Using Semantic Textual Similarity to Improve Back Translation and Validation of Instruments in American Sign Language (ASL)

Wilson, J. F., Embree, J. A., Huber, J., Taylor, K.C., & D. Guthmann
Substance Abuse Resources & Disability Issues (SARDI) Program
Boonshoft School of Medicine, Wright State University

I. Abstract

Numerous barriers exist when attempting to provide culturally appropriate assessment instruments for persons who are Deaf. In addition to a scarcity of specialized instruments, validated versions of translated instruments are also rare, contributing to an increased likelihood of errors in screening, costs time and money for providers, and can contribute to Deaf consumers disenfranchisement with pursuing treatment. Instruments were translated; a process called forward translation, into ASL gloss by a team of Culturally Deaf experts. Thirty Bi-lingual Deaf and adult Children of Deaf Adults (CODA) participated in back translation. Lastly, cognitive interviews were conducted with culturally Deaf (n = 20) people across the United States, and field-testing was conducted (n = 450+). Instruments included the Satisfaction With Life Scale (SWLS), Rosenberg Self Esteem Scale (RSES), Beck Depression Inventory (BDI), Alcohol Use Disorders Identification Test (AUDIT), Drug Abuse Screening Test (DAST), Wide Range Interest and Occupation Test (WRIOT), Operational Network Evaluation Tool (O-NET), and Global Appraisal of Individual Needs (GAIN). Translations were compared to original text and ASL gloss using a semantic textual similarity (STS) analysis combining distributional similarity and latent similarity analysis. Preliminary results suggest these scores accurately predict similarity ($p < .05$) between ASL gloss and back translations, as well as agreement among back-translators. Although STS is not a substitution for a skilled panel of native-signers, the scores provided offer an objective way of comparing large samples of back translation content and reduce subjectivity among researchers during the process of choosing the most linguistically appropriate items.

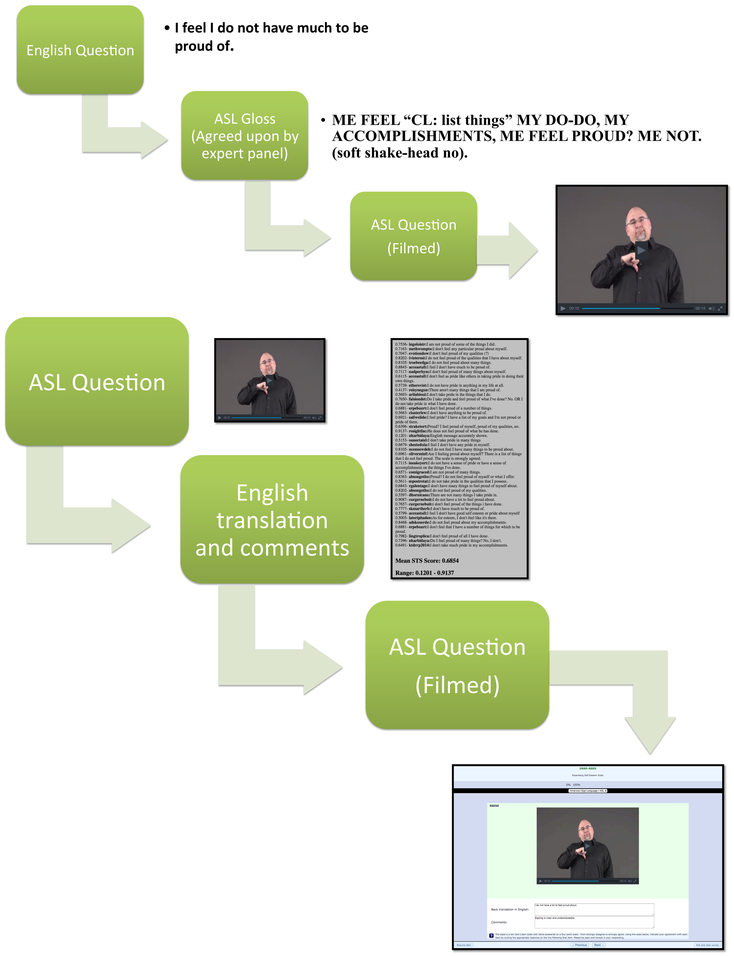
II. Population

- Individuals (n = 30), both Deaf and Children of Deaf Adults (CODA), participated in back translation. Bi-lingual Deaf and CODAs were chosen because their native language is ASL and they have English fluency.
- Cognitive interviews were conducted with culturally Deaf ($N = 20$) people across the United States to test furthur for reaganal signing differences.
- Lastly, field-testing was done through a secure online portal with respondents from 48 states ($N = 450+$). Each instrument was tested with people from a pool of approximately 1000 Deaf users of ASL.

