
CHAPTER 4

Method for assessing the complexity of an IT project for video game localization

Maksym Yevlanov
Nataliia Vasylytsova
Iryna Panforova
Dmytro Osypchuk

Abstract

The object of the study is the process of project assessment and control.

During the study, the task of assessing the complexity of an IT project for video game localization was solved. In this direction, there has been little research recently conducted in the IT industry, mainly aimed at solving organizational problems of managing relevant IT projects. Research aimed at solving problems of assessing time spent on implementing IT projects for video game localization is mainly of an applied nature. Therefore, there is a constant need to conduct scientific research and process the experience of individual specialists and IT companies acquired during the planning and implementation of IT projects for video game localization.

According to the results of the study, a set of main factors that influence the implementation of a third-level IT project for video game localization was identified, and a system for their classification was proposed. Based on the results of identifying and classifying these factors, a method for assessing the complexity of an IT project for video game localization was developed. This method consists of four stages, which, in turn, are divided into 12 activities. The features of each of the activities of the developed method are described in detail.

Experimental verification of the obtained results was carried out in two stages. At the first stage, the developed method was used to assess the duration of the Ukrainian localization of the video game "Don't Starve Together". The predicted assessment of the translation duration of this game was 32 hours 57 minutes. The actual translation duration was 26 hours 26 minutes. The predicted translation duration exceeded the actual one by 24.7%. At the second stage, the effectiveness of using the localization characteristics obtained during the assessment of previous

IT projects was tested in the project on the Ukrainian localization of the game "Sid Meier's Civilization® VI". The predicted translation duration was 11 hours 11 minutes. The actual translation duration was 9 hours 57 minutes (the difference with the predicted value was 12.1%).

Additionally, the assessment results obtained using the developed method were compared with the assessment results obtained using the common methodology for assessing time spent on software localization. The results showed that using the developed method allows for the most accurate estimates of time spent on third-level IT projects for video game localization.

Keywords

Localization, assessment, IT project, translation unit, video game.

4.1 Introduction

Developers of modern IT products are increasingly focused on consumers from the global, rather than local market. However, the adaptation of IT software products to local markets (localization) remains a necessity caused by the prevalence of languages, cultural diversity and the specifics of local legislation. One of the most complex types of such IT products, from the point of view of localization, are computer games (or video games). The term "computer game" should be understood as a separate class of application software that involves the interaction of the user (player) with other players or a computer in order to achieve a goal set by the rules of the game [1].

Video game localization is a relatively new area of activity in the IT industry. Researchers have recognized that since the 2000s. IT companies have begun to apply a new approach known as "deep localization" [2, 3]. This approach requires bringing the game closer to the consumer in each individual localization (everything that does not contradict the game world itself and can facilitate player immersion can be revised and adapted in accordance with what is considered a more successful local influence) [3]. Therefore, modern research suggests dividing the global problem of "deep localization" into two separate aspects [2, 4, 5]:

- "linguistic" localization (which mainly concerns the translation process);
- "product localization" (which consists in adapting a product or service to the culture and language of customers in a specific target market [4, 5], even deeply changing the relevant structural and artistic aspects of the product).

IT projects for video game localization also differ significantly from other IT projects for the creation, modification or development of IT products. Among the

differences of IT projects for video game localization, it is worth noting the special requirements that are imposed on the planning of such projects. In particular, during planning, an important task is to assess the complexity of video game localization. Such an assessment requires a deep understanding of the various factors that affect the successful implementation of relevant IT projects. An objective and accurate assessment of the complexity of video game localization allows for increased accuracy in planning time and personnel needs for an IT project, and also affects the accuracy of assessing financial costs for this IT project.

The current state of research in the field of software localization for IT products and video games, as a separate class of such products, demonstrates some progress in the development of methods and tools for assessing the complexity of IT projects. Separate works by such leading institutions and organizations as the Localization Industry Standards Association (1990–2011), the International Game Developers Association (IGDA), and some localization service providers are devoted to solving the issues of effective management and assessment of localization projects. However, the problem of increasing the accuracy of assessing the complexity of IT projects for video game localization has not been fully resolved, in particular, due to the growth of globalization processes and the complexity of video games. At the same time, given the high requirements for translation quality and technical adaptation, correct assessment of complexity becomes critically important for the successful management of modern IT projects. Accurate forecasting of localization operations is critically important, since errors in assessment can lead to underassessment or overassessment of resources (team, budget, etc.) or time required for project implementation. These errors, in turn, negatively affect project planning processes and end-user satisfaction. Therefore, research on the development of new and improvement of existing methods for assessing the complexity of IT projects for video game localization is relevant from both theoretical and applied points of view.

4.2 Analysis of current research in the field of software and video game localization

In general, the number of studies in the field of software and video game localization should be recognized as insufficient. Thus, in [2] it was recognized that for the period 1981–2021 in well-known journals, conferences and seminars indexed in the three main scientific search engines (Scopus, WoS and Google Scholar), only 45 studies were published on this topic. A similar conclusion was made in [6]: out of 468 researched scientific articles devoted to the processes of game development,

localization was considered in only a few of these articles. At the same time, attention was mainly paid to the following issues [6]:

- research on regional aspects of game localization;
- research on individual concepts related to game localization.

An important problem was and remains the problem of effective organization of localization work. An equally important issue is the possibility of connections between localization and other work of an IT project on video game development. Unfortunately, in the early 2010s, the relationship between localization and software development processes was largely ignored in research. In practice, localization was often separated from development [4]. In 2013, [7] proposed an extended video game value chain diagram that included a game localizer (Fig. 4.1).

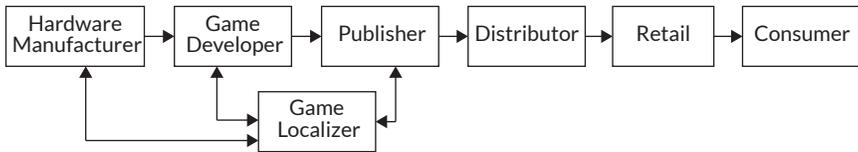


Fig. 4.1 Diagram of the extended video game value chain
Source: [7]

As can be seen from this diagram, the game localizer function is excluded from the main chain of the IT company. In the process of localizing a game, the localizer could interact with the publisher, the game developer, and the hardware manufacturer [7].

The reasons for this separation are the perception of localization as a labor-intensive and expensive activity, often performed by external partners of IT software development companies. Such external partners have limited access to developers and rely on documentation and formalized processes. This contrasts with the minimal documentation and on-site presence cultivated in Agile IT project management methods [8].

