GEORGETOWN UNIVERSITY | beeckcenter social impact + innovation | Digital Benefits NETWORK

SEPTEMBER, 2022

Designing for Multilingual Translation

Complex benefits information creates unnecessary barriers for residents and navigators who must understand what's relevant to them so they can receive benefits. For non-native English speakers, these barriers are exacerbated.

While many state benefits agencies have made significant progress to make benefits information more readily accessible and easy to understand, translating content from English to other languages remains an area that needs reform. Transforming all of your agency's content can seem daunting–especially when you want the translated content to also be in plain language–but it is a necessary task to ensure equitable delivery of these benefits. Non-native English speakers are often part of the population that needs these benefits the most but for whom they are most inaccessible.

Contributing Authors

Ariel Kennan
Fellow, Digital Benefits Network, Beeck Center for Social Impact + Innovation

Elle Meyers
Beeck Center Student Analyst

Translation Approaches

Translation greatly increases the accessibility of content for people who do not speak English as their first language. When your content is written in plain language, it is easier to translate. However, translation can be expensive and add another step to your process. There are a few ways to make translation more accessible to your organization:

+ **Create a glossary of common words and phrases.**
  A glossary can help track words that your organization commonly uses and may be specific to the context you work in. For example, the Center for Participatory Change has created English-to-Spanish glossary for terms used in their community organizing.

+ **Identify materials that are accessed most frequently.**
  Use web analytics, flyer distribution counts, or other metrics to identify the web pages with the highest number of visitors and commonly used printed materials and prioritize them for translation. You may find that there are pages on your website that are infrequently accessed that can be served by a translation API (see below) so you can prioritize your translation budget for the most accessed pages.

+ **Conduct research to better understand needs for non-native English speakers.**
  Prior to undertaking the development of the custom translation model, the City of San José worked with Code for America to better understand low-income and non-English speaking residents' digital service access needs. They conducted in-person interviews with 19 residents and utilized interpreters to help facilitate interviews with Spanish- and Vietnamese-speaking residents. They found that many residents preferred a plain language, simple English version as the machine-translations were inaccurate or not context specific. Additionally, they found that some digital phrases like "sign in" or "log out," which are commonly seen in English, sometimes don't exist for speakers of specific dialects, such as pre-1975 Vietnamese speakers.

Image of various devices displaying ACCESS NYC's website in a different language on each screen.

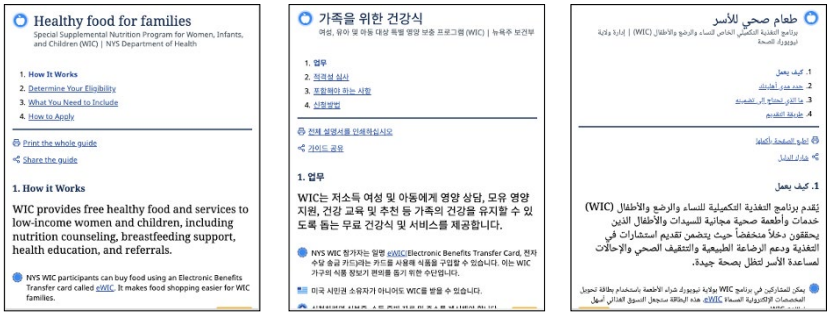
+ **Design for equitable multilingual content.**

In 2016, the New York City Mayor's Office for Economic Opportunity (NYC Opportunity) set out to redesign ACCESS NYC, New York City's public benefits information site and eligibility screener. The goal was to make it simpler for residents to get easy-to-understand information about more than 80 benefits and programs, screen for potential eligibility for more than 40 benefits, and find help nearby. NYC Opportunity worked with a team of professional translators to make ACCESS NYC available in seven languages in 2017, adding four more languages in 2018. ACCESS NYC now supports 11 languages: English, Arabic, Bengali, Chinese (Traditional), French, Haitian Creole, Korean, Polish, Russian, Spanish, and Urdu. It is the first website in New York City to be fully compliant under Local Law 30, which mandates that city agencies make public services available in these languages, enabling 86 percent of residents to navigate in their primary language.

