

**WEBSITE LOCALIZATION OF KHAN ACADEMY ARMENIA AND ITS IMPACT ON
ENGLISH INTO ARMENIAN TRANSLATION INDUSTRY**

By

Gayane Aramyan

Presented to the

Department of English & Communications

In partial Fulfillment of the

Requirements for the Degree of Bachelor of Arts

American University of Armenia

Yerevan Armenia

May 14, 2019

Abstract

This research is an examination of practices of localization of Khan Academy educational website from the English language into Armenian. This study addresses the best way to translate, to determine better processes of the software application development. The conducted case study is using Khan Academy and Crowdin website localization platforms. The practical application goes through localization procedures of Khan Academy Armenia and evaluating its effectiveness. At the same time, the paper measures the impact of localization on the translation industry and in which ways technology brings completely new approaches in translation and website localization. The examination concludes that overall Khan Academy localization processes would be more productive and cost-effective if the company built its customized platform.

Table of Content

Introduction.....	4
Literature Review	7
Description of the Project	14
Conclusions and Limitations.....	21
Appendices.....	23
Bibliography	29

Introduction

In this case study, I conducted a research paper which examines the localization processes of Khan Academy from English into Armenian. As project manager of Khan Academy Armenia, in the past six months I was evaluating both technical and management procedures of localization processes.

Khan Academy is a non-profit organization which provides free online educational resources for everyone, anywhere. It offers online practice exercises, video tutorials and individualized learning platform where students can organize their learning processes. They provide online content from subjects like mathematics, science, arts and humanities, and more. The platform is designed for students, parents, and teachers (Khan Academy, n.d.). The Khan Academy organization was founded by Salman Khan in 2008 with the mission to provide a world-class education to everyone, anywhere (Noer, 2012). They initially had the mission to make the platform accessible in any language, and today more than 36 languages localize the website content.

From 2017, Ayb Educational Foundation joined to the localization project of Khan Academy Armenia. It made accessible the content of Khan Academy from mathematics, humanities, finance and computing subjects, translating and localizing the video, exercise, and article content. As Khan Academy localization is a non-profit project, it gives access to anyone who wants to make accessible the educational content in its local language. Initially, the team of Khan Academy was publishing video tutorial content on Youtube and at the end of 2017, released the demo version of the website (Ayb Educational Foundation, n.d.). This research paper examines the localization processes of Khan Academy Armenia from January 2019 to April 2019 providing in-depth analysis of both internationalization and localization processes of international Khan Academy.

Armenian team of Khan Academy localized more than 2500 videos, and more than 1000 tutorials and articles (Khan Academy International, n.d.). The online videos are both available on Youtube and Khan Academy website. The Khan Academy Armenia has two separate teams for video

and exercise, article translation. Exercises and articles are translated within the localization platform of Crowdin.org and video localization with the help of Camtasia and other technical tools. Each team has subject specialists, translators and managers who make sure that the content is appropriately localized. January to May 2019 the team extended its subjects, translating more than 1,3 million words and 1000 localized videos. It was also planned for the platform to become a LIVE site, releasing a mobile version of Khan Academy Armenia. However, localizing a large amount of content within a short period could be challenging, and throughout management processes, the team modified, reflected, reconsidered specific translation strategies. The new approaches are also coming from extensive research of the localization field and its organizational differences from translation processes. It discusses how it is essential for the Armenian language to conduct more extensive localization projects which can bring new platforms for the local users, and more importantly develop and modernize the content of Armenian.

The localization industry emerged at the end of the 20th century when computer software companies wanted to reach the global market and make accessible the platform in different languages. However, localization not only changed the technological perspectives of computer software but also changed the industry of translation in various aspects. Synthesis of linguistics and technology brought new translation approaches, and today it is apparent that technical developments may also shape the contextual aspects of languages. In past five months, the localization team of Khan Academy Armenia re-evaluated the effectiveness of translation procedures, compared its own experience with international organizations and found out different gaps related to localization of software from English into Armenian. The updated model of the translation will allow not only to localize the content of Khan Academy more effectively but also suggest a model for implementation in any localization project.

Localization, in general, has a complex work structure. Before the project, localization platform provides translators, editors, content reviewers different online tools, automation solutions, where they can efficiently optimize their work processes. They must make sure that the entire

environment of the platform functions smoothly and team members work sufficient enough to deliver a high-quality product. At the same time, localization highly depends on the product development team, and the localization management team must make sure that the platform technically supports the work of the localization team.

In the upcoming 30 years, the technology industry will be more focused on the development of software programs where localization will play a huge role in there (Franklin, 2018). It is one of the fastest growing industries in the world, and many localizing countries need to take this into account and consistently work on content localization and product development processes, where they have not only new locale software products but also have opportunity to follow the dynamics of modification, the evolution of the language.

After the emergence of localization platforms in the 1980s, the industry in the last 40 years rapidly developed and brought more efficient technical and content solutions. Within a complex environment, companies designed unified translation tools which could serve for internalization purposes. Even though many large software companies, like Microsoft, Oracle were executing localization processes internally, after while noticing apparent inefficiency of the approach, they shifted their strategy, outsourcing translation to have a higher quality of localized content. With that, they also developed globalization management systems (GMS), in which managers create a collaborative environment between developers and translators where they evaluate the efficiency of their work processes.

Today in Armenia translation from English into Armenian is the fastest growing industry. However, software localization is not developing efficiently, opposed in many countries where it is one of the fastest growing industry. Localization from English into Armenian, in general, has a low quality and this is coming from a couple of reasons. First, localization is a teamwork of content specialists, translators, editors and project managers. It needs to be a long-term, continuous project because software platforms are constantly changing, and the localization team continually needs to correct, revise the content. Khan Academy Armenia is one of the localization projects which went

through these procedures and in the past five months actively evaluated the effectiveness of the technical abilities of the localization platform.

Literature Review

Today the translation industry widely discusses perspectives of localization and their impact on the development of the translation field. However, it is essential to note that designed mechanisms and management of localization are not newly designed things, and it has been in the industry for more than three decades. The Localization Industry Standards Association (LISA) provides terminologies which give a clear understanding of how generally the field operates. There are three essential terms: internationalization, localization and translation. Internationalization defines as, “The process of generalizing a product so that it can handle multiple languages and cultural conventions without the need for re-design. Internationalization takes place at the level of program design and document development” (Esselink, 2000).