This situation has hardly improved in the first half of the 2020s. In [9] it is noted that as of 2023, there are almost no studies that analyze localization from the perspective of video game development. According to the results of a survey of developers of video games available in Spanish on Steam, in [9] it was found that the choice of translator is influenced by the size and experience of the development company. The earlier the localization is planned, the greater the likelihood that the game will be translated by a professional who cannot play the game before its translation. In general, from the results of the processing of the surveys presented in [9], it is

possible to conclude that it is desirable to consider the localization of a video game as a subproject or a separate IT project, which is carried out in parallel with the IT project for the development of the video game. However, the models and methods of managing an IT project for the localization of a video game are formed exclusively on the basis of the private experience of individual specialists and IT companies and have hardly been studied by scientists.

This conclusion is confirmed by the results of field research by a specialist in the field of video game localization M. Toftedahl. Based on the results of this research, published in [10], he proposed to present the organizational structure of an IT company for the development of games in the indie sector in the form of a diagram shown in Fig. 4.2.

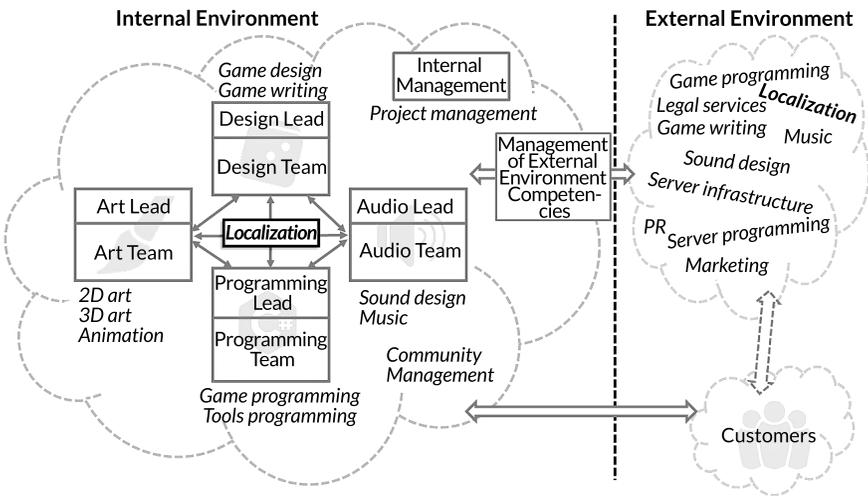


Fig. 4.2 Structure of an indie game development company based on field research in Sweden and China
Source: [10]

As can be seen from the diagram in Fig. 4.2, game localization is a work or separate IT project that is performed both by the IT development company and by the external (relative to the IT development company) environment. The interaction of the IT development company with such external projects in [10] is proposed to be organized through the management of the competencies of the performers of these external projects. However, in the studied cases of video game localization, the results showed that localization is a more demanding IT project than initially planned.

Even if the support of competencies from the external environment is used, management and development should also largely take place in the internal environment of the IT development company [10].

In general, based on the results of the research conducted in [10], the following general conclusions were formulated:

- game localization is one of the important development tasks that must be planned in a timely manner;
- game localization is a task which complexity is underestimated;
- localization is not the main (critical) element of the IT project for developing an indie video game, it is implemented as needed;
- in the context of indie game development, game localization is synonymous with translation, more complex aspects of localization related to game mechanics and game design (the so-called culturalization of the gameplay) require additional research.

The factors identified as contributing to success in game localization in [10] are:

- cultural sensitivity;
- effective communication and collaboration;
- use of localization tools and technologies.

The author of [10] believes that these factors are likely to be relevant to game localization efforts in different contexts and game genres.

Among the problems faced by game localization professionals, [10] specifically mentions:

- tight localization deadlines;
- the need for effective localization project management.

These problems, according to the author of [10], are likely common to the entire industry.

The main tools used in the process of video game localization will remain machine and automated (computer-aided translation, CAT) translation systems. However, the use of such systems has its own characteristics, which depend on both the translator and the language into which the text is translated. The experience of using machine translation systems in the process of software localization under the conditions of daily work of professional translators for languages with low resources and rich morphology (Czech, Estonian, Hungarian, Latvian, Lithuanian and Polish) is reviewed in [11]. The results of experimental studies showed that the use of machine translation suggestions in addition to the use of CAT increased the productivity of translators in all assessment experiments. However, significant differences were observed in the productivity of different translation tasks and in the results of different translators. In [11] it was proposed to explain these differences

by the work mode and skills of individual translators. It should also be noted that the error rate increased in all of the experiments (except one), while remaining at the "Good" quality level. This result was explained in [11] by the fact that translators tend to trust the suggestions coming from the machine translation tool and do not check them sufficiently, even if they are marked as machine translation suggestions. It should be recognized that with the development of the capabilities of artificial intelligence tools, interest in the use of such tools in the field of software and video game localization has increased. Thus, in [12] an example of tuning the original neural machine translation model RNN-encoder-decoder with the addition of categorical information related to the problem of rare words (words that relate to a certain subject area, which is a common phenomenon in the text of software) is considered. However, this model has not been tested for video game localization and, according to the authors, provides successful translation results mainly for English and other official languages used by the United Nations [12].

The use of GPT capabilities for video game localization was investigated in [13]. The results show that the GPT-4 model is capable of providing multilingual output in a structured form. The GPT-4 model seems to have the ability to support the resolution of transcreation-related problems such as brainstorming, comparison, and coherence, although the results are not always satisfactory, and there are errors including conceptual repetition, formatting issues, and inappropriate or unusual portmanteau. In terms of output quality, the English-Chinese combination appears to be less creative compared to adaptations from English into French or Spanish. However, to draw any conclusions on this issue, further comparative qualitative and quantitative studies are needed, as well as reception studies among users to find out their preferred translations, as creativity is a highly subjective matter [13].

However, the use of GPT-4 for video game localization is not free from many drawbacks, among which [13] particularly highlight the following:

- the strong dependence of localization performance on the specific arrangement of tokens in the GPT-4 model;
- the limitation of the amount of context that can be used to improve coherence can lead to transcreations that deviate from previous work or do not correspond to the setting of the video game;
- GPT models can produce hallucinatory or nonsensical results or contain grammatical or spelling errors;
- GPT models may not take into account culturally sensitive issues, which can lead to the perpetuation, reproduction, and even reinforcement of stereotypes and prejudices that were present in the training data of these models;

- the different amount of data available in different languages affects the quality of results for certain language pairs during translation and cultural adaptation;
- GPT models risk violating copyright and intellectual property laws, which can expose users to financial and legal consequences;
- GPT models are currently limited by paid access, and the number of monthly messages allowed by more powerful models is limited, which negatively affects the localization market and workforce;
- opacity, together with the randomness of the results, limits the interpretability and applicability of the GPT model, as well as the reproducibility of the results.

