ANALYZING CROWDSOURCING MODEL IN TRANSLATION (SUBJECT: TED TRANSLATION)

АНАЛІЗ ВИКОРИСТАННЯ МОДЕЛІ КРАУДСОРСИНГУ В ПЕРЕКЛАДІ (НА ПРИКЛАДІ ПЕРЕКЛАДІВ ДЛЯ TED)

Debela A.S.,
orcid.org/0000-0001-7690-9948
Second-year master's Student at the Faculty of Linguistics
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

Ivashkevych L.S.,
orcid.org/0000-0001-7166-5331
Candidate of Philological Sciences, Associate Professor,
Associate Professor at the Department of Theory, Practice and Translation of German
National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

The article explores the principles of crowdsourcing and their use in translation. Crowdsourcing can be considered part of so-called community translation, a relatively new approach in the translation field of mass voluntary translation, in which everyone is involved and there are no limits on the amount of their contribution. Relevant publications are analyzed, showing the reasons for and the relevance, the features and the exiting experience of crowdsourcing in translation, and its advantages as compared to professional translation. Particular attention is paid to crowdsourcing translation of subtitles for TED Talks annual conference, analyzing the approaches to creating, translating, and quality assurance of subtitles. In particular, the roles in the quality control process are divided into a transcriber (who transcribes the video of the talk and creates subtitles), a translator, an editor, and a language supervisor or a language coordinator (who does the final review of translated and edited subtitles and approves them for publication). Strategies and tactics for using crowdsourcing in subtitling translation and its peculiarities are investigated. The main problems encountered by a linguist in this type of translation are analyzed. These include, in particular, specific features of some preachers' speech: accent, peculiar pronunciation, junk words, speech volume. Wordplay is often a complication. We also analyzed the translation requirements for TED Talks subtitles and compared them with the standard localization requirements for the translation industry. It was found that part of the requirements for volunteer subtitling for TED Talks is the same as those of the translation industry, namely: adequate translation of content, preservation of formatting, localization of time format, units of measure, etc. At the same time, some factors differ: volunteers must also adhere to subtitling requirements, for example: considering the length of subtitles, number of characters, reading speed, etc. The article concludes with the possible development of crowdsourcing in the translation industry in the future and associated changes in quality assurance.

Key words: crowdsourcing, community translation, subtitling, transcribing, TED Talks.

Статтю присвячено дослідженню принципу краудсорсингу та його використанню в перекладі. Краудсорсинг можна вважати частиною так званого «громадського перекладу», відносно нового в перекладацькій сфері підходу масового волонтерського перекладу, у якому беруть участь всі охочі і немає обмежень щодо обсягів їх внеску. Проаналізовано відповідні публікації, показано причини появи та актуальності, особливості та вже наявний досвід застосування краудсорсингу в перекладі та його переваги в порівнянні із професійним перекладом. Особливу увагу у статті присвячено краудсорсингу в перекладі субтитрів до відеозаписів виступів щорічної конференції TED Talks: проведено аналіз підходів до створення та перекладу таких субтитрів та до забезпечення їх якості. Зокрема, йдеться про розподіл ролей у процесі контролю якості субтитрів на транскрибатора (розшифровує відеозапис виступу та створює субтитри), перекладача субтитрів, редактора та мовного експерта (виконує фінальну перевірку перекладених та відредагованих субтитрів та надає допуск до публікації). Досліджено стратегії й тактики використання краудсорсингу в перекладі субтитрів, його особливості. Проаналізовано основні проблеми, з якими стикається лінгвіст під час такого виду перекладу. До таких належать, зокрема, характерні особливості мовлення певних мовців: акцент, своєрідна вимова, «слова-паразити», гучність мовлення. Часто складністю стає гра слів. Також було проведено аналіз вимог до перекладу субтитрів для TED Talks та співставлення їх із стандартними для галузі перекладу локалізаційними вимогами. Виявлено, що частина вимог до волонтерського перекладу субтитрів для TED Talks збігається з вимогами перекладацької індустрії, а саме: адекватна передача змісту, дотримання форматування, локалізація форматів та одиниць вимірювання, дотримання правопису цільової мови тощо. Водночас відмінним фактором є те, що волонтерам слід дотримуватися ще й вимог до створення субтитрів, наприклад, зважати на довжину субтитрів, кількість знаків, враховувати швидкість читання тощо. Стаття завершується висновками про можливий розвиток краудсорсингу в перекладі в майбутньому та пов'язані із цим зміни в забезпеченні його якості.