Essentially, internationalization is related to the first stage development processes of the platform, in literature the short version “i18n” is also used, because there are 18 letters within internationalization word. Even though the definition mentions that the platform must be designed in a way that it would not need to be redesigned, however in many cases it shows the opposite, where the product needs constant changes and revisions. Linguists, translators, managers and programmers in localization projects need to establish collaborative work environment for having high-quality platform. Usually, localization comes after processes of internationalization, where translators and editors are actively involved in the processes. “Localization is taking a product and making it linguistically and culturally appropriate to the target locale where it will be used and sold” (Esselink, 2000). Internationalization and localization are interrelated processes, and as platforms are improving and becoming more complex. These systems require participation of different professionals, such as project managers, software engineers, quality insurance specialists, desktop publishers, translators, editors, and reviewers. So, it is not only necessary how technologically equipped professionals are,

but also how they work and collaborate, for better-localized product outcome. Globalization and Localization Association defines localization as, “Industry which helps companies, governments, and non-profit organizations create global content for audiences for any culture who speak any language” (GALA, 2012).

In localization platforms, translation documents divide in strings, and contextual information of the string is always included. Providing information about user interface string context can be one of the challenging issues. However, a lack of meaning, in any case, gives a poorly translated content. Localization team not only translates but also conveys the right information across linguistic and cultural barriers (De la Cova, 2016).

In some cases, local computer-aided translation was not welcomed innovation, especially in post-Soviet Union countries. The issue was not related to the problem of the lack of standardized translation. During the twenty century, Soviet Union countries started to actively translate art and literature texts. In that period linguists developed an intuitive understanding of translation. Machine translation (MT) was not seriously considered and even considered an illusion of accuracy (Malmkjær, 2011). The criticism over MT mitigated when dominant language, in this case, English language, implemented norms of translation, such as contextual simplification in a software environment. As machine translation does not deal with post-editing processes, it is not dealing with how translators do their work at the cognitive level. However, it can estimate the user’s interaction with MT tools in processes of cognitive studies (Schwieter, 2017).

Computer-aided translation mechanism can be complicated enough not only for translators and linguists but also for engineers. In the early stages of localization development processes, software companies tried to develop professionals who were both linguists and programmers. However, programmers are usually deeply merged in technical skills, but after while companies like Microsoft, Oracle and Google started to provide detailed guidelines on how to internationalize their product properly (De la Cova, 2016). In the last two decades, this approach proved that the quality of product largely depends on the implementation strategies and how seriously software companies want

to reach the global market. For instance, if the engineering team is not adapting target marketing requirements, in the end, it may cause problems later on. There are several essential requirements in localization from which it is worth to mention: characters sets, space and word breaking, date format, calendar, number format, periods, commas, date formats (Chan, 2015). The absence of these requirements can also come from a lack of understanding of how the software functions. It is highly recommended that in localization those engineers who will participate in development processes already familiar with existing platform architecture. Otherwise, a software company may use more than required human resources and get low quality internationalized product.

Translation Memory (TM) is essential for any translation platform. Through the accumulation of translated data, it can suggest or automatically translate the content. However, linguists argue that these tools may have an adverse effect on translators, especially for beginners. They can become too much dependent on the tool, and it can demotivate them to use their other cognitive skills (LeBlanc, 2013). Translators need to remember that in any case, they must work as attentively as in traditional translation processes because programmed tools can cause technical issues and without attentive examination, they may do an unclear translation.

In large amount of accumulated data TM can be a useful tool, however terminology or glossary usage within the platform also a crucial aspect in localization. Lack of terminologies can create inconsistent translated content. Even though editors and content reviewers go through all the translation, however, they may not have a consistent memory, and in a considerable amount of translated data, they may also approve inconsistent translation content. As terminologies may also contain a vast amount of data, they may also need to be categorized. The corpus-based approach is one of the solutions, where it allows us to identify subject-specific terms and will enable translators to quickly choose the most convenient word for the context (Bowker, 2002).

Even though today there is no specific profession as translator-engineer, however engineers went deep into the mechanism of linguistics, to come up with unified technical solutions. Back in the

early development processes, internalization would play a key role in product design processes, especially developers who were writing codes in Java language.

“The designers of Java realized early on that support for global software development would be an important language feature storing all characters as 16 bits encoded in Unicode was a key Java 1.0 design decision to support internationalization” (Deitsch, 2001). In Computer Science a bit or binary digit is the smallest unit which has either 0 or 1 unit, which is mainly used for storing data. Later on, in the upcoming Java version, it would have more advanced features, but this unified solution allowed for any language to be accessible in software platforms.

Usually, translation takes place in the environment of two languages. In some cases where the translator knows more than two languages, may also check and compare with other languages. However, they work with limited data, and their comparison is not as unified as in localization case. Localization platforms primarily are designed in globalization point of view where all translations are available and unified which is more productive and efficient to solve a lot of technical problems. Cross-language can be primarily a powerful tool for question and answer systems, as translators usually work with fragments of texts and different language translators can be more helpful to each other (Jiménez-Crespo, 2011).

Even though machine translation innovation fundamentally changed the industry of translation, however in its initial stages linguists for decades were skeptical about its potential. Many professionals thought that the difference between machine translation and post-editing stages might not be that different, then starting from scratches. This solution could be a problem for translators because internationalization companies were not presenting their approaches and solutions related to machine translation mechanisms. One of the most persuasive arguments was the accumulation of errors in translation memories (Pym, 2014). However, the criticism related to translation memory and machine translation was not valid, and none of it proved.

In localization processes understanding the usability of the product can be eliminated in the processes. However, it plays an essential role in evaluating the quality of localization and translation

and it is coming from key term, as user experience (UX) which is defined as, “A person’s perceptions and responses that result from the use or anticipated use of a product system or service” (Suojanen, 2015). UX can be evaluated both examining the dynamics of the software usage data and also randomly testing out the product and through survey questions asking more specific questions about the platform.

The localization industry in the last three decades rapidly developed, the however overall sector is an immature stage, especially in preparation and education of future translators. The link between research and reality may not always be evident. Engineers may develop a powerful platform and linguistics may express their concerns about the software; however, without designing specific implementation of researches, they will not necessarily have an impact on the industry. Some part of the problems is coming from the nature of the advancement of technologies and their prospective effect on web localization (Jiménez-Crespo, 2013). And for sure upcoming research will be a result of new technological solutions.

Today globalization allowed translators to introduce out-of-ordinary localization concept where through innovations and tools to provide much more open standards, like OASIS-based standard or creative commons (Cronin, 2013). In a vast amount of accumulation of English based data on the internet to be able to translate in other languages, initiatives like community-based platforms or crowdsourcing can be a great solution. For instance, google translate is primarily community-based translation software and through many volunteer contributions, it continually becomes more useful.