III. Procedure

English versions of instruments were translated, in a process called forward translation, into ASL gloss by a team of Culturally Deaf experts. This process was completed for the Satisfaction With Life Scale (SWLS), Rosenberg Self Esteem Scale (RSES), Beck Depression Inventory (BDI), Alcohol Use Disorders Identification Test (AUDIT), Drug Abuse Screening Test (DAST), Wide Range Interest and Occupation Test (WRIOT), Operational Network Evaluation Tool (O'NET), and Global Appraisal of Individual Needs (GAIN). Upon completion of the forward translation, another team of ASL experts, both Deaf and Hearing, recorded a video of each question in ASL. During this part of the process, ASL grammar and structure are fine-tuned. The back translation consists of two parts – to translate the ASL version of what is seen into English and to provide any comments with regards to ASL structure, grammar, and syntax. These translations are compared to the original text and the ASL gloss, using a semantic textual similarity (STS) analysis, combining distributional similarity and latent similarity analysis (LSA). Based on this analysis, the expert panel decides whether any of the questions need to be re-recorded, and the process is repeated. After all portions of an instrument are equivalent in meaning to the original, the videos continue to the process of cognitive interviewing and field-testing.

V. Forward and Back Translation

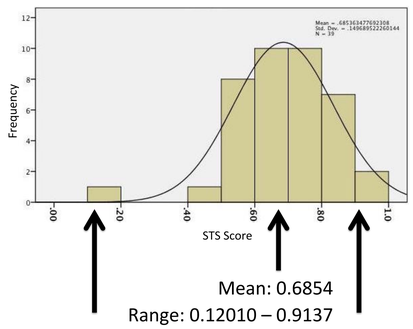


VI. Interpreting Scores Example (RSES)

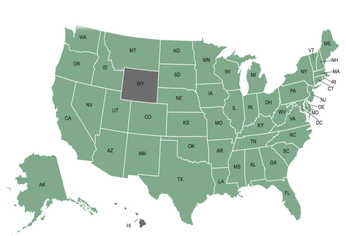
One of the reasons to use Latent Semantic Analysis (LSA) to measure semantic similarity is to improve analysis of back translation results and help decide which items need to be revised and reshot. Quantifying the results can reduce subjectivity and researcher bias, and aggregate scores can make it possible to look at large sets of data to understand how in item failed to perform so that it can be improved in future reshoots.

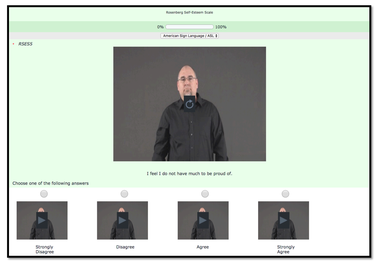
ASL Gloss → Back Translation
0.6238, *SD* = .1398

English → Back Translation
0.6854, *SD* = .1497

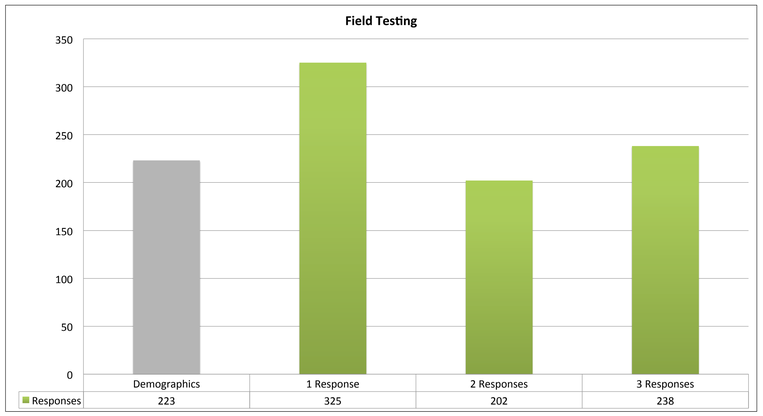


VII. Field Testing





Field testing was conducted for five of the instruments thus far. The GAIN-SS (n = 566), SWLS and RSES (n = 443), and the AUDIT and DAST (n = 434) were administered to Deaf participants in 48 States.



VIII. Results

Preliminary results suggest that STS scores can accurately predict similarity between ASL gloss and final back translations, as well as agreement among back-translators. STS scores comparing original English and final back translations are less promising, and suggest more complex conceptual differences between English and ASL. Although these scores cannot replace human interpretation, they do represent a step forward in reducing subjectivity among researchers doing this type of work, and can be a valuable tool to hel;p improve the accuracy and accessibility of instruments in ASL.

IX. Conclusions

Although STS is not a substitution for a skilled panel of native-signers, the scores provided offer an objective way of comparing large samples of back translation content and reduce subjectivity among researchers during the process of choosing the most linguistically appropriate items. The possibility of crowd-sourcing back translation efforts and using natural language processing (NLP) holds particular promise in reducing costs and minimizing differences in regional signs that can contribute to reduced understanding.

X. References

Sinclair, S.J., Blais, M.A., Gansler, D.A., et al. (2010). Psychometric properties of the Rosenberg self-esteem scale: overall and across demographic groups living within the United States. *Evaluation & Health Professions, 33*, 56–80.

Funding for this program was made possible by a grant (90DP0067) from the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR).