

Usability Test Report for Google Translate

Brian Bobo
TCC 6120 Fall 2015
Final Project Report

December 19, 2015

Executive Summary

Usability testing of Google Translate was conducted in Mexico City in December, 2015. One pilot tester and three test participants completed nine tasks and a brief survey in each 45-minute session. All participants were either bilingual or multilingual, and the three test participants were all employed either as language instructors or translators. All participants shared a common fluency in English and Spanish, so those languages were used for translation tasks. The test goals were to discover problems related to interface learnability and ease-of-use when performing common translation tasks, as well as issues with the understandability and perceived accuracy of the machine translation.

Participants generally found the interface easy to learn and use for simple text translation, but less so for more complex speech and document translation. Even the participants who otherwise performed well and successfully completed all tasks were slowed down by some interface elements. Most of the interface issues identified during this test were related to actionable properties of specific interface elements not being clearly communicated to the user, causing irritation and delay. Participants also felt that the translation quality was sufficient for informal use, such as when traveling, but probably would not be accurate enough for professional use.

This report describes the purpose, objectives, and methodology of the test. Participant profiles and details of test procedures and tasks are included. Task performance and survey data are presented and discussed, and recommendations for improvement are proposed and rated for severity. Copies of the test script and participant satisfaction questionnaire are included in the appendices.

A short [video summary](#) of the test and this report is also available.

Introduction

Google Translate is a multilingual machine translation service. It enables users to translate text, speech, handwriting, images, websites, and documents in 90 languages. The platform offers web and mobile interfaces as well as an application programming interface (API). Google Translate uses statistical analysis rather than rule-based algorithms. The service does not attempt to translate directly from the source language to the target language. Instead, it translates from the source language to an intermediate language, and then from the intermediate language to the target language.

Usability tests are a means of evaluating a product or service by testing it with representative users performing typical tasks. These tests yield quantitative and qualitative data that help evaluators identify problems and measure user satisfaction. The resulting data are then analyzed, and recommendations for improvement are proposed based on the results.

I conducted a usability test of Google Translate using three typical multilingual users, each with varying levels of fluency in their non-native languages. Users were observed performing tasks on a laptop computer using the Google Translate web interface. User task completion success rate, time on task, and number of errors were recorded. A survey was also administered to measure user satisfaction.

Purpose and Objectives

The primary purpose of this usability study was to identify problems that users might experience when performing typical text, speech, and document translation tasks using the Google Translate service. Secondary intents included assessing the usability of non-translation interface functions, as well as measuring user satisfaction with the interface and the machine translation results. The test goals were to discover problems related to interface learnability and ease-of-use when performing common translation tasks, as well as issues with the understandability and perceived accuracy of the machine translation.

To be effective, cross-platform statistical translation tools must be easy to learn and use. Google relies on users in the translate community to improve the tool. Machine translations are only about 70 percent accurate, so the easier the tool is to use, the more active users there will be, and their increased input will in turn even further improve the accuracy and usability of the service.

Success rate and time on task were the primary metrics used to measure participant performance. Significant errors were also logged and reported, and user satisfaction survey results were tabulated to indicate participants' perceptions of learnability, ease-of-use, and translation quality.

Methodology

Standard usability testing methodology was used to evaluate the Google Translate service. Representative users were observed performing typical translation tasks. Since all participants were proficient in both English and Spanish, those two languages were selected as the source and target languages for the test tasks. Participants were tested individually, and employed a concurrent think-aloud protocol while executing tasks to help the observer understand their behavior. Participants also completed a user satisfaction survey. The qualitative and quantitative data from the test were then analyzed and summarized to make recommendations for improvement.

Participants

All study participants were bilingual or multilingual. Two were female and one was male. All users were proficient using web and mobile interfaces in general. All were familiar with Google Translate, and had used it for occasional text translations. None were proficient with the service, and none had used any of Google Translate's advanced features, nor had any of the participants used it for anything other than simple text translations.

Participant 1 is from the United States. Her first language is English, and she also speaks Spanish and French with some proficiency. She teaches English to Spanish-speaking

business clients in Mexico City. Participant 2 is from Mexico. His first language is Spanish, and he also speaks English with some proficiency. He works at the same company as Participant 1, and also teaches English to Spanish-speaking business clients. Participant 3 is from the United Kingdom. Her first language is English, and she also speaks Spanish, Italian, and French with near-native fluency. She is a professional document translator.

Additionally, a pilot subject participated in a preliminary test. The pilot participant is from Greece. His first language is Greek, and he is also fluent in English and Spanish. He is an economist who teaches at a university in Mexico City. User profiles are compared in the following table:

	Native Country	First Language	Additional Languages	Occupation	Google Translate Proficiency
Pilot Participant	Greece	Greek	Spanish, English	Economist	Beginner
Participant 1	United States	English	Spanish, French	Language Instructor	Beginner
Participant 2	Mexico	Spanish	English	Language Instructor	Beginner
Participant 3	United Kingdom	English	Spanish, French, Italian	Translator	Beginner

Table 1: Participant Profiles

Roles

Since this was an independent project, I acted as the test facilitator, note-taker, audio recorder, and data analyst.

Test Environment and Equipment

The test environment was relaxed and informal. The pilot participant and Participant 3 were tested at their respective places of residence. Participant 1 and Participant 2 were

tested in a small conference room at their place of business. In all cases, the testing environment was quiet and private during the tests, and no interruptions occurred.

The tests were conducted using a laptop computer with a 15-inch screen. The Google Translate web interface was accessed using the Google Chrome web browser, version 47.0.2526.106 m. Users were allowed to choose the pointing device used during the test (either the laptop's default touchpad or a wireless mouse) depending on their preference. The built-in laptop microphone was used as the sound input both for speech translation tasks and participant voice recording. The open source audio recording and editing program Audacity was used for recording and playback of the test sessions.

Procedure

At the beginning of each test, the participant was briefed on the purpose of the test, and of usability testing generally. It was explained to them that it is the interface being evaluated rather than the participant. Each participant was then asked to sign an informed consent statement acknowledging that their participation is voluntary and that an audio recording would be made. Participants were also provided with an explanation of the concurrent think-aloud protocol. Even though all participants were familiar with Google Translate, none were proficient users, so each was given a brief overview of the service and its different translate functions. Participants were then given the opportunity to ask questions.

Once the participant was ready to begin, they were reminded to vocalize their thoughts, the recording was started, and the participant was asked to read aloud the title and description of the task. Participants completed the tasks one at a time in sequence. The facilitator remained a silent observer during task execution, and only provided assistance if the participant was unable to move forward with the task. During task performance, the facilitator observed and took notes of any important details that were not vocalized.

Following task completion, participants were asked to complete a short user satisfaction survey. After completing the survey, participants were asked about their overall test experience during a short debriefing session.

Tasks

Translation tasks included translating text (single word, complete sentence, and short passage), translating speech (single word and complete sentence), translating a web page, and translating an uploaded document. The test also included one interface function that doesn't involve translating (saving a translation to a phrasebook).

Users were asked to complete the following eight tasks:

Task 1	Translate the following words from English to Spanish by typing them into the text box, one at a time: strategy, soap, Thursday, monkey, avocado
Task 2	Translate the following words from English to Spanish by speaking them into the microphone, one at a time: strategy, soap, Thursday, monkey, avocado
Task 3	Translate the following words from English to Spanish by typing them into the text box, then translate the resulting Spanish word back to English, one at a time: cranberry, straight, biscuit
Task 4	Translate the following sentence from English to Spanish by copying and pasting it into the text box: In industrial settings social pressures can lead to misinterpretation, mistakes, and accidents.
Task 5	Translate the following sentence from English to Spanish by speaking it into the microphone: In industrial settings social pressures can lead to misinterpretation, mistakes, and accidents.
Task 6	Translate the following short passage from English to Spanish by copying and pasting it into the text box: Everyday tasks are not difficult because of their inherent complexity. They are difficult only because they require learning arbitrary relationships and arbitrary mappings, and because they sometimes require precision in their execution.
Task 7	Translate the following web page from Spanish to English: http://www.museofridakahlo.org.mx/esp/1/el-museo/la-casa/la-casa-azul
Task 8	Translate the following document from English to Spanish: concussion_fact_sheet.pdf (Original document at http://www.cdc.gov/headsup/pdfs/highschoolsports/athletes_fact_sheet-a.pdf)
Task 9	Translate the English word "phrasebook" to Spanish by typing it into the text box, and save the translation to your phrasebook.