However, the authors of the study [13] note that the results obtained and the identified limitations of the application of GPT models for video game localization were established by them on the example of one video game. Therefore, additional research is required to clarify the conclusions regarding the possibility and feasibility of using GPT models for video game localization. In general, [13] recommends using the capabilities of GPT models to interact with such categories of localization specialists as interns and beginners. Thanks to the capabilities of GPT models, these categories of specialists can increase their productivity and creativity, while maintaining control over the result using spoken machine transcreation based on prompts [13].

Summing up the interim results of the analysis of research on game localization tools, it should be recognized that the main localization tools remain the human translator and the CAT and machine translation tools he/she uses. The main advantages of using such tools are their speed and cheapness compared to human translation. However, the use of machine translation has a number of disadvantages, among which it is especially worth noting:

- the impossibility of a complete and high-quality understanding of the context;
- the inability to interpret idioms and cultural features;
- the inability to convey the literary features of the text;
- difficulties with the translation of highly specialized texts;
- the inability to be creative.

Therefore, it is impossible to fully localize a video game using machine translation tools, especially while maintaining the consistency of terminology, ensuring its correctness, and taking into account textual and linguistic features. Editing such a text requires a large number of editors, and its result will contain more undetected errors. To combine the advantages of each method, machine translation tools can be used in some cases at the translator's discretion to speed up the translation of individual translation units, rather than the text as a whole. The term "translation unit" (TU) should be understood as the smallest indivisible text that can be represented by a word, phrase, sentence, or set of sentences [14].

Thus, based on the results of the analysis, it is possible to conclude that there is a need for a modern perception of video game localization as a separate type of IT project. The main resources of these IT projects are game text translation specialists and CAT and machine translation tools. Modern Agile and hybrid methodologies can be used for the general management of such IT projects. However, specific methods and models for assessing, planning, and managing the work of such projects remain almost unexplored.

Therefore, the aim of this study is to develop a method for assessing the complexity of video game localization work with increased accuracy. This increase is planned to be achieved by taking into account a larger number of factors affecting the complexity of localization work than in currently existing methods.

To achieve the aim, it is proposed to solve the following objectives:

- develop a method for assessing the complexity of an IT project for video game localization;
- carry out experimental verification of the developed method.

4.3 Basic research method

The object of the study is the process of project assessment and control. The purpose of this process is to ensure the balance and feasibility of plans, determine the status of the project, its technical implementation and implementation of processes, ensure implementation according to plans and schedules within the projected budgets for technical tasks [15].

The main hypothesis of the study is the hypothesis of the possibility of increasing the accuracy of forecasting the duration of the IT project for video game localization by means of:

- research of factors that affect the complexity of localization work;
- development of a method for assessing the complexity of the IT project for video game localization.

The main assumption of the study is the assumption about the method of implementing the IT project for video game localization. According to the results of the analysis of modern scientific research presented in **Section 4.1**, this method is considered as the translation of game texts by a human translator using machine translation tools.

A study of current experience in quantitatively assessing time costs as an indicator of the complexity of an IT project for video game localization [16–19] showed that the main method of such assessment remains the method based on

the number of words. This method is based on calculating the indicator of translation time costs t_{est} according to the formula

$$t_{est} = w/s, \quad (4.1)$$

where w – the number of words in the text being translated; s – the recommended translation speed.

Different variations of this methodology can use different analogues of the indicators of the number of words in the text being translated and the recommended translation speed. Thus, instead of the indicator "Number of words in the text being translated", the following indicators of text volume can be used:

- "Total number of words in the localization file" (indicator $W1$);
- "Number of unique words in the localization file" (indicator $W2$);
- "Number of words in unique TUs of the localization file" (indicator $W3$).

The software translation speeds recommended in [16–19] can be characterized by the following indicators:

- "Interface translation speed" (indicator $S1$) [16, 17];
- "Speed taking into account the complexity of the text" (indicator $S2$) [18];
- "Baseline speed" (indicator $S3$) [19].

A comparison of the above speed indicators is presented in **Table 4.1**.

The considered indicators $W1$ – $W3$ and $S1$ – $S3$ can be compared depending on whether they take into account certain factors of localization complexity. The results of the comparison of the complexity factors considered in [16–19], which are taken into account by variations of the method based on the number of words, are given in **Table 4.2**.

The variants of the translation duration assessment methodology defined in [16–19] offer slightly different indicators for assessing the translation duration. However, all of these variants state that the duration of editing is equal to half the duration of the text translation work.

Table 4.1 Comparison of translation speed indicators proposed in [16–19]

Indicators	Recommended speed, words/day	Recommended speed, words/hour	Availability of variability
$S1$	2000	250	Present (depends on the type of content)
$S2$	1200–4000	200–500	Present (depends on difficulty)
$S3$	2500	300	Absent

Source: [16–19]

Table 4.2 Results of the comparison of the complexity factors, taken into account by variations of the method based on the number of words

Factor	Indicator		
	W1	W2	W3
Text volume	Takes into account	Takes into account	Takes into account
Text complexity	Doesn't count	Doesn't count	Doesn't count
Speed of the method application	Very high	High	High
Need for preliminary data processing	Absent	Necessary	Necessary
Taking into account the linguistic features of the text	Doesn't count	Doesn't count	Doesn't count
Uniqueness of the text	Doesn't count	Takes into account	Partially takes into account

Factor	Indicator		
	S1	S2	S3
Text volume	Takes into account	Takes into account	Takes into account
Text complexity	Partially takes into account	Takes into account	Doesn't count
Speed of the method application	High	Low	Very high
Need for preliminary data processing	Absent	Necessary	Absent
Taking into account the linguistic features of the text	Doesn't count	Partially takes into account	Doesn't count
Uniqueness of the text	Doesn't count	Partially takes into account	Doesn't count

The main drawback of the considered indicators of text volume and translation speed is that they do not take into account the complexity of the text (with the exception of S2). The factors considered in **Table 4.2** are sufficient for comparing the indicators with each other. However, the list given in **Table 4.2** is not an exhaustive list of factors that determine the duration and complexity of translation work.