It is tested that translation processes can reduce translation workload about three times (Frederking, 2004) if localization is equipped with proper tools. Still, there is no in-depth investigation for understanding the needs of Machine translation (MT) needs from a user perspective for enhancing the platform. In certain cases, theories can be behind the development processes. For instance, MT evaluation, theories are far more developed than the available MT tools (Wilks, 2010). However, before any platform and software development scientists start to explore the theoretical part

and it may be a better solution if translators, localization teams also actively start to participate in these processes to come up with better solutions.

It is possible to conclude that translation is essentially a localizing practice (Venuti, 2013). Like localization, translation itself is dealing with adaptation of the context to the local language, market, and culture. It creates a discursive environment between languages where they give and receive information, and inappropriate adaptation, it can create more diverse discourse within the local culture.

For having a good quality localized platform, there are two main aspects to be considered during internationalization: high-quality localization management procedures and evaluation of product development processes (Dunne, 2011). During localization management, computer-assisted translation (CAT) can be a great tool both for translators, and at times it is a cost-effective solution for the project. There are also other technical aspects which in any localization projects must be considered, such as user interface text display, product glossary, locale-specific display of calendar, date, time and also graphics files display with texts. These features are usually set up in any software, because different languages have their specific terminology, calendar, or time and localization team must make sure that the same software is as interactive as in dominant or main language. Even though it seems that quality of translation in localization processes might seem an obvious issue, however in its earlier stages it was not a big issue. Mari Taanonen from Aalto University found out in “Targeting language markets through Web localization” research that, “The quality of language seems to have a great impact on online customer experience and linguistic errors undermine the credibility of the service” (Taanonen, 2012). These issues can come from both bug issues and poor translation and both programmers and translators must make sure that the platform displays properly for users.

But in overall project management perspectives, localization team must clearly understand the scope of the localization, understand clearly the culture of a software company and make sure that they appropriately adapt the content into the local language.

In the case study of “Khan Academy Videos in Chinese: A Case Study in OER Revision,” Allen Rao, John Hilton, and Sarah Harper found out that localization can be a helpful tool for revision of Open education resources (Rao, Iii, Harper, 2017). This is especially important because in the case of video localization, original videos may not be revised, and they need to be changed accordingly. It is both not cost effective and hard process and that is why video localization can be a good solution for revision of the English content of Khan Academy. Still, there needs to be a mutual collaboration between content creators and translators, to have intended productive results of content revision.

In localization one of the challenging issues is loss of translation. Translators may decide to keep or remove certain content, depending on how hard it is to localize in certain language. Miguel A. Jiménez-Crespo in his “‘Loss’ or ‘lost’ in translation” case study found out that due to the hypertext nature of website any localization processes lead to a certain content loss. Eventually, this problem leads to lower interactivity than the original website content (Jiménez-Crespo, 2012). In the complex environment of localization with the improper management structure and technical abilities interactivity issue can become a more serious problem. However, these studies need to take further investigations, because non-profit website localization content is not examined deeply enough.

Methodology

The case study was a direct exploration of the examined project; I am the project manager and advocate of Khan Academy Armenia, directly researched the case study by examining the internationalization tools of Khan Academy, measuring their effectiveness and also suggesting a new localization management mechanism. Primarily, most of the study I investigated in the Crowdin.org software localization platform, particularly Khan Academy software localization project with Crowdin.org platform, “Crowdin is a localization management platform for mobile, web and desktop apps, games and help desks. Their platform provides centralized data from where it is possible to examine various translation patterns” (Translation and Localization Management Software, n.d.).

Besides Crowding.org, I examined other localization platforms that Khan Academy engineers built and compared all the platforms and what type of features, services they provide, tools like online translation services, translation memory (TM), machine translation (MT), terminology and glossary solutions. During management processes, I also examined the video localization criteria and with the collaboration of content specialists designed more efficient solutions. New models were designed based on the previous practices of video localization teams and collaboration, experience exchange with translators of Khan Academy.

Through direct involvement, I actively participated and examined translated data of Khan Academy from English into Armenian. In the past six months in the project there were involved more than 20 translators, videographers, and content specialist from Khan Academy Armenia. With more than 3 million words translation of Khan Academy from the English language into Armenian, it was possible to examine different translation patterns. Besides translation patterns, it was also possible to explore the technical abilities of the platform and being able to contact with the engineering team of the Crowdin. This way it was possible to examine the effectiveness of the localization solutions the company provides. With that, during the research, the paper explored the other additional internationalization platforms that Khan Academy engineering built.

To evaluate the quality of translation both translation tools and localization management systems of the project were examined. Through direct examination and participation, it was possible to reflect and reevaluate the effectiveness of localization management processes and to identify a clear distinction between technical development perspectives of the platform and localization management effectiveness.

Description of the Project

In this research, I studied and examined the efficiency of localization that Khan Academy team offers and as part of the international team from Armenia what kind of new localization management solutions I prefer. In the study, I separated software tools from management solutions,

because technical solutions are usually designed before any translation processes take place. In this case, I can only evaluate the effectiveness of TM, MT, terminology and other provided tools. We need to take into account that any technical change is a long-term process and may not be changed as quickly as localization management processes. However, these two described processes are interconnected, the efficiency of localization management depends on the technical abilities of the software, and it must always reflect the needs of the translation projects.

Crowdin has conducted more than 79 thousand localization projects, such as Microsoft, Reddit, Code.org, MIT OpenCourseWare and also Khan Academy is their largest public localization projects (Translation and Localization Management Software, n.d.). Through source file management both managers, translators and editors can organize and plan ahead their translation tasks. Usually the structure of translation files is grouped before the localization processes, that is why it may cause additional problems for managers, because during the process they may want to change source files structure. But even by changing certain structures, still the localizing company may have concerns about the efficiency of file management. Khan Academy started to build additional platforms which offered better file management system.

Currently, Khan Academy Armenia team localized more than 3 million words out of 17.5 million words content of the platform. Crowdin considers the base platform where localization is being conducted (See more in Appendices, Figure 1). It is essential for any software company to consider localization as its development key component. Through instant communication, information exchange of the platform on daily basis can enhance and bring high-quality content to a user.

The platform of Crowdin's Khan Academy content is divided into source string. A contributor can translate or edit the content, and the string will appear precisely in the same localization of the platform, but in a localized form (See more in Appendices, Figure 2). In each file, translators see divided strings. They not only translate the specific content but also have access to entire translation history, terminology library, which can help them to have more consistent, correct localization

content. They both see whether string translation goes into other files automatically, and also who translated or edited before them. They are also able to see what other translation options are available within the entire files and also it suggests machine translation if it has enough accumulated data. And also, through comments localization teams can discuss different issues related to that specific string (See more in Appendices, Figure 3).