Table 2: Task Descriptions

Survey

After completing the test tasks, users were presented with a questionnaire intended to measure their satisfaction with the usability of the interface and the quality of the translation. The full questionnaire can be found in Appendix B.

Results and Discussion

Task Completion Success Rate

The following table shows the number and percentage of tasks, per participant, that were successfully completed without facilitator intervention.

Participant	Task 1	Task 2	Task 3	Task 4	Task 4	Task 5	Task 6	Task 7	Task 8	Task 9
Pilot	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
1	✓		✓	✓	✓	✓	✓			
2	✓		✓	✓	✓	✓	✓			
3	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Success	4	2	4	4	4	4	4	2	2	2
Rate	100%	50%	100%	100%	100%	100%	100%	50%	50%	50%

Table 3: Task Completion Success Rate

Time on Task

The following table shows the time, in seconds (rounded to the nearest whole second), that each user needed to complete each task.

	Pilot	Participant 1	Participant 2	Participant 3	Average
Task 1	31	38	41	23	33
Task 2	73	97	86	34	73
Task 3	58	58	56	49	55
Task 4	13	13	11	8	11
Task 5	18	33	27	14	23
Task 6	16	18	14	10	15
Task 7	28	26	30	13	24
Task 8	24	50	44	17	34
Task 9	39	54	51	22	42

Table 4: Time on Task

Errors and Issues

The pilot tester and Participant 3 were both able to complete all tasks without any assistance from the facilitator. Participant 1 and Participant 2 both required some prompting to complete four of the tasks (the same four tasks).

Task 2 (speech translation of single words) proved problematic because users did not realize that they needed to either clear the text box between words, or stop and restart the speech input between words. If they failed to do so, the service would simply add the next word to the previous translation.

When translating a web page (Task 7), the user pastes or enters a URL into the source language text box, and then a link to the translated page is displayed in the target text box.

Participants 1 and 2 did not seem to recognize that they needed to click the link to get the translated page, and simply waited for something else to happen.

Similarly, Task 8 required users to upload a document for translation. Once the document was uploaded, neither Participant 1 nor Participant 2 realized that they needed to press the blue “Translate” button to display the translated document. This is probably because the instant translation feature was turned on for all tasks. So, during all previous tasks, text would be translated as it was being typed. Even though the blue “Translate” button is still visible in instant translation mode, it need not be pressed to translate text, so the participants may have become conditioned to ignore it.

Participants 1 and 2 also had issues with Task 9 (saving a translation to the phrasebook). There is a star icon in the target language text box that must be clicked to save the translation to the phrasebook, but this representation did not seem sufficient to suggest this capability to the users. If the user opens the actual phrasebook using its associated icon, they will see a message instructing them to click the star to save the translation, but this is not readily apparent. Even the participants that completed this task without prompting spent some time hovering over the icons and reading the tooltips to figure out how to complete this task.

Participant 3 had the fastest task completion times of all participants. This is largely because she was the only participant that cleared the source language text box by clicking the “X” symbol displayed in the upper right corner of the box. Many users relate this symbol with closing a window or box, rather than simply clearing it, so the other users backspaced through their previous entries to clear them. Also, participant 3 was the only one to use the “swap languages” button when switching the source and target languages, whereas the other participants manually re-selected the languages using the buttons above the source and target language text boxes.

All users commented about the inconsistency of the translations when executing the “round trip” translation task (Task 3). Often, words translated from English to Spanish will not translate back to the original English word.