According to the data in **Table 4.1**, it can be concluded that in [16–19] the recommended translation speed is set at or close to 300 words per hour. In cases where the translation work is already in progress, the average translation speed can be taken as the recommended speed. However, this indicator will be individual for each individual translator.

The review of the methodology based on the number of words allowed to draw the following conclusions:

- this methodology was and remains the most common means of calculating the duration of an IT project for the localization of an IT product;
- this methodology does not take into account a number of indicators that directly affect the duration of translation;
- this methodology is a quick, but not accurate way of calculating the predicted duration of translation in an IT project for the localization of an IT product;
- among the studied variations of this methodology, only variations based on the S2 indicator take into account the complexity of the text when calculating the duration of the localization of an IT product.

Therefore, when developing a method for assessing the complexity of an IT project for the localization of a video game, the considered features of the methodology based on the number of words should be taken into account in order to eliminate the identified shortcomings of this methodology.

It is proposed to divide the experimental verification of the research results into two stages:

- the main stage;
- the additional stage.

The purpose of the main stage is to verify the effectiveness of the developed method for assessing the complexity of an IT project for video game localization in comparison with the existing method based on the number of words. To verify this effectiveness, it is proposed to use the following indicators: absolute localization duration, relative error in the prediction of localization duration, relative average error in the translation duration of localization files.

The purpose of the additional stage is to verify whether the values of the proposed individual translation speed coefficient (ITSC), which were obtained within the IT project for the localization of one video game, can be used in the IT project for the localization of another video game. This will allow to investigate the possibility of applying the proposed ITSC in games that differ in genres and text styles.

Crowdin [20] was chosen as the CAT tool for video game localization used during the experimental verification of the research results. It is a cloud platform for managing translations and localization. It is suitable for localizing desktop and mobile applications, websites, and software documentation. Crowdin is a Ukrainian development and performs the same functions as Phrase Strings [21], but has both a paid subscription and a free version. The free version allows to create public and private projects up to 60,000 words. Crowdin contains information on over 176,000 localization projects and over two million registered users.

4.4 Results of the development of a method for assessing the complexity of an IT project for video game localization

4.4.1 Identification of factors affecting the complexity of video game localization work

As noted above, the list given in **Table 4.2** is not an exhaustive list of factors that determine the duration and complexity of translation work. Therefore, to develop a method for assessing the complexity of an IT project for video game localization, a special study was conducted to identify the main classes and groups of factors that can affect the progress and results of such an assessment [22].

According to the results of this study, it is proposed to divide the entire set of main factors that affect the progress and results of assessing the complexity of an IT project for video game localization into three main classes (**Fig. 4.3**):

- class of technical factors;
- class of factors set by the customer;
- class of factors determined by volume and content limitations.

It is proposed to include a set of limitations in the class of technical factors, which are primarily determined by the architecture of the IT product being localized, as well as the capabilities of the software used for localization. It is proposed to divide this set into the following groups of factors:

- "Selected CAT tool";
- "Localization file structure";
- "Variable organization";
- "Limitations on the number of characters".

The group of factors that describe the selected CAT tool characterize the environment in which the members of the localization team will work. A common feature for all CAT tools is that they allow to upload localization files in the original language, translate and edit text in a convenient interface, and at the output receive already formatted localization files in the target language. Most of these tools have the ability to connect translation memory and machine translation, distribute tasks between employees, track project progress, and allow to create and use dictionaries and glossaries.

The main differences between CAT tools are in tariff plans and functional features, such as: version control systems, automatic payment generation, project and participant statistics, etc.

The group of factors "Localization file structure" is determined by the developer. It is proposed to divide this group into the following subgroups of factors: distribution of TUs by files, quality of description of individual TUs.

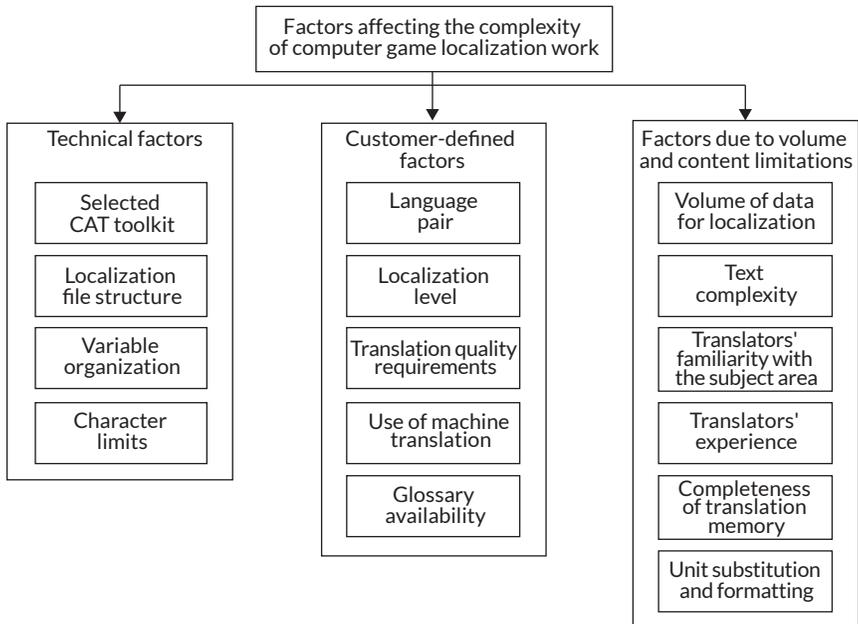


Fig. 4.3 Classification scheme of the main factors that influence the progress and results of the assessment of the complexity of an IT project for video game localization

The subgroup of factors "Distribution of TUs by files" characterizes how exactly TUs are distributed by developers across the file system. This distribution can occur in one of three ways:

- all TUs are contained in a single localization file;
- TUs are distributed across a set of several files;
- TUs are stored in a folder tree, where each folder is responsible for a separate group of localization files, united by common features.

The quality of the description of individual TUs involves their meaningful naming. Such naming simplifies the translator's understanding of the context and simplifies finding the TU's place in an IT product (in particular, in a video game).

If the localization files are organized as a tree, it is possible to determine the purpose of each TU group. Each TU has a key by which it is stored and inserted into the IT product. If the keys have meaningful names, the localization team can automatically distribute TUs to localization files. This will allow each employee to be responsible for translating files containing related TU groups. In addition, such a structure

allows to quickly understand the context of the TU being translated and, if necessary, find the desired TU in the IT product.