The redesigned site included a complete rewrite of more than 450 pages of public benefits content to ensure it was accurate, written in plain language, and available in commonly spoken languages. ACCESS NYC's typography also created dignity across all languages by using Google Noto. Using this open-source typeface, which was developed specifically for multilingual communications, characters in all languages are stylistically consistent on ACCESS NYC. This way, visual readability does not change when the language changes.

Find Out More on the Digital Benefits Hub

Read more about the ACCESS NYC redesign



Example of a program guide on ACCESS NYC in English, Korean, and Arabic.

+ **See if your organization or agency has a master services agreement for translation services.**
Some governments make master service agreements with professional translation companies to provide translation services. You may be able to buy into the organization's agreement rather than contracting with them directly, which can save time and money.

+ **Integrate a translation API.**
While human-translated content is a best practice and the easiest to read, it may not be possible to use for your entire project. By using a translation API, you can offer full pages or supplement pieces of human-translated content across your websites and print materials, using computer code that automatically translates your content. There are many translation APIs available as open source or paid resources. Your organization will need to evaluate which solution works best for your needs based on considerations such as quality of translations for the languages you plan to support, integration with your technology, and cost.

+ **Build a custom model to better target your translated content to the needs of your audience.**
The San José custom translation model reflects the dialects unique to the Spanish- and Vietnamese-speaking communities of San José. It is also trained with words and phrases specific to the city government context and designed to continuously improve.

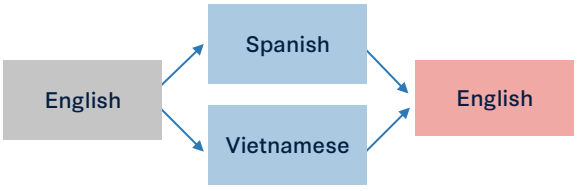
The custom San José translation model is built using Google AutoML, which has extensive documentation on preparing training data, creating and managing datasets and models, and evaluating models. San José's team found that in order to create good machine learning content, humans must be involved at every step to do tasks like collecting data, translating phrases, collecting errors, and retraining the model.

As a first step to building the model, the team collected high quality sentence pairs of English sentences matched to their Vietnamese or Spanish translations. They built these up from the existing free Google Translate model, which provided a baseline, but also increased the amount of additional content that needed to be collected. The base had better quality Spanish translations, requiring the new model to only need about half as much additional Spanish data than Vietnamese data in order to build to a quality that tested well with residents.

The San José team took into consideration nuance in language that reflect the cultures and dialects present in their community, such as orienting toward Mexican Spanish, as almost 20 percent of all immigrants to Santa Clara County are from Mexico. They also included multiple phrases for words that didn't have a one-to-one translation, such as graffiti in Vietnamese. Unfortunately, the San José team found that machine learning is challenged by the "slang, spelling and grammatical errors, idioms and figures of speech" that are often present in resident requests. City staff emails also include government jargon, but the model excels at translating within the rules of the bureaucratic language.

Use Case
Residents submit non-English service requests

Source Text
San José app service request descriptions



Use Case
City staff respond in English

Source Text
100 sentences from Customer Call email scripts

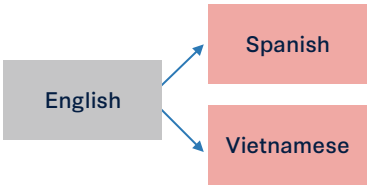


Diagram demonstrating the translation directions and sources utilized in the San José custom model development.

Find Out More on the Digital Benefits Hub

Read more about the San José custom translation model.

+ **Use hybrid approaches.**

The New York City Office of Technology and Innovation has been exploring how to keep content automatically and continuously up-to-date in the 11 languages most commonly spoken by New Yorkers. It does so using a combination of neural machine translation (NMT) and human translation vendors. Through a partnership with United States Digital Response, the NYC[x] Innovation Fellows built an open source prototype called Easy Localization System Access (ELSA) which can pull English content from any source, such as code, text file, or Wordpress and send it to any third-party translation API. It can also track changes between versions of content and only send the minimal amount of content for translation to help save money and time. The process can be fully automated or inserted as part of an editorial workflow where a person or series of reviewers who speak the language can review and approve the content before it is published.

Get in Touch

Our Digital Benefits Network team is here to help!

Visit us at the Digital Benefits Hub

Please contact us with any thoughts, questions, or potential collaborations via email at digitalbenefits@georgetown.edu