Some of the interface elements discussed above that caused confusion for participants are shown in the image below:

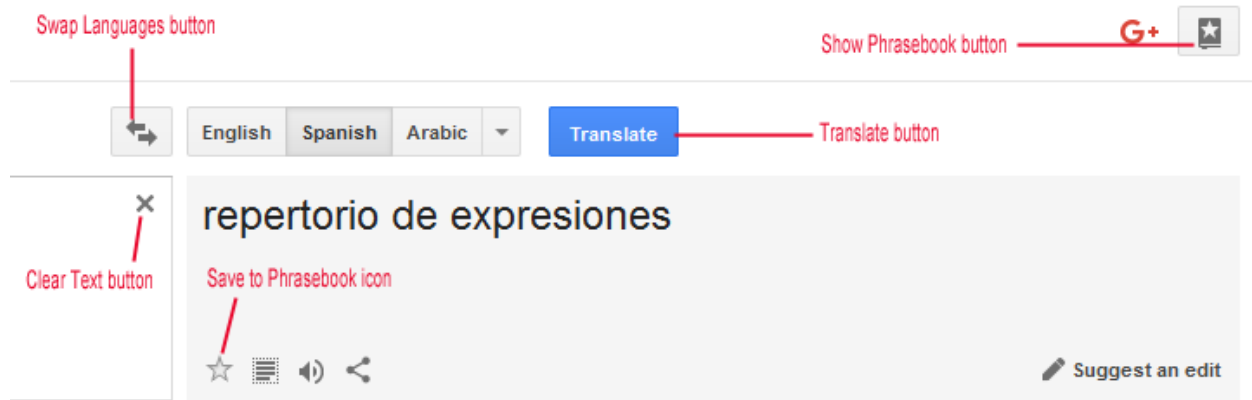


Figure 1: Google Translate Interface Elements

Survey Results

The following table shows the results of the participant satisfaction questionnaire that was administered after task testing:

	SD	D	N	A	SA
The interface was easy to use		1	1	2	
The interface was easy to use for text translation				3	1
The interface was easy to use for speech translation		2	1	1	
The interface was easy to use for document translation		1	1	2	
I easily learned the functions I needed to use to complete tasks	1	1	1	1	
I was never confused by the interface while completing tasks	1	1	2		
I would rely on the translations while traveling			1	3	
I would rely on the translations for professional work	1	2	1		
Translated web pages and documents were mostly readable and understandable				4	
Translated web pages and documents were easy to read and highly accurate	1	2	1		
Word translations were mostly accurate			1	3	
Sentence and short passage translations were mostly accurate		2	2		

SD=Strongly Disagree, D=Disagree, N=Neutral, A=Agree, SA=Strongly Agree

Table 5: Survey Results

All participants found the interface easy to use for text translation tasks, but less so for speech translation tasks. The interface was not generally rated highly for learnability, and there were clearly some confusing interface elements. Most would rely on the translation while traveling, but not for professional work. For all tasks, translations were perceived as generally accurate, but not highly accurate.

Recommendations

The following table presents the proposed recommendations for interface improvement, the justification for each change, and the severity rating for the corresponding issue:

Recommendation	Justification	Severity
Use a different icon (such as a broom) or a text cue to indicate the clear text box function	Most users did not immediately associate the “X” icon with clearing the field and backspaced through the text to clear instead, which increased time on task	Low
Disable or remove the “Translate” button when the service is in instant translation mode	Users become accustomed to the instant translation, and become blind to the button’s functionality when it is actually needed	Medium
Use an icon other than a star (maybe a book with an arrow pointing to it) to indicate the Save to Phrasebook functionality	None of the users readily associated the star symbol with save to phrasebook action	Medium
Similar to the above recommendation, use an icon other than a star to represent the Open Phrasebook functionality	None of the users readily associated the star symbol with the open phrasebook action	Medium
Add a small text cue near the Swap Languages button	The arrows pointing in opposite directions would seem to be clear, but most users ignored the button because they did not immediately recognize its function	Low
Add a text cue to press the Translate button after uploading a document	Users who have become blind to the Translate button’s function while working in instant translate mode need an additional prompt to complete the action	Medium
Add a text cue to click the link after a URL is entered for web page translation	Even though users clearly see the link displayed in the target language box, they expect to be taken automatically to the page and don’t realize that they have to click the link	Medium