These are general tools that every localization platform provides, however from linguistic perspectives not always these tools can serve the localization teams in the most efficient ways. For a better outcome, it is crucial to have active communication between localization product developers and project managers. Also, as any change or development plans of the platform can become long-term, slow process, it is also crucial for project managers to communicate and explain the current structure of the platform and how they must adjust to it.

As localization teams in past 1.5 years have also access to publish their own translated content into their locale, Khan academy headquarters built an additional tool which organizes the files of the website more efficiently and managers can see not only word count of the translation, but also where and how the translation appears (See more in Appendices, Figure 4). Throughout these practices, the management team based on the general practice and research designed guidelines and structures that the organization can use internally.

Before going into translation processes, it is important to take into consideration several technical issues which can hinder further translation procedures. Before any localization processes the platform where managers, translators, and editor can conduct their work, they are relying on a platform which is designed for the international market. In internationalization processes, platform developers try to find efficient solutions for all languages where they can translate the content efficiently, fast, automated and consistent. However, even in internationalization individual languages can be a priority if they plan to use it after the release of the product. And those languages which join much later, they can have more significant issues. It will become the ultimate task to project

managers to have constant communication with engineering team and adequately address the problems.

Localization translation management team of Armenia can be considered an outsourced organization. Structurally any local language team is not directly working with developers. However, they are in active contact with them. In Khan Academy platform there are two types of materials translators are localizing: text and video. Before 2019, the team had enough resources to work with translators, train them, however with a significant amount of translation it is not possible to manage them separately. There were a couple of reasons. First, if there are more than 15 translators in the project, there is a high chance that four to five translators will be replaced by others. In this case, it is even worse to work individually with translators, mainly if they translate for short-term work, individual approach is not cost-effective for the project.

The project managers must make sure that any newcomer gets required guideline. However, they may get the instructions, but not master them properly, that is why it is essential to create a community of translators who help each other to get to know all the tools required. Otherwise, translators can work slower and do poor-quality translations. So, Khan Academy Armenia localization team kept individual meeting sessions with translators but also started to provide different materials and tutorials before the meetings; that way one-on-one meetings would become questions and answer sessions.

Also, a vital aspect of translation is consistency in language, terminology. Most of the tools for localization were built for keeping up with a consistent, high-quality translation. In general, translation is associated with high-quality translators, but in localization case, it is only one component. In localization case team structure is more complicated with more supportive technical tools and more communication is required between specialists.

In the past six months localization reflected the effectiveness of the structure of the team and evaluated that the structure is efficient enough to provide the desired result. The most evident proof of managers was in some cases more than twice translated materials could be reviewed additionally by

managers. It was necessary for managers to intervene in translation processes. However, they were also required to build a consistent structure where workflow goes from one team group to another. In the past six months after evaluation of localization procedures, management team changed overall the processes. (See more in Appendices, Figure 5).

Automated translation is a standard feature of Crowdin that offers to any localization project. If the translator and editor find within the string multiple contexts that the translation will appear, then the function is considered to be the automated translation. However, as within one project, various languages participate, it is not possible to switch off the feature of automatic translation. In general, this solution considers to be efficient; however, for languages which have the case of a noun, adjective, pronoun, numeral or participle, the automated translation method is not a right solution. This feature is especially problematic when the name needs to be localized in some instances, however as it is also automatically being translated in the biography section of the team, then within the mathematics text problems, translators cannot conduct efficient localization (See more in Appendices, Figure 10).

If the automated translation can be used regarding how much translation was accumulated, in machine translation (MT) and translation memory (TM) case accumulated translation data is required. These tools provide a translation suggestion based on the previous data of translation, and it is especially helpful when the localization team wants to reduce the cost of translation and this way, they can detect the identical or similar content where translators apply these tools. In Crowdin, case translators can see machine translation suggestion and automatically use in their translation, in case of Khan Academy localization platforms, it suggests smart translation feature where they can automatically translate more than one string. These two features translators and editors can use at the same time, however as the tool does not provide an error-free outcome, they always need to double check. Sometimes it is debatable whether these tools are helping to reduce the translation cost, because if they also generate wrong translation suggestions, then there is always required to double

check the content. More investigation is needed, and more engineer works to improve the features of MT and TM.

The other automated feature which is absent in localization platforms of Khan Academy is final editing tools. This is not related to translation quality, but minor issues like spacing, capitalization, full stop. Editors can sometimes miss these grammatical mistakes, and machines can be more accurate and at the same time faster than humans. Full stop issue can be debatable because not in all cases machine can be reliable. However, the spacing issue can be a helpful solution. Other automated localizations, like date, calendar, time adapting for each locale also resolved much later in localization of Khan Academy, and unfortunately, Crowdin was not supporting these solutions. However, they had to be set up in the localization platform in its initial stages (See more in Appendices, Figure 11). Even after debugging this issue, the localization management and quality assurance team must make sure that new emerged bugs are quickly resolved by engineers.

Technically localization platforms do not support all these steps and managers must make sure the processes go smoothly between teams. One of the important changes was adding review specialists between translation and editing steps. Reviewers will make sure that subject terminology, context is localized correctly, make a comparison between English, locale subject content and make changes accordingly. With that, editors concentrate on language content editing and through constant communication with reviewers they make sure to have high-quality content. After that, quality assurance specialists will make sure that localized content is properly visible on the website and report any emerged site bugs. After that localization managers publish the content on the website. All described steps in a newer version of steps were generally included, however in some cases it was additional workload for some team members, and after reviewing successful models of localization teams, the team will get more efficient, faster and high-quality content (See more in Appendices, Figure 6).

In the case of video localization, depending on video material managers may decide to either subtitle, dub or create entirely new types of video materials. It is also vital to notice that platform that

is built for website text localization may support video localization in specific cases. In Crowdin's case, it does not provide specialized tools for videos, and there are a couple of reasons. First, all Khan Academy video materials are uploaded in Youtube platform, which means that for localization platforms they only need to connect the link to the platform. Second, video materials usually take a relatively large amount of space compared to text, images, so in this case, data is stored on Youtube eventually so that their website can function correctly.

Before January 2019, Khan Academy Armenia's team was mainly dealing with one type of video materials where the team designed specific procedures related to that. However, in the past five months after expanding subject video content, the team dealt with different formats of videos and using the same approach was not effective enough. Compared to text localization where the individual method was not necessary, in video localization, depending on the format of video it is essential to train and extensively work with videographers. There were three types of video formats where the Armenian team dealt with; video tutorials where videographer can completely remake the entire material, videos where only voice and sound can be localized, and the video material are edited, and videos which can be only subtitled.

