Transcription Error Comparison in Spanish Using the Wav2Vec Model

Marcus Berger and Elizabeth Nichols

U.S. Census Bureau

2020 Census Questionnaire Assistance (CQA)

Assisted callers over the phone to complete the 2020 Census
- Most callers first reached an Interactive Voice Response (IVR)
- Callers who had additional questions not addressed by the IVR could request to speak with a Customer Service Representative (CSR)
- With the caller's permission, these calls were recorded
- Assistance was available in English and 12 non-English languages

CSRs handled calls for a variety of topics, including but not limited to:
- Direct collection of census data
- Resolving technical issues with the online instrument
- Robocalls
- Calls from the media

CSRs relied on scripts, and were required to read these scripts verbatim to callers

Machine Learning and CQA Files

Not ideal

- Scripts required expert review and pretesting
- Quality monitors were in place to ensure script adherence
  - Costly and ad hoc
- Post-production behavior coding is used to evaluate respondent questions from recorded calls
  - Labor intensive and not in real time

Benefits of CSR scripts

- Provides a baseline for analysis of machine transcription (on the CSR side)

Benefits of Machine Transcription

Allow for callers to receive appropriate assistance through IVR to reduce reliance on CSRs

Evaluate CSRs for their ability to stay on script

Improve CSR ability to find relevant FAQs
- Could potentially prepopulate FAQs based on the caller's question as the caller is speaking
- Reduces the need for CSR to have to look up relevant FAQs

Focus on Spanish

This analysis focuses on comparison of machine transcribed calls compared to a manually transcribed baseline in Spanish

Overall transcription quality is often assessed by word error rate
- Word error rate is the sum of the number of substitutions, insertions, and deletions, divided by the total number of words

Focus on Spanish in particular
- Word error rate of 0.45 in analyzed calls

This analysis was a textual analysis of machine transcription against a manual transcription baseline
- We did not listen to audio for this particular analysis

Machine Learning and CQA Files

2020 Census Call Center audio files were collected between March and October 2020

- Calls were transcribed using the wav2vec transcription model
- Calls were clipped into 30-second snippets
  - 150 Spanish snippets were analyzed by human coders as a baseline comparison for machine transcriptions
  - The Spanish machine transcriptions had a higher word error rate than the English
  - Each manual transcription was done by two transcribers to ensure accuracy

This presentation is not a quantitative analysis of word error rate

- This is a qualitative analysis of the types of errors that show up in machine transcriptions

Manual transcription process

Each 30-second snippet was transcribed by two human transcribers fluent in Spanish

Certain conventions were followed to be able to match the format of the machine transcription

- No use of punctuation
- Proper use of accent marks
- Spelling out the names of numbers and letters
  - E.g. "twenty twenty" instead of "2020"
- Consistent spelling of filler words
  - We established a shared document with out own standardized spelling
  - E.g. "Umm", "Hmm", "Ahh"

Spanish errors

In some cases, short unstressed words/syllables were incorrectly machine-transcribed

Machine Transcription	Manual Transcription
"ese" (that)	"este" (this)
"vimos" (we saw)	"vivimos" (we live)
"un" (a)	"algún" (some)
"o que" (or that)	"ok" (ok)
"la" (the)	"una" (a)

Spanish errors

In other cases, there was a loss of grammatical information

Machine Transcription	Manual Transcription
"seleccione" (selection)	"selecciones" (selections)
"puede" (he/she can)	"puedo" (I can)
"viene" (he/she comes)	"vienen" (they come)

United States® Census Bureau

Spanish errors

In some cases, the machine transcription contained an entirely different and unrelated word

Machine Transcription	Manual Transcription
"hacer" (to do/make)	"a ser" (to be)
"es tan" (it is so)	"están" (they are)
"damas" (ladies)	"además" (besides)
"que frío" (how cold)	"qué sé yo" (what do I know)
"duro" (hard)	"dudo" (doubt)
"favorita" (favorite)	"ahorita" (about now/right now)

United States® Census Bureau

Main Takeaway - Understanding Limitations

Quantitatively, the Spanish data had an average word error rate of 0.45

None of the errors analyzed appeared to significantly alter the content of the speech

- None of the segments analyzed contained any "threat calls", so it is unclear whether these transcription errors would fail to capture such urgency

Further research

Further research could analyze whether errors in transcription could pull up the correct FAQs for CSRs

If not, further analysis could investigate whether the incorrect FAQ is displayed because of a transcription error or an error in the keyword search

- In case it is a transcription error, what kind of transcription error
- Puerto et al. (2023) has taken some first steps at this research

Potential shortcomings

This was only an analysis of transcripts, not audio recordings
- Some regional dialects of Spanish "aspirate" the "s" sound in some contexts
- Does not take into account difficult to hear speech
  - Either low volume or jumbled words
  - The human coder may have been able to use context to establish the word(s)
- Potential human error in manual transcription

AI moves fast!
- The models we used have potentially come a long way in their accuracy of transcription since this work was conducted

Thank you!

marcus.p.berger@census.gov

References

- Nichols, E., Puerto, M., Sadacca, B., Genter, S., Zajac, K. (2023). "Using Machine Learning for Quality Assessments of Call Center Interactions." Presented at *Federal Committee on Statistical Methodology Conference*, October 2023.

- Puerto, M., Nichols, E., Genter, S., Zajac, K., Sadacca, B. (2023). "Applying machine learning language models to document linkage to evaluate FAQ coverage." Presented at *Federal Committee on Statistical Methodology Conference*, October 2023.

- Puerto, M., Nichols, E., Chapman, C., Sadacca, B., Zajac, K. (2024). "Evaluating different Speech to Text Transcription Models." Presented at *Federal Computer Assisted Survey Information Collection (FedCASIC),* April 2024.