:

1. When was the last time you've been to the dentist
2. Are you taking or have recently taken any prescription or over the counter medicine(s)

Post-Operative Instructions:

1. Blood clots in the tooth socket is part of the normal healing process, do not disturb it
2. After a tooth is removed, you may have some discomfort and notice some swelling. This normal.

GRADING SCALE

ADEQUACY SCALE

- 5 = 100% of information conveyed from the original
- 4 = 75% of information conveyed from the original
- 3 = 50% of information conveyed from the original
- 2 = 25% of information conveyed from the original
- 1 = 0% of information conveyed from the original

MEANING SCALE

- 5 = Same meaning as original
- 4 = Almost the same meaning as the original
- 3 = Partially the same meaning as the original
- 2 = Misleading information added/omitted compared to the original
- 1 = Totally different meaning from the original

SEVERITY SCALE

- N/A = Sentence basically accurate
- 5 = Error, no effect on patient care
- 4 = Error, unclear effect on patient care
- 3 = Error, delays necessary care
- 2 = Error, impairs care in some way
- 1 = Error, dangerous to patient

MATERIALS & METHODS

- The efficiency of translation assessment was measured using a pre-developed scale by Khanna et al, which analyzed the translations base on their adequacy, meaning, severity (0-5 scale)
- Categories of documents translated: home care instructions, medical history questions, and post-operative instructions.
- Each document was translated into Spanish and Russian by both Pocketalk and human translator, with text transcripts produced
- The blinded researchers then rated each phrase in the translation transcript using the pre-developed scale (71 total)
- The adequacy, meaning, and efficiency of translation were assessed with F-Test, the degree of variance between the two groups determined, and the type of T-Test to use determined (Non-Equal Variance vs. Equal Variance)
- T-Test used to determine the significance of the results with an alpha of 0.05
- Results compared based on translation method, category of document, and original language of translation
- Results also correlated with word count and frequency of different error types

RESULTS

- Differences in Adequacy, Meaning, and Severity scores between instant and human translation were not statistically significant when translating from Russian to English (t(140) = 0.72, p = 0.47, t(135) = -0.21, p = 0.83, t(140) = 0, p = 0.1 respectively).
- Adequacy, Meaning, and Severity scores were higher (at a statistically significant level) for the instant translator than the human translator when translating from Spanish to English (t(127) = -4.71, p < 0.01, t(130) = -2.36, p = 0.02, t(110) = -2.81, p = 0.01 respectively).
- When comparing instant translator performance between languages, differences in Meaning and Severity were not statistically significant while the Adequacy score was significantly higher for Spanish translation than Russian translation (t(140) = -2.45, p = 0.02).
- When comparing human translator performance between languages, differences in Adequacy, Meaning, and Severity were not statistically significant between Russian and Spanish translations.
- For the Russian translation, the most common instant translator error types were substitutions (17) and omissions (11) while the most common human translator error types were substitutions (13) and additions (13)
- For the Spanish translation, the most common instant translator error types were substitutions (11) and omissions (3) while the most common human translator error types were additions (18) and omissions (12).
- Weakly negative correlations (R² value of 0.09, at greatest) were found between word count and adequacy/meaning scores for both languages and translator types

MEAN SCORE & STD OF RUSSIAN AND SPANISH TRANSLATION

Russian Translation

	Adequacy	Meaning	Severity
Instant	4.53 (0.53)	4.58 (0.71)	4.66 (0.61)
Human	4.46 (0.63)	4.61 (0.87)	4.66 (0.72)

Spanish Translation

	Adequacy	Meaning	Severity
Instant	4.75 (0.50)	4.73 (0.76)	4.82 (0.52)
Human	4.27 (0.70)	4.38 (1.00)	4.46 (0.92)

POCKETALK DEVICE



References

- Divi C, Koss RG, Schmalz SP, Loeb JM. Language proficiency and adverse events in US hospitals: a pilot study. *Int J Qual Health Care*. 2007 Apr;19(2):60-7. doi: 10.1093/intqhc/mzl069. Epub 2007 Feb 2. PMID: 17277013
- Ortman, Jennifer M., and Hyon B. Shin. "Language Projections: 2010 to 2020." American Sociological Association Annual Meeting. 2011
- U.S. Census Bureau. 2019. "Language Spoken at Home. American Community Survey." Retrieved from <https://data.census.gov/cedsci/table?q=United%20States%20language&tid=A CSST1Y2019.S1601&hidePreview=false>
- Khanna RR, Karlner LS, Eck M, Vittinghoff E, Koenig CI, Fang MC, Accuracy of GoogleTranslate™. *J. Hosp. Med* 2011;9:519-525. doi:10.1002/jhm.898
- American Dental Association. (2012). "Health History Form – English." Retrieved from <https://ebusiness.ada.org/Assets/Docs/S117.pdf>
- American Dental Association. (2018). "Home Oral Care Recommendations." Retrieved from https://www.ada.org/-/media/ADA/Science%20and%20Research/Files/HomeOralCare_Web_Final_1jul2018.pdf?e=en
- American Dental Association. (2005). "How to Floss." Retrieved from https://www.ada.org/-/media/ADA/Science%20and%20Research/Files/watch_materials_floss.ashx
- Center of Disease Control and Prevention. (2021). "Oral Health Tips." Retrieved from <https://www.cdc.gov/oralhealth/basics/adult-oral-health/tips.html>
- Center of Disease Control and Prevention. (2021). "Children Oral Health." Retrieved from <https://www.cdc.gov/oralhealth/basics/childrens-oral-health/index.html>

CONCLUSION

- Instant Translation devices may be a viable alternative to human translators in the dental healthcare setting particularly in Spanish, but not necessarily Russian
- Most common error type for instant translators was substitutions and the most common error type for human translators were additions (in both languages)
- Overall, instant translator technologies have a promising future in their application within the health-care industry as they can help close gaps in our care for patients that do not speak English fluently.