Ключові слова: краудсорсинг, громадський переклад, субтитрування, транскрибація, TED Talks.

Problem statement. Educational platforms are growing in popularity these days. At the same time, there is a growing demand for translation to disseminate information. Volunteers, including inexperienced linguists, are often recruited to translate materials. In this article, the main problems faced by the linguist in this type of translation are analyzed.

Analysis of recent studies and publications. Several publications are devoted to the principle of crowdsourcing, including translation. Besides J. Howe [5], who first used the word "crowdsourcing", it is worth mentioning such researchers as M. O'Hagan [8], R.V. Povoroznyuk [2], M. Taibi [12], K. Losse [7], I.S. Bakhov [1], etc.

Highlighting the previously unresolved parts of the overall problem. The article focuses on the specifics of using the principle of crowdsourcing in the subtitling of video for the annual TED Talks conference.

Presentation of the main material. The term "crowdsourcing" was firstly used in 2006 by Jeff Howe, Wired magazine editor, in his article "The Rise of Crowdsourcing" [5]. The main idea of crowdsourcing is outsourcing a task to an undefined but large group of people ("crowd"). Crowdsourcing allows a large amount of work to be done and a big number of specialists to be involved.

This type of work organization allows creating a large network of providers. It is usually used in large-scale projects to delegate a task to a group of providers. These can be professionals in a particular field as well as beginners and people with no relevant education, but who have basic knowledge and skills.

The crowdsourcing method is also widely used in the field of translation. M. O'Hagan in his work "Evolution of User-generated Translation: Fansubs, Translation Hackings and Crowdsourcing" [8] studies the phenomenon of "community translation". He describes the experience of Facebook, a social network, that "launched crowdsourcing in 2007 to let their users translate selected parts of the Facebook website into different language versions" [8, p. 112, 7, 1]. 2008 Facebook involved over 30,000 users around the world to translate its website into 16 languages. Volunteers did not need to translate any specific amount of text. They could choose any fragment or string they wanted to translate [10, p. 938].

As a result, user knowledge helped not only to translate the Facebook site but also to localize it so that the site met the users' expectations. An analysis of this experience shows that Facebook users were better at some tasks than professional translators because they knew the social network system. It was therefore much easier for volunteers to find the equiv-alent of a particular option on Facebook in the target language [1, p. 18].

According to the developers of Facebook, their main aim was not to save money. Most of the money was spent on developing the platform, where translators worked [8, p. 112].

In the beginning, students were involved in the project, but soon everyone could join in. Of course, professional linguists controlled the process and checked all translations, but this was one of the last steps of this project.

R.V. Povoroznyuk uses in her article "Interlinguistic dimension of TED Talks online-platform" (укр. «Інтерлінгвальний вимір онлайн-платформи TED Talks») [2, p.103] the concept "community translation" (укр. «громадський переклад»), also known as "collaborative translation" (укр. «колаборативний переклад»). She says that this type of translation plays a big role for many platforms and resources. A scientist considers crowdsourcing to be an online collaboration between two or more translators including editors.

Povoroznyuk refers to O'Hagan and Taibi's definitions of community translation. M. Taibi notes that this type of community translation is indeed to provide accessibility of information in multilingual communities [12, p.15].

M. O'Hagan defines this as the community translation for community needs. That is, the process can be considered volunteer work, which is usually done by Internet users [9].

There are a lot of online resources and materials that need to be translated into a big number of languages. So, a big number of platforms use crowdsourcing for such tasks. For example: TED [15], Coursera [3], GlobalVoices [4], Wikipedia [18], Khan Academy [6] etc.

One such platform is TED Conference (Technology Entertainment Design) [15]. This is an annual conference that is known for its lectures – TED talks. These Talks cover many areas of activity and discuss actual issues, talking about important discoveries, research, social problems, etc. Therefore, to disseminate information to large audiences, there is a need to translate the lectures into multiple languages. And as a consequence, the need for a huge number of translators, transcribers, editors, etc. arises.

The organizers of the TED Translators community, to ensure a high quality of translation, have structured the roles of volunteers and the requirements for each of them [15;17].