The subgroup of factors "Variable organization" characterizes the quality of variable naming and the number of these variables. This affects readability and understanding of the context. A large number of variables in a TU negatively affects its readability, as does their meaningless naming. High-quality organization of variables allows to better understand the context in which the TU is used, simplifies the coordination of words in a sentence, and increases its readability by translators and editors.

The subgroup of factors "Character Limits" characterizes the correspondence of the size of the translated content to the interface element in which it needs to be placed. First of all, the restriction is imposed on those TUs that are used to describe user interface elements. This achieves the correctness of the display of interface elements and texts. However, such coordination complicates the localization process by the need to fit the translated text into a limited set of characters. This may require the use of synonyms, alternative forms, abbreviations, but should not affect the content embedded in the TU.

The class of factors established by the customer proposes to include a set of restrictions that guide the localization process of an IT product and are determined by the customer. This set is proposed to be divided into the following groups of factors:

- "Language pair";
- "Localization level";
- "Localization quality requirements";
- "Permission to use machine translation";
- "Glossary availability".

The group of factors "Language pair" characterizes the pair of languages with which the translator works, namely: the original language and the target language. The analysis showed that today for most software products that are localized, the original language is English, however, in some cases the original language may be different. The localization process can be complicated depending on the language pair, namely the features of the target language and the original language, which differ. Such features may include:

- belonging of a pair of languages to different groups that determine the way words are formed in the language (agglutinative, inflectional or root languages);
- differences in the agreement of parts of a sentence depending on the gender and number of the noun;
- direction of writing the text, etc.

The group of factors "Localization level" characterizes the type of content that will be localized. In total, it is proposed to consider four levels of localization of video games, the choice of which depends on the volume of content to be localized [23].

The first level of localization involves the availability of the product in the local market in its original form. With the advent of digital distribution platforms such as Steam, Xbox, etc., the distribution of games has become much easier and does not require additional costs for delivering physical media to stores.

The second level of localization involves the localization of the game packaging (in the case of digital media – the store page) and documentation, but not the game content itself.

The third level of localization involves full text localization of the game, but not the sound. In this case, the interface and subtitles of the game must be fully localized.

The fourth level of localization involves full localization of the game, which includes text localization, localization of sound, graphical elements of the game, localization of documentation and other additional materials.

Thus, the first level of localization does not require translation, the second and third levels affect the amount of data for localization, which depends on the size of the project team. However, if the localization level is defined as the fourth, then the localization project team is joined by voiceover specialists (voiceover team), which will require additional interaction and cooperation with this team.

This study proposes to focus on the third localization level.

The group of factors "Localization quality requirements" is determined by the customer and depends on the budget and time allocated to the IT localization project. Depending on these requirements, such factors characterize the features of testing and elements of the software product that must be tested. Thus, testing of localization results can include both individual game elements and the game as a whole. In addition, if additional resources are available, additional stages of localization testing and text verification (so-called proofreading) can be carried out. Depending on the availability of human resources, testing of localization results can be carried out by the localization team, customer employees, or third-party contractors.

The group of factors "Permission to use machine translation" can be determined depending on the requirements set by the customer. Using machine translation can significantly speed up the translation process, but it increases the number of errors in it, because machine translation tools cannot yet understand the context, unlike humans. Thus, a text translated by machine translation will require additional attention during editing. Because of this, most often, machine translation is used by translators to speed up the translation in the following form: a quick translation of a separate TU is performed by machine translation, immediately after which

the TU is edited by the translator in order to correct errors, agree on proper names, take into account the character limit, etc.

The group of factors "Glossary availability" characterizes a document containing a list of terms and proper names that occur in the game, their transcription and explanations. The glossary should be translated before the localization files are translated. The glossary can be formed as follows:

- provided by the localization customer, since it is enough to form it only once in the original language, after which it can be provided to all localization teams;
- formed (developed) by the localization team "from scratch", which will require additional time.

The class of factors caused by volume and content restrictions is proposed to include a set of restrictions that are linguistic in nature and depend on the content that needs to be translated. This set is proposed to be divided into the following groups of factors:

- "Volume of data for localization";
- "Text complexity";
- "Translators' familiarity with the subject area";
- "Experience of translators";
- "Completeness of translation memory";
- "Unit substitution and formatting".

The group of factors "Volume of data for localization" consists of one indicator with the same name. This indicator is the main indicator that characterizes the required amount of work that needs to be done. On its basis, it is proposed to assess the time that will be spent on the implementation of a video game localization project. Most often, this indicator is measured in the number of words, but can also be measured in the number of TUs or characters.

The group of factors "Text complexity" shows how specialized the text being translated is. If the text contains a lot of technical terms or specialized vocabulary, localization may require finding or creating equivalents for these terms in the target language. For this, special industry dictionaries can be used or specialists working in a related field can be involved for additional consultations.

In video games, most often, the complexity of the text is associated with the literary features of the genre. Games set in the Middle Ages are characterized by the use of archaisms, historicisms, and archaic languages. Games set in fictional worlds are characterized by references to the original sources of the genre (if any) and the use of the appropriate terminology base.

The group of factors "Translators' familiarity with the subject area" in the case of video games characterizes:

- familiarity with the game itself or with the genres to which the game belongs;

- level of understanding of the context and plot of the game, which affects the understanding of the terminology and context of the terms used in the game;
- assessment of different translation options;
- course and result of choosing the translation option that best matches the original, helping to preserve the authenticity and atmosphere of the original.

Some games use special game systems (for example, such as *Dungeon & Dragons*). In such cases, it is necessary to coordinate the terminology with the localized original source (if the latter is available).

The group of factors "Translators' experience" characterizes the impact on the technical understanding of localization aspects and the speed of decision-making by the translator. Experienced translators who have knowledge of the gaming industry, its terminology, trends and features can better understand the context of the game and provide higher-quality localization. At the same time, the ability to work with CAT tools and effective management of their own resources can speed up the localization process. The group of factors "Completeness of translation memory" characterizes the impact on localization of previously acquired experience. Translation memory is a database containing a set of previously translated texts. It allows the user of the CAT tool to substitute a translation from memory when the current translation unit matches a previously translated one and, if necessary, refine it. The fullness of the translation memory significantly speeds up the localization process. Translation memory is filled by translators during the translation process of the game, but it can be transferred from previous localization projects that were carried out by the team. Such a transfer is useful for accelerating the translation of terms that are common to many games, but it is even more useful when transferring between games belonging to the same genre, since the terminological bases of such games overlap.