Table 6: Recommendations

The following severity rating scale was used to rate the impact of issues:

- High: Significant impact on the user's ability to complete a task. Would result in task failure for most users or cause extreme irritation.
- Medium: Slows down or confuses the user. Causes task failure for some. Delays and irritates most.
- Low: Might annoy the user, but would not present an obstacle to task completion. Could cause some hesitation.

Conclusion

Most of the participants found the interface easy to learn and use for simple text translations, but less easy to learn and use for speech and document translation. The primary interface problems lie in the way that affordances are communicated to the user. Some interface elements are confusing or do not clearly suggest to the user the action that they represent. The time on task data and the satisfaction survey results reveal that even the participants who otherwise performed well and successfully completed all tasks were slowed down by some interface elements. The recommendations in this report address these particular elements.

During performance of the test tasks, users sometimes vocalized their thoughts about the quality of the machine translation. They also indicated their satisfaction with translation quality on the satisfaction survey. As might be expected, they felt that the translation quality was good for single words or short blocks of text, and was suitable for casual use such as traveling. But they felt that the overall translation quality denigrated when translating long passages of text or whole documents, and that they probably would not rely on the service for professional work. However, no recommendations for improving translation quality were included in this report. Accuracy problems with machine translations are well-documented, and are algorithmic and computational issues that affect all such services. Recommendations for improving translation quality are well beyond the scope of this test and report.

Appendix A – Usability Test Script

Thank you for volunteering to participate in this usability test of Google Translate. Usability tests are a means of evaluating a product or service by testing it with representative users performing typical tasks. These tests yield data that help evaluators identify problems and measure user satisfaction. The data are then analyzed, and recommendations for improvement are proposed based on the results.

As you know, Google Translate is a multilingual machine translation service. It enables users to translate text, speech, handwriting, images, websites, and documents in 90 languages. Today, we will be using the Google Translate web interface, and we will be translating text, speech, and documents from English to Spanish and vice versa.

I will be observing you and taking notes as you complete the assigned tasks. Please remember that we are evaluating the interface, and not your performance, so there is no need to worry about how you are doing.

A concurrent think-aloud protocol is used during the test, which means that you should feel free to vocalize your thoughts and frustrations as you complete each task. In general, though, you should not ask me for assistance unless you feel that you are truly stuck and cannot complete the task without help.

There are nine tasks, followed by a brief questionnaire. The entire session should take about 30 to 40 minutes to complete.

Also, I will be making an audio recording of your comments during the test. This is only for evaluation purposes, and will not be distributed to anyone else. If you are comfortable with this, I have a consent form I would like for you to sign.

Do you have any questions before we begin?

Appendix B – User Survey

GOOGLE TRANSLATE USABILITY TEST PARTICIPANT QUESTIONNAIRE

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
The interface was easy to use	☐	☐	☐	☐	☐
The interface was easy to use for text translation	☐	☐	☐	☐	☐
The interface was easy to use for speech translation	☐	☐	☐	☐	☐
The interface was easy to use for document translation	☐	☐	☐	☐	☐
I easily learned the functions I needed to use to complete tasks	☐	☐	☐	☐	☐
I was never confused by the interface while completing tasks	☐	☐	☐	☐	☐
I would rely on the translations while traveling	☐	☐	☐	☐	☐
I would rely on the translations for professional work	☐	☐	☐	☐	☐
Translated web pages and documents were mostly readable and understandable	☐	☐	☐	☐	☐
Translated web pages and documents were easy to read and highly accurate	☐	☐	☐	☐	☐
Word translations were mostly accurate	☐	☐	☐	☐	☐
Sentence and short passage translations were mostly accurate	☐	☐	☐	☐	☐