Translator/Transcriber: Translators and volunteers create subtitles. They should be fluent in work languages: for transcribers, it is necessary to

be fluent in the transcription language (the original language of TED talk), for a translator – in both (source and target);

Reviewer: Reviewers are experienced volunteers who review transcripts or translations. These volunteers check subtitles not only for mistakes in grammar but also the length of subtitles, reading speed, etc. They should have advanced skills in source and target languages and have a "strong understanding of TED guidelines and best practices" [16]. They also can provide feedback for inexperienced volunteers. Volunteers are recommended not to start the review before they have translated or transcribed at least 90 minutes of talk;

Language Supervisor: Language Supervisors (or Language Coordinators [17]) are experienced volunteers who help with linguistic questions, post-editing, or technical issues. They do a final check on the transcript or translation and approve it for publication. Language Coordinators need to have 40 published reviews, also they are "selected for their expertise and their history of helping their language community in the Open Translation Project" [16].

There is also Mentoring program on the TED platform. Volunteers can apply for this program as mentees to improve their subtitling and reviewing skills.

The TED Translators community is growing steadily. And, as already noted, the use of crowdsourcing involves delegating the task to the "crowd" and it is not always the case that the task is carried out by professionals. In translation, non-professional translators may also be involved in the process. This is why there is a need for certain rules and requirements for the task to ensure high-level quality translation and avoid mistakes.

Since translation for TED Talks is not just text translation, but the actual creation of subtitles, subtitling peculiarities should also be taken into account. To do this, some subtitling tips have been developed:

1. "When a subtitle is longer than 42 characters, break it into two lines.

2. Never use more than two lines per subtitle.

3. Keep broken lines as close in length as possible.

4. Keep 'linguistic wholes' together when breaking lines.

5. Keep the subtitle reading speed at a maximum of 21 characters/second.

6. Compress subtitles over 21 characters/second" [14].

The organizers also provide a style guide that TED volunteers have to follow.

1. Informal over formal. The translator should choose informal terms, not formal or academic.

2. Modern over traditional. The linguist should use modern terms.

3. Personal over generic. The translator should avoid word-for-word translation, but keep the speaker's original color, style.

4. Global over regional. Universal terminology must be used.

5. Idioms. Linguists should avoid word-for-word translation, but try to find an equivalent in the target language. If not, adapt the translation for the readers.

6. TED. TED is always written as "TED".

7. Titles of works. The translator should check the official translation for books, movies, magazines, and poems in the target language. If not, should not translate.

8. Proper nouns. For a language that does not use the Latin alphabet, the name is transliterated. The same for places if there is no equivalent in the target language.

9. Punctuation. Linguists should use the target language's native punctuation.

10. Character sets. The translator should use the standard Unicode characters.

11. Units of measurement. Units of measurement need to be converted more understandable to readers in the target language [13].

Consequently, it can be seen that the translation requirements for TED are quality requirements generalized to the translation industry. Rules for the adequate translation of content, preservation of formatting, localization of time format, units of measure, etc. should be adhered to. But at the same time, a special feature of translation for TED is the presence of an additional factor. Creating and translating subtitles is extra work for the volunteer because not only the rules regarding translation but also the subtitling rules need to be taken into account. This includes the number of characters and lines, line length, etc.

But even when following all the guidelines, linguists face some difficulties when working with subtitles. For transcribers, it can be difficult to create a transcript for video when words cannot be distinguished.

This causes inconsistencies in the transcript – misspelling. Such mistakes cause difficulties during the translation process itself. But such mistakes are corrected at the translation and editing stage.

Also, a difficult point when creating a transcript can be the specifics of the speaker's original speech: junk words (for example: "well", "oh", "so", "right", etc.), pronunciation, voice volume, accent, etc.

Wordplay also belongs to such features. Wordplay has always been one of the difficulties in translation since it is not always possible to convey the meaning of a saying in another language. For example, Sir Ken Robinson in his speech "Do schools kill creativity" [11] tells the story about his son and his role in the Nativity play. That was a part where three kings

come in bearing gifts, gold, frankincense, and myrrh. But they got the words mixed up: "The three four-year-boys came in and the first boy said, "I bring you gold". And the second boy said, "I bring you myrrh". And the third boy said, "Frank sent this" [11].

The phrase "Frank sent this" shows in this case the boy's misunderstanding of the meaning of the word "frankincense". In Ukrainian, the linguist translated this phrase literally: "Френк передав ось це". Unfortunately, the wordplay broke down during the translating of this part of the speech.

Thus, it can be seen that during translation, such features of the speaker's speech can cause difficulties and, as a result, in some cases, inconsistencies.