The group of factors "Replacement and formatting of units of measurement" characterizes the possibility or necessity of switching from one system of measurements to another (for example, from the metric system to systems of measurements based on the English system and vice versa). Such a transition includes formatting of measures of weight, length, as well as monetary units, date and time format, etc. Such formatting may include making changes to the text that is directly related to the units of measurement.

The proposed system for classifying factors that affect the progress and results of assessing the complexity of an IT project for video game localization allowed to establish the main factors that should be taken into account when developing a method for assessing the complexity of an IT project for video game localization.

4.4.2 Results of determining the content of the stages and activities of the method for assessing the complexity of an IT project for video game localization

The developed method for assessing the complexity of an IT project for video game localization is proposed to be presented as a sequence of the following stages:

- Stage 1: preliminary data preparation;
- Stage 2: assessment of the complexity of translating TUs contained in the localization files;
- Stage 3: obtaining the predicted duration of translating localization files;
- Stage 4: calculation of the total predicted duration of the project).

In accordance with the provisions of PMBOK [24], each stage of the developed method was divided into separate activities as follows:

Stage 1:

- Activity 1 "Preliminary preparation of data for further processing".

Stage 2:

- Activity 2 "Determination of TU similarity";
- Activity 3 "Assessing the complexity of the TU translation and calculating the complexity of the localization file".

Stage 3:

- Activity 4 "Calculating the baseline predicted translation speed of the localization file";
- Activity 5 "Calculating the baseline predicted translation duration of the localization file";
- Activity 6 "Determining the ITSC";
- Activity 7 "Calculating the individual predicted translation speed of the localization file";
- Activity 8 "Calculating the individual predicted translation duration of the localization files".

Stage 4: "Calculating the overall predicted duration of the project":

- Activity 9 "Calculating the overall predicted duration of the translation work";
- Activity 10 "Calculating the overall predicted duration of the editing";
- Activity 11 "Calculating the predicted duration of the IT project";
- Activity 12 "Calculating the average translation speed".

The activities are defined as follows:

1. Activity 1 "Preparation of data for further processing" consists in analyzing the input data and bringing it into a form convenient for further processing.

The input data of Activity 1 are provided in the form of localization files received by the technical specialist from the customer or from the program itself.

Pre-preparation of data involves structuring the TU (if this has not been done) and bringing the localization files to the sizes determined by the experts (members of the localization team) convenient for further processing.

The result of Activity 1 is structured localization files prepared for further processing.

2. Activity 2 "Determination of TU similarity" consists in determining the similarity of sets of words included in different TUs using the Jaccard measure.

The input data set for Activity 2 is the result of Activity 1.

The Jaccard measure (coefficient) $J(A,B)$ is calculated using the formula

$$J(A,B) = \frac{A \cap B}{A \cup B}, \quad (4.2)$$

where A – the set of words included in the first TU being compared; B – the set of words included in the second TU being compared.

The second TU is considered similar to the first if the Jaccard coefficient value exceeds 0.69. Then the same score will be applied to the second TU as for the first. This Jaccard coefficient value was obtained as a result of a study of the Crowdin CAT system settings [20].

The result of Activity 2 is a set of TUs that require further assessment.

3. Activity 3 "TU translation complexity assessment and localization file complexity calculation" consists in assessing the translation complexity of each unique TU included in the localization file. Based on the assessment results, the file complexity score and the predicted translation time are calculated.

The input data set for Activity 3 is the results of Activity 2.

The assessment of the complexity of the TU translation is proposed to be carried out by an expert on a nine-point scale. Possible assessments of complexity on this scale and their explanations are given in **Table 4.3**.

The localization file complexity assessment $grade_f$ is proposed to be presented as a weighted average assessment of all elements included in the file and determined by the formula

$$grade_f = \frac{\sum_{i=1}^m (grade_i \cdot n_i)}{\sum_{i=1}^m n_i}, \quad (4.3)$$

where $grade_i$ – the expert assessment of the complexity of the i -th TU; n_i – the size of the i -th TU (number of words); m – the number of TUs in the localization file.

The result of the execution of Activity 3 is a set of localization file complexity assessments.

Table 4.3 Developed scale for assessing the complexity of the translation of translation units

Difficulty group	Common features	Expert assessment	Excellent signs
TUs of low complexity	Translation of the TU does not require searching for additional information	1	Variables do not complicate the understanding of the TU, there is no need to adapt the text
		2	Variables make it somewhat difficult to understand the TU, there is no need to adapt the text
		3	Variables make it somewhat difficult to understand the TU, there is a need to adapt the text
TUs of medium complexity	In the TU, there are terms that require clarification, or non-standard lexical and grammatical constructions	4	Variables make it somewhat difficult to understand the TU, there is no need to adapt the text
		5	Variables to some extent complicate the understanding of the TU, there is no need to adapt the text
		6	Variables greatly complicate the understanding of the TU, there is a need to adapt the text
TUs of high complexity	In the TU, there are terms that need to be clarified, and non-standard lexical and grammatical constructions	7	Variables make it somewhat difficult to understand the TU, there is no need to adapt the text
		8	Variables to some extent complicate the understanding of the TU, there is no need to adapt the text
		9	Variables greatly complicate the understanding of the TU, there is a need to adapt the text

4. Activity 4 "Calculation of the basic predicted translation speed of the localization file" consists in calculating the basic predicted translation speed of the localization file v_f .

The input data set for Activity 4 is the results of Activity 3 and data on the size (number of words) of the assessed localization files.

According to [18], a translator can process 500 words per hour in the case of the simplest texts and 200 words per hour in the case of the most complex texts. For texts of average complexity, this indicator is 300 words per hour [16–19].

To establish the dependence of translation speed on the assessment of the complexity of the translation of the text, it was proposed to build a special model.

According to the developed scale for assessment (**Table 4.3**), it was established that the expert assessment of translation complexity 1 corresponds to a speed of 500 words per hour, the assessment of 5–300 words per hour, and the assessment of 9–200 words per hour. Thus, when the value of the function argument is 1, the function value is 500, when the value of the function argument is 5 – the function value is 300, and when the value of the function argument is 9 – the function value is 200.