The first video material type is subject related tutorials where an individual created the original English video itself. These videos include original voice recording which is explaining topic material on the virtual blackboard. In this case, the topic review specialist works with videographers providing feedback of the video. For videographers as the format of videos are similar, there is a general review system that video review specialist uses (See more in Appendices, Figure 7).

For video feedback, it is essential to provide a particular rubric system, because reviewers may concentrate on specific issues, and missing other important aspects. Reviewers provide feedback for each mentioned category, and also with individual meetings going through all the elements to make sure that videographers understand the strength and weakness of their videos. However, reviewers may provide detailed feedback, but they also need external feedbacks to offer a better content review. This way videographers get new types of comments on their videos and review

specialists update their approaches. However, content reviewers before providing external feedback, must themselves go through comments, as they are not experts in the field.

Depending on video format, localization management may decide how many specialists need to be involved in the processes. One universal model will not work for videos, and the localization team must make sure that they have enough resources and human capacity to localize the video. In any case, the team must make sure that they went through all video materials and identified the necessary support for localizing it. This is an important aspect because, in English videos, production teams have their unique video style, animation, audio recording and more importantly, language style. So, for getting high-quality video, the localization management team must make sure that they answer questions related to audio, text and visual components of the video (See more in Appendices, Figure 8).

Conclusions and Limitations

This study examined the internationalization processes of Khan Academy and Crowdin localization company. Examining other localization examples and comparing their management procedures, the paper concludes that it would be more efficient and cost-effective if Khan Academy built its localization platform without outsourcing it to other organizations. Software companies may outsource their internationalization processes to other organization if they have a relatively small project and they do not have a significant amount of content on their website. Khan Academy website has a vast amount of text, and video data and localization is one of their organizational priorities.

As Khan Academy is a non-profit organization, in initial stages it would not be cost-effective to build their internationalization platforms. One of the evidence is currently their engineer's direct involvement in the platform development processes. As Crowdin is not able to provide sufficient platform support for translators, Khan Academy team additionally built platforms to support localization management and translation processes.

In initial stages of Khan Academy localization, there was a small community of translators, and there was mostly volunteer community contribution. However, after while not only community grew, but also localization project teams emerged in different countries. Today, Khan Academy localization team has more than 10,212 community of translators, 112 localization managers and annually it translates 57,650,000 words, out of which 36,126,000 is approved and eventually published in their locale platform (See more in Appendices, Figure 9). If these teams are not equipped with useful technical tools and proper localization guidelines, they may spend more than required time and resources during translation processes.

The research concludes that for large projects like Khan Academy outsourcing localization project was not an efficient and cost-effective solution, as many software companies experience proved, that no other engineer can better understand the structure of the website than their own ones. Internationalization processes must simultaneously take place during product development stages. This will be both more efficient and convenient for the engineers and localization management team.

Discussed tools like machine translation or translation memory must take into account that they are always becoming better and work more efficient. The limitation also applies to localization work processes where managers must make sure that their approaches are up to date and they use all the available tools that platforms are providing. Also, localization development largely depends on how much active the industry is within the language market. In the case of the Armenian language, it is a newly emerging industry, and in case of rapidly developing new projects, it will take time to adapt the systems of software localization culture within the market.

Appendices

Figure 1: Official Khan Academy translators languages data screenshot from Crowdin