Conclusions from this study and prospects. Crowdsourcing as a type of translation is very popular and widely used when it comes to large-scale projects, making translating materials into several languages. It allows more volunteer translators to be involved, thus increasing the target audience of the resource. This widens the range of topics and attracts more experts on certain topics. We believe that as the range of topics and the number of translators increases, so will the number of instruments and tools for quality monitoring. Therefore, the prospect of future research is to analyze translation quality evaluation criteria for large-scale projects where crowdsourcing is used.

REFERENCES:

1. Бахов І.С. Новаторські підходи до оцінки якості перекладу в Facebook. *Science and Education a New Dimension. Philology.* 2018. Vol.VI (51), no. 176. URL: https://www.researchgate.net/profile/Ivan-Bakhov/publication/328374161_Innovative_approaches_to_assessing_quality_of_translation_on_Facebook/links/5f1f477245851515ef502647/Innovative-approaches-to-assessing-quality-of-translation-on-Facebook.pdf (access date: 10.10.2021).

2. Поворознюк Р.В. Інтерлінгвальний вимір онлайн-платформи TED Talks. *Одеський лінгвістичний вісник.* № 10. Том 2. 2017. С. 103–106.

3. Coursera. URL: https://www.coursera.org/ (access date: 10.10.2021).

4. GlobalVoices. URL: https://globalvoices.org/ (access date: 10.10.2021).

5. Howe J. The Rise of Crowdsourcing. *Wired Magazine.* 2006. Vol. 14, no. 6. P. 1–4. URL: https://sistemas-humano-computacionais.wdfiles.com/local--files/capitulo%3Aredes-sociais/Howe_The_Rise_of_Crowdsourcing.pdf (access date: 05.09.2021).

6. Khan Academy. URL: https://www.khanacademy.org/ (access date: 10.10.2021).

7. Losse, K. Achieving Quality in a Crowd-sourced Translation Environment. Keynote Presentation at the 13th Localisation Research Conference Localisation4 All, Marino Institute of Education. Dublin, 2008, October 2-3.

8. O'Hagan M. Evolution of User-generated Translation: Fansubs, Translation Hackings and Crowdsourcing. *The Journal of Internationalization and Localization.* 2009. Vol. 1. pp. 94–121. URL: https://www.jbe-platform.com/docserver/fulltext/jial.1.04hag.pdf?expires=1601823405&id=id&accname=guest&checksum=30FC9EF90A0F0CD9EBFC81DEE32D2343 (access date: 05.09.2021).

9. O'Hagan, Minako. Community Translation: Translation as a social activity and its possible consequences in the advent of Web 2.0 and beyond. *Linguistica Antverpiensia, New Series–Themes in Translation Studies.* 2011. Vol. 10. URL: https://lans-tts.uantwerpen.be/index.php/LANS-TTS/article/view/275/173 (access date: 28.10.2021).

10. O'Hagan, Minako. "Translations| Massively Open Translation: Unpacking the Relationship Between Technology and Translation in the 21st Century". *International Journal of Communication.* 2016. Vol. 10. URL: https://ijoc.org/index.php/ijoc/article/view/3507/1572 (access date: 28.10.2021).

11. Robinson K. Do schools kill creativity? TED Talks. 2006. URL: https://www.ted.com/talks/sir_ken_robinson_do_schools_kill_creativity (access date: 25.10.2021).

12. Taibi M. Ozolins U. Community Translation. New York: *Bloomsbury Academic.* 2016. 200 p. URL: https://books.google.com.ua/books?hl=uk&lr=&id=gzM0CwAAQBAJ&oi=fnd&pg=PP1&dq=Community+Translation&ots=G7jxAMQmC7&sig=wIWksgiJTr3wdCp5zVieyx-edfo&redir_esc=y#v=onepage&q=Community%20Translation&f=false (access date: 10.10.2021).

13. TED. Guidelines. URL: https://www.ted.com/participate/translate/guidelines#h2--subtitling (access date: 10.10.2021).

14. TED. Subtitling tips. URL: https://www.ted.com/participate/translate/subtitling-tips (access date: 05.09.2021).

15. TED. URL: https://www.ted.com/ (access date: 05.09.2021).

16. TED. Volunteer Roles. URL: https://www.ted.com/participate/translate/volunteer-roles (access date: 05.09.2021).

17. The OTP Learning Series 02: User roles and workflow. URL: https://www.youtube.com/watch?v=s2J6bb1Uh0U&list=PLuvL0OYxuPwxQbdq4W7TCQ7TBnW39cDRC (access date: 05.09.2021).

18. Wikipedia. URL: https://en.wikipedia.org/wiki/Main_Page (access date: 10.10.2021).