Given the nonlinear nature of the relationship between the assessment of complexity and translation speed, the method of function approximation was used and it was established that this function is a branch of a hyperbola, which is located in the first quarter of the Cartesian coordinate system. Since the difficulty scores range from 1 to 9, the graph of the function is limited to the range $[1, \dots, 9]$ along the x -axis. Based on these data, a system of equations was developed to determine the parameters of the hyperbola passing through the specified points. This system of equations has the form:

$$\begin{cases} 500 = \frac{a}{1-h} + k, \\ 300 = \frac{a}{5-h} + k, \\ 200 = \frac{a}{9-h} + k. \end{cases} \quad (4.4)$$

The system of equations (4.4) has the following solution: $a = 4800$, $h = -7$, $k = -1000$. Therefore, the basic predicted translation speed of the localization file v_f is proposed to be determined by the formula

$$v_f = \left(\frac{4800}{\text{grade}_f + 7} \right) - 100. \quad (4.5)$$

The result of Activity 4 is the set of calculated basic predicted translation speeds of the localization files.

5. Activity 5 "Calculation of the basic predicted translation duration of the localization file" consists in calculating the basic predicted translation duration of the localization file t_f .

The input data for Activity 5 are the result of Activity 4 and the size (number of words) of the localization files.

The calculation of the basic predicted translation duration t_f is carried out by the formula

$$t_f = \frac{n_f}{V_f}, \quad (4.6)$$

where n_f – the number of words in the localization file.

The result of Activity 5 is the set of calculated basic predicted translation durations of the localization files.

6. Activity 6 "Determining the ITSC" consists of comparing the predicted translation duration value with the actual translation duration value of the data sample, which will allow to adjust further estimates of the translation duration of the localization files.

The input data for Activity 6 are sets of values of the actual and predicted translation duration of the test data sets.

The ITSC for a localization file $l_{f(x,y)}$ is defined as the ratio

$$l_{f(x,y)} = \frac{t_{f_{est}}}{t_{f_{real}}}, \quad (4.7)$$

where $t_{f_{est}}$ – the predicted translation duration of the localization file; $t_{f_{real}}$ – the actual translation duration of the localization file; x – the translator performing the translation; y – the expert performing the assessment.

The ITSC for pairs "translator-expert" $l_{(x,y)}$ is calculated by the formula

$$l_{(x,y)} = \frac{\sum_{i=1}^t l_{f_{(x,y)}}}{t}, \quad (4.8)$$

where $l_{f_{(x,y)}}$ – the ITSC value of the localization file obtained for the i -th pair of "translator-expert"; t – the number of localization files included in the test data set.

This calculation method is due to the fact that in different projects for the same pair of "translator-expert" who assess the TU complexity, knowledge in the subject area may differ. In addition, it reduces the amount of additional work performed, since the test sample is already part of the localization files of the project.

The result of Activity 6 is a set of calculated ITSC for each pair of "translator-expert".

7. Activity 7 "Calculation of individual predicted translation speed of a localization file" consists in calculating individual coefficients of predicted translation speeds of localization files.

The input data for Activity 7 are the results of Activity 4 and Activity 6.

The individual predicted translation speed v_f^* is determined by the formula

$$v_f^* = I_{test} \cdot V_f, \quad (4.9)$$

where I_{test} – the ITSC of the test data set.

The output of Activity 7 is a set of calculated individual predicted translation speeds for localization files.

8. Activity 8 "Calculation of individual predicted translation duration for localization files" consists in calculating the values of individual predicted translation durations for localization files, taking into account the difference in domain knowledge and experience of the translator and the expert who assesses the complexity of the translation.

The input data for Activity 8 are the sizes (number of words) of localization files and the results of Activity 7.

The individual predicted translation duration t_f^* is calculated by the formula

$$t_f^* = \frac{n_f}{v_f}, \quad (4.10)$$

where n_f – the number of words in the localization file.

The output of Activity 8 is a set of values for the calculated individual predicted translation times for localization files.

9. Activity 9 "Calculate the total predicted translation time" consists of calculating the sum of the translation times for the localization files into which the input data set was divided.

The input data for this activity are the glossary translation time and the results of Activity 8.

The total predicted translation time $t_{translation}$ is determined by the formula

$$t_{translation} = t_{glossary} + \sum_{j=1}^q \sum_{i=1}^b t_{f_{ij}}^*, \quad (4.11)$$

where $t_{glossary}$ – duration of glossary translation; $t_{f_{ij}}^*$ – duration of translation of one i -th localization file in the j -th category; b – number of localization files into which the j -th category is divided; q – number of categories into which the localization files are divided.

The result of Activity 9 is the calculated total predicted duration of work on the translation of the text provided for by the IT project for video game localization.

10. Activity 10 "Calculation of the total predicted duration of editing" consists in calculating the total predicted duration of editing in the IT project for video game localization.

The input data for Activity 10 is the result of Activity 9.

The total predicted duration of editing is determined by the formula [16–18]

$$t_{\text{editing}} = 0.5 \cdot t_{\text{translation}} \quad (4.12)$$

The result of Activity 10 is the total predicted duration of work on editing all the text provided for by the IT project for video game localization.

11. Activity 11 "Calculation of the projected duration of the IT project" consists of calculating the projected duration of the IT project for video game localization t_{total} .

The input data for Activity 11 are the results of Activities 9 and Activity 10, as well as the time allocated by the customer for testing the game localization.

The projected duration of the IT project for video game localization t_{total} is determined by the formula

$$t_{\text{total}} = t_{\text{translation}} + t_{\text{editing}} + t_{\text{testing}}, \quad (4.13)$$

where t_{testing} – the duration of localization testing.

The duration of localization testing is determined by the resources allocated by the customer and can be assessed by the localization team, development staff, or an external contractor.

The output of Activity 11 is the calculated projected duration of the entire IT project for video game localization.

12. Activity 12 "Calculation of the average translation speed" consists of calculating the average translation speed within the entire game localization project v_{avg} . The average translation speed in an IT project for video game localization can be used for further application in a word-based method to calculate the duration of other localization projects.

The input data for Activity 12 is the total volume of translated text and its translation duration.

The average translation speed v_{avg} is calculated using the formula

$$v_{\text{avg}} = \frac{n_{\text{total}}}{t_{\text{translation}}}, \quad (4.14)$$

where n_{total} – the number of words in the video game text for translation.

The result of Activity 12 is the value of the average translation speed within the IT project for video game localization.

The developed method takes into account the following factors that affect the complexity of video game localization work:

- "Localization file structure";
- "Variable organization method";
- "Language pair";
- "Translation quality requirements";
- "Volume of data for localization";
- "Text complexity";
- "Translators' familiarity with the subject area";
- "Translators' experience".

4.5 Experimental verification of the obtained results

Based on the proposals of **Section 4.3**, the experimental verification was divided into two stages: main and additional.

The IT project for video game localization "Don't Starve Together" [25] was selected for the main stage. At the same time, as indicated in **Section 4.3**, it is proposed to use the following indicators:

- absolute localization duration;
- relative error of localization duration prediction;
- relative average error of localization file translation duration.