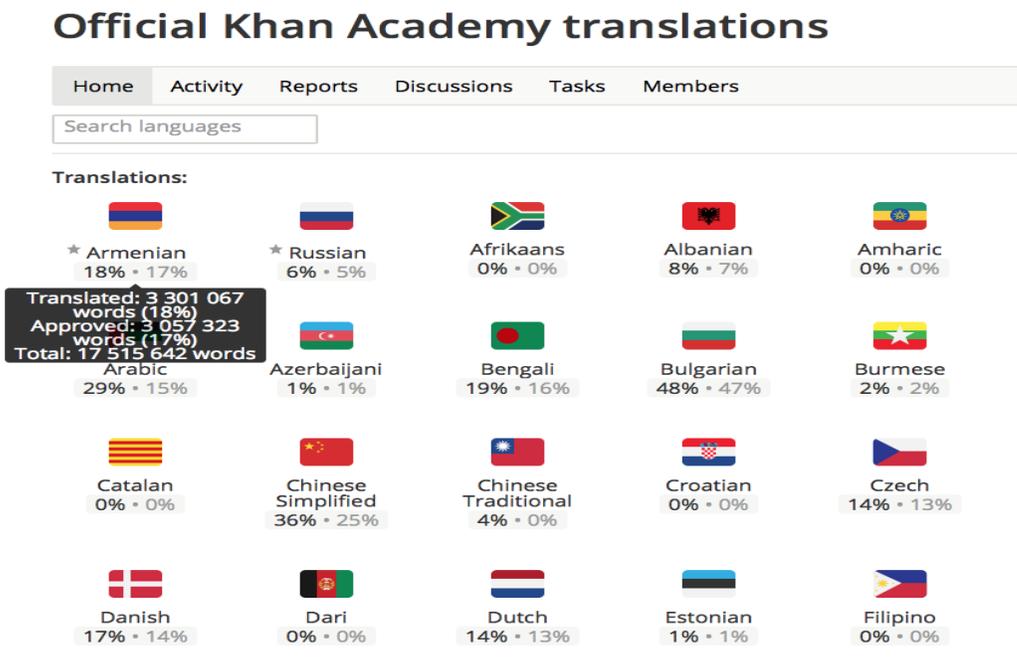


Figure 2: Official Khan Academy Armenia Crowdin XML online files

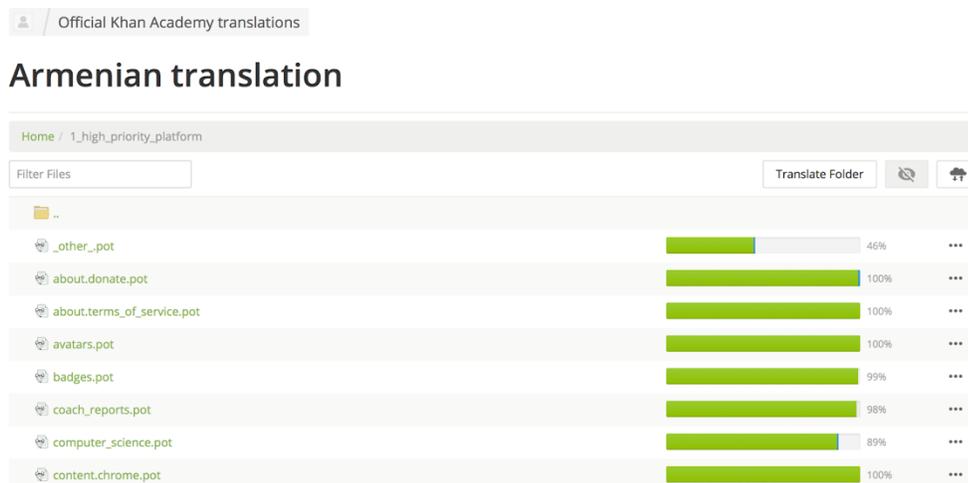


Figure 3: Official Khan Academy Crowding strings translation platform structure

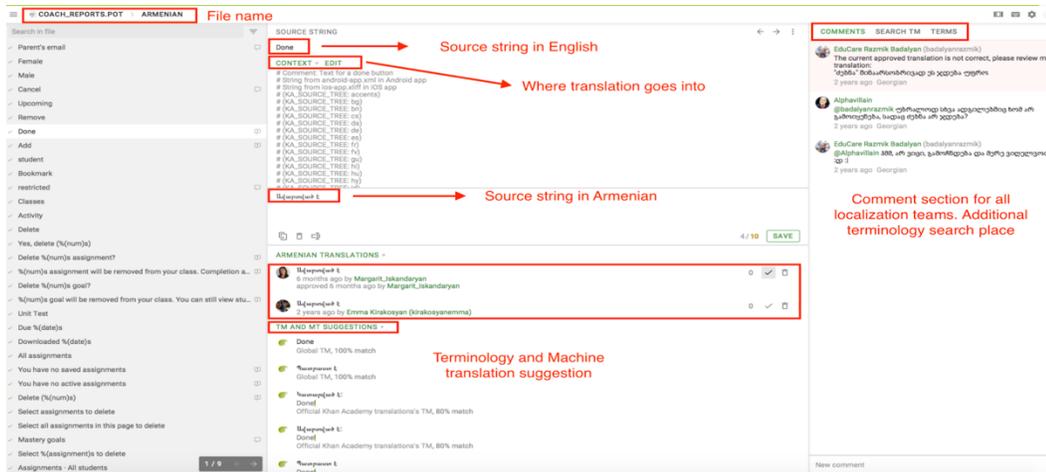


Figure 4: Official Khan Academy platform additionally developed for translators by Khan Academy engineers

Translate into Armenian

Change language

VIDEOS

Subtitles

Captions

OTHER CONTENT

Exercises

Descriptions

Programs

Articles

PLATFORM

Website

Mobile

DATA FRESHNESS

Dubbed videos were last published to the site 10 hours ago.

Subtitles and captions were last fetched from YouTube and uploaded to KA's database 6 days ago.

Approved strings were last fetched from Crowdin and uploaded to KA's database 5 hours ago.

Statistics (% translated, etc.) on the dashboard were last computed from data in KA's database 4 hours ago.

Exercises

Unchanged, revised, added Untranslated

Exercises in Math	TOTAL WORDS	UNTRANSLATED	READY TO APPROVE	PROGRESS	VISIBLE
Total	1644390	1526196	12365		
Count money (U.S.)	1101	17	0		
Estimate to add and subtract multi-digit whole numbers	482	320	0		
Find missing factors (1-digit multiplication)	478	139	0		
Patterns in multiplication tables	1407	860	245		
Tell time to the nearest minute	1819	1785	0		
Time differences (within the hour)	975	963	0		
Telling time word problems (within the hour)	1678	1463	0		
Word problems with mass	1646	1646	0		
Estimate volume (milliliters and liters)	1127	1091	0		
Word problems with volume	2176	2176	0		
Graph data on line plots	2238	2220	9		
Round to nearest 10 or 100 on the number line	1701	1626	0		
Rounding challenge	1323	1258	0		
Compare with multiplication word problems	1447	1065	150		
Equivalent fractions 2 (denominators 10 & 100)	217	160	0		
Common fractions and decimals	558	544	0		
Estimating length (mm, cm, m, km)	843	827	0		
Estimating length (in, ft, yd, and mi)	800	786	0		
Estimating time (seconds, minutes, and hours)	609	609	0		
Convert to smaller units (g and kg)	729	729	0		
Convert to smaller units (mL and L)	729	715	0		
Convert to smaller units (sec, min, & hr)	850	822	0		
Convert money word problems	2098	1998	100		

Figure 5: Official Khan Academy Armenia’s localization team structure from 2017-2018

Old model of Khan Academy Armenia’s localization processes

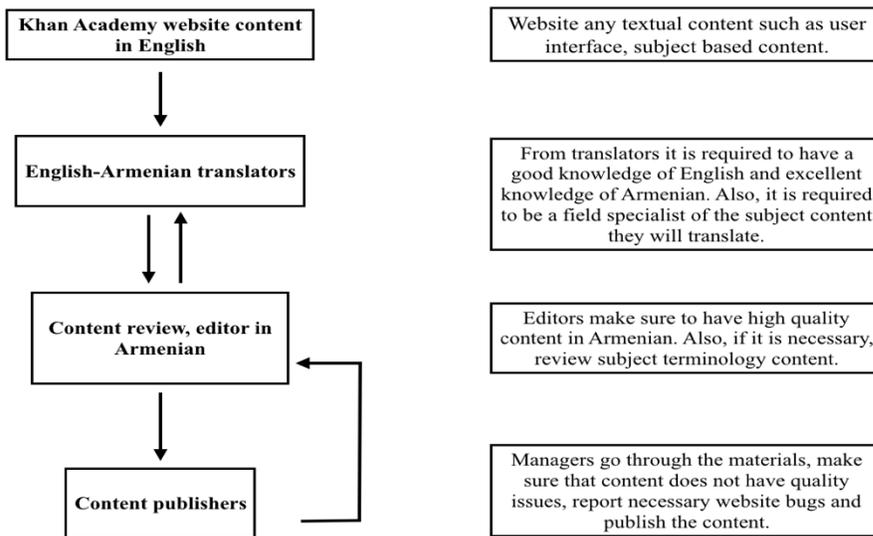


Figure 6: Official Khan Academy Armenia’s localization newly adapted team structure

New model of Khan Academy Armenia’s localization processes

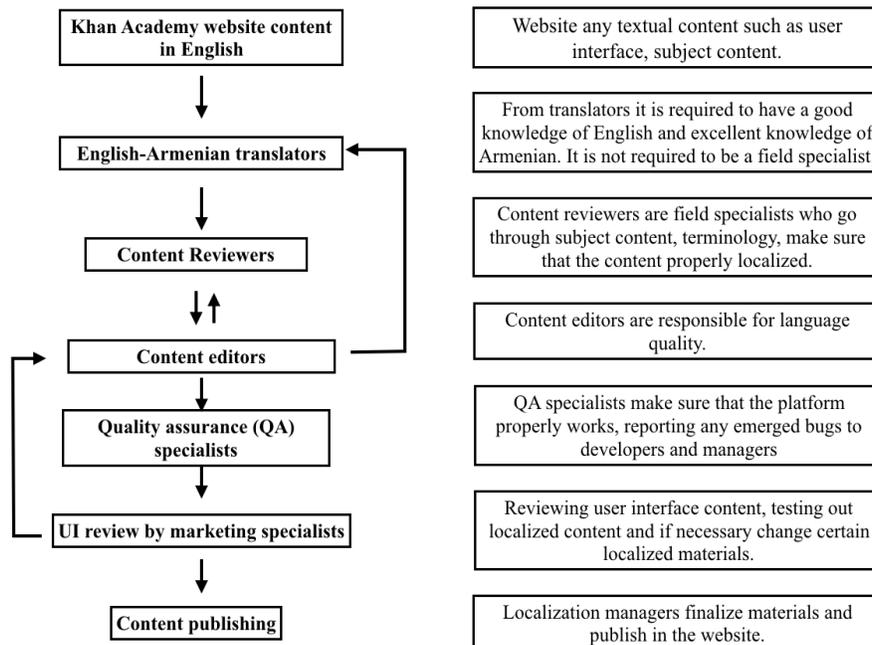


Figure 7: Official Khan Academy's designed feedback mechanism for video tutorial localization

Video quality evaluation - Khan Academy Armenia
Feedback content for subject video tutorials

Content	Understandability of the concepts	The absence of conceptual mistakes	Color coding
Translation and adaptation	Accuracy of terms' translation	Communication of the message, the quality of adoption	Consistency
Speech and diction	Natural and motivating diction	Dynamic diction (rhythmic, vigorous)	The absence of redundant words
Language	Literate writing	Literate speech	
Technical	Recording device problem	The quality of editing	Malfunctioning of tablet, improper color palette, arrangement problem, poor handwriting)

Figure 8: Official Khan Academy's designed feedback mechanism for video localization with different visual formats

Khan Academy videos with different speakers, visual and animation Breaking down the components of the video

1. Audio component	Do interviewers, speakers visible in the video? If yes, then provide how many people are involved and their gender, and approximate age	What is the pace of the video, is it possible to dub properly?	Is there sound effect, music in the video? If yes, is it possible to use the same music or alternative one?
2. Text component	How many words the video has?	Is the language of the video is formal/not formal?	Specific notes to translators, editors regarding to the content of the video
3. Visual component	Does the video contain certain visual, texts, animations? - if yes, is it possible to localize them?	Is there any visual that needs to be excluded during the localization? - if yes, provide details	Is the format of the video outdated? If yes, then when it was made, and is it possible to remake the video material?

In ideal case the video must be dubbed properly and text visuals localized. If there is video animation, visuals which is impossible to localize, then only audio will be dubbed. If the video has certain visual and animations which is not possible to localize and audio part of the video is not possible to dub properly, then the video will be subtitled.

Figure 9: Official Khan Academy's annual translated and approved data

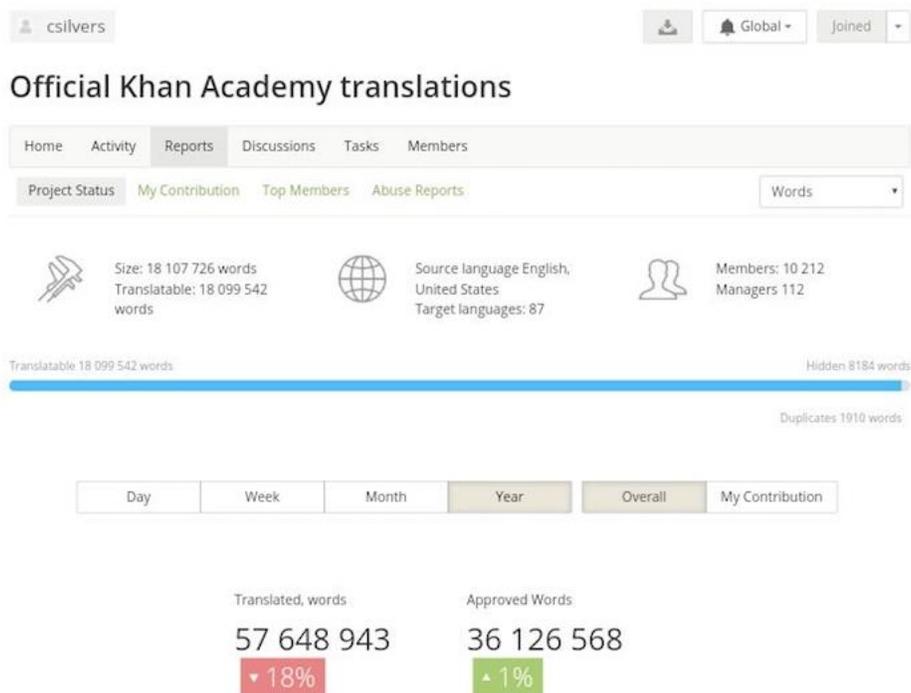


Figure 10: Official Khan Academy's Automated translation example

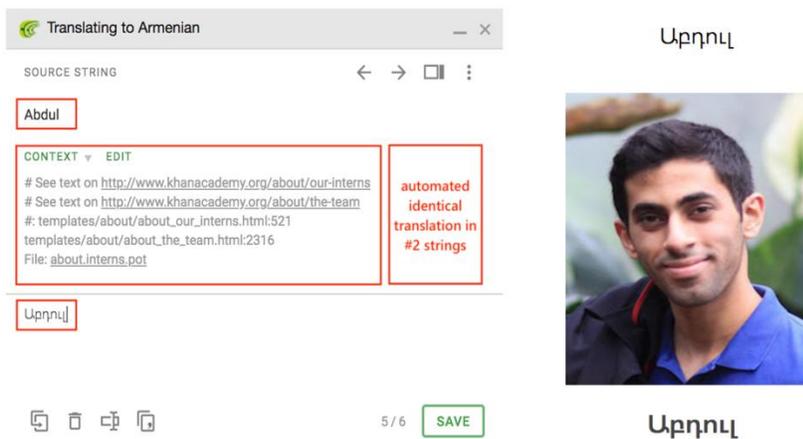
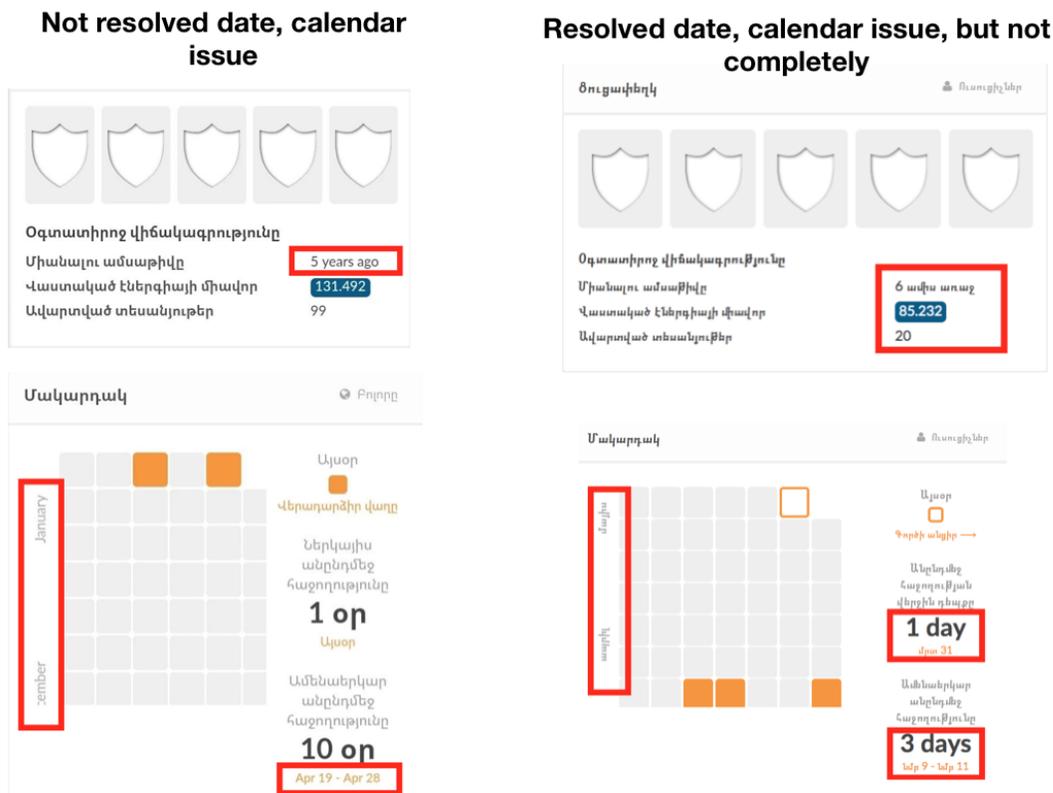


Figure 11: Date, calendar issue in Khan Academy Armenia website



Bibliography

- Bowker, L., & Pearson, J. (2002). *Working with specialized language: A practical guide to using corpora*. London: Routledge.
- Chan, S. (2015). *The Routledge encyclopedia of translation technology*. London: Routledge.
- Cronin, M. (2013). *Translation in the digital age*. London: Routledge.
- Deitsch, A., & Czarnecki, D. (2001). *Java internationalization*. Sebastapol: O'Reilly.
- De la Cova, E. (2016, November). Translation Challenges in the Localization of Web Applications. *Sidebar 27*, 135-166.
- Dunne, K. J., & Dunne, E. S. (2011). *Translation and localization project management: The art of the possible*. Amsterdam: John Benjamins Pub.
- Esselink, B. (2000). *A practical guide to localization (pp. 25-54)*. Amsterdam: John Benjamins.
- Franklin, D. (2018). *Megatech: Technology in 2050*. London: The Economist.
- Frederking, R. E., & Taylor, K. B. (2004). *Machine translation: From real users to research: 6th conference of the Association for Machine Translation in the Americas, AMTA 2002, Washington, DC, USA, September 28 - October 2, 2004: Proceedings*. Berlin: Springer.

Globalization and Localization Association (GALA) (2012). *About the Localization Industry*.

Retrieved from <http://www.gala-global-org/about-localization-industry>.

Jiménez-Crespo, M. A. (January 2012). 'Loss' or 'lost' in translation: A contrastive genre study

of original and localised non-profit US websites. *The Journal of Specialised*

Translation, (17), 136-163. Retrieved from

https://www.jostrans.org/issue17/art_jimenez.pdf.

Jiménez-Crespo, M. A. (2011, January). To adapt or not to adapt in web localization: a

contrastive genre-based study of original and localized legal sections in corporate

websites. *The Journal of Specialised Translation*, 15, 2-27.

Jiménez-Crespo, M. A. (2013). *Translation and web localization*. London: Routledge.

Khan Academy. (n.d.). Retrieved from <https://www.khanacademy.org/>

Khan Academy International. (n.d.). Retrieved from <https://international.khanacademy.org/>

LeBlanc, M. (2013). Translators on translation memory (TM). Results of

an ethnographic study in three translation services and agencies. *Translation and*

Interpreting 5 (2), 1-13.

Malmkjær, K., & Windle, K. (2011). *The Oxford handbook of translation studies*. Oxford:

Oxford University Press.

- Noer, M. (2012, January 19). One Man, One Computer, 10 Million Students: How Khan Academy Is Reinventing Education. *Forbes*. Retrieved from <https://www.forbes.com/sites/michaelnoer/2012/11/02/one-man-one-computer-10-million-students-how-khan-academy-is-reinventing-education/#208eefd444e0>
- Ayb Educational Foundation. (n.d.). Retrieved from <https://foundation.ayb.am/en/projects/khan-academy-armenia>
- Pym, A. (2014). *Exploring translation theories*. London: Routledge.
- Rao, A., Iii, J. H., & Harper, S. (2017). Khan Academy Videos in Chinese: A Case Study in OER Revision. *The International Review of Research in Open and Distributed Learning*, 18(5). doi:10.19173/irrodl.v18i5.3086
- Schwietzer, J. W., & Ferreira, A. (2017). *The handbook of translation and cognition*. Chichester: Wiley-Blackwell.
- Suojanen, T., Tuominen, T., & Koskinen, K. (2015). *User-centered translation (pp.13-29)*. London: Routledge.
- Taanonen, M. (2014). Targeting language markets through Web localization: A case study of a digital cloud service. Retrieved from http://epub.lib.aalto.fi/en/ethesis/pdf/13794/hse_ethesis_13794.pdf
- Translation and Localization Management Software. (n.d.). Retrieved from

<https://crowdin.com/features>

Venuti, L. (2013). *Translation changes everything: Theory and practice*. London: Routledge.

Wilks, Y. A. (2010). *Machine Translation: Its Scope and Limits*. New York: Springer.