The localization files for both stages were obtained directly from the game files. Thus, in the game "Don't Starve Together", each available language corresponds to a single localization file containing all TUs in this language.

An example of a TU from a localization file in the original language is shown in **Fig. 4.4**.

The TU shown in **Fig. 4.4** consists of several parts that have a certain meaning:

- keys (lines starting with "#."), which are unique identifiers needed to substitute values into the game interface;
- the TU context (lines starting with "msgctxt") indicates the context (application element) to which the string belongs;
- the original string (lines starting with "msgid") in English, which needs to be translated;
- the translation string (lines starting with "msgstr"), into which the text in the target language is placed.

```
#. STRINGS.ACTIONS.ABANDON  
msgctxt "STRINGS.ACTIONS.ABANDON"  
msgid "Abandon"  
msgstr ""
```

Fig. 4.4 Example of a Translation Unit in the original language from the "Don't Starve Together" game

In the case under consideration, the values of the keys coincide with the values of the context. This limits the context provided, but it is sufficient to understand which element of the game is being translated. The approach, in which the translation does not replace the original text, but duplicates it, increases the size of the localization file, but increases reliability (if the translation is missing, the value in the original language will be substituted).

During the execution of Activity 1, it was established that in the selected game, each available language corresponds to one localization file, which contains all TUs in this language. This file contains 72049 TUs with a total length of 515500 words. TUs contain meaningful key names, so they were automatically distributed among 113 localization files, each of which represents a certain subcategory. It was decided to reduce the amount of data for testing the developed method. 29 localization files were selected and it was decided to do only the translation without editing and testing. These files contain 2348 TUs with a total length of 18348 words (3.5% of the total number of words in the game). The localization files did not contain a separate glossary, so it was generated during the translation of the localization files using CAT tools.

During the execution of Activity 2, it was found that among the 2348 TUs there are 2200 unique TUs with a total length of 17815 words. The Jaccard coefficient (4.2) was used to determine the similarity of TUs.

Examples of the results of Activity 1 and Activity 2 are given in **Table 4.4**.

During Activity 3, an expert assessment of the translation complexity of TUs included in the localization files was conducted.

During the experimental test, the speed of expert assessment was measured. This speed ranges from 80 words per minute to 110 words per minute. This indicator may vary depending on the level of familiarity of the expert with the subject area. During the work on assessing the translation complexity of TUs, the expert's familiarity with the subject area increases. It is expected that the assessment speed will increase with the increase in the level of familiarity of the expert with the subject area.

As a result of Activity 3, according to formula (4.3), weighted average estimates of the complexity of the specified localization files were obtained.

Table 4.4 Characteristics of localization files

Name of file	Quantity of TUs	Quantity of words	Average size of TU
boarlord	65	377	5.8
carnival_crowkid	75	555	7.4
carnival_host	35	351	11.2
character_aboutme	19	317	16.68
character_descriptions	22	380	17.27
Name of file	Quantity of unique TUs	Quantity of words in unique TUs	
boarlord	64	375	
carnival_crowkid	74	551	
carnival_host	34	337	
character_aboutme	19	317	
character_descriptions	22	380	

During Activity 4, according to formula (4.4), the basic predicted translation speeds of localization files were determined.

During Activity 5, the baseline predicted translation time for each of the considered localization files was determined according to formula (4.5).

During Activity 6, the translation time for a test sample of 5 localization files with a total length of 1980 words was generated, translated, and measured. A comparison of the baseline predicted and actual translation times for these files is given in **Table 4.5**.

Table 4.5 Localization file characteristics

Name of file	Quantity of words	Basic predicted translation duration, min	Actual translation duration, min
boarlord	377	61	40
carnival_crowkid	555	92	47
carnival_host	351	55	36
character_aboutme	317	56	37
character_descriptions	380	72	49

After comparing the predicted and actual translation times of the localization files, the ITSC for long strings (2.5 words and more) was set to $I_{(xy)} = 1.5$. Given that files with a short average string length (less than 2.5 words) account for 806 words

out of a total of 17.815 words, it was decided not to calculate the ITSC for short strings using a separate test sample. Experimentally, during translation, the ITSC for short strings was found to be $I_{(x,y)} = 0.8$.

During Activities 7 and 8, individual predicted translation speeds and times were calculated and adjusted for a specific translator. In this study, one person acted as both the expert and the translator during the experimental validation. Therefore, the ITSC obtained as a result of Activity 6 is higher than that which can be obtained in a real project. The reason for this is that at the beginning of the translation work, the translator, who is also an expert, is already familiar with the text.

Examples of the results of Activities 4–8 are given in **Table 4.6**. In addition to the basic predicted translation speed and duration and the individual predicted translation speed and duration, which are calculated using the developed method, **Table 4.6** contains the actual indicators of translation speed and duration.

The individual predicted translation speed, according to [18], ranges from 200–500 words per hour. It was experimentally established that the real translation speed is higher than that suggested by the source [18]. The reasons for this may be:

- experience of translators;
- familiarity of translators with the subject area;
- peculiarities of video game translation as a type of software;
- peculiarities of the text being translated.

Table 4.6 Predicted and actual translation speed and duration

Name of file	Basic predicted speed of translation, words/min	Basic predicted translation duration, min	Individual predicted speed of translation, words/min
boarlord	367	61	550.5
carnival_crowkid	358	92	537
carnival_host	343	55	514.5
character_aboutme	319	56	478.5
character_descriptions	300	72	450

Name of file	Individual predicted translation duration, min	Real translation speed, words/min	Real translation duration, min
boarlord	41.09	565.5	40
carnival_crowkid	62.01	708.51	47
carnival_host	40.93	585	36
character_aboutme	39.75	514.05	37
character_descriptions	50.67	465.31	49

During the execution of Activity 9, the translation duration in the video game localization project was calculated. Since the glossary was generated and translated during the translation of the localization files, the translation time of the glossary was included in the translation time of the localization files. The predicted translation duration, which was calculated using formula (4.10), was 32 hours 57 minutes.

The video game localization IT project was limited to the translation of the localization files only. As a result, the editing time for Activity 10 is zero.

The testing time is determined by the customer, in this experiment it is also zero. Therefore, the predicted duration of the IT project for video game localization, calculated during the execution of Activity 11, is equal to the translation time, namely 32 hours 57 minutes.

The actual translation duration was 26 hours 26 minutes. Thus, the predicted translation duration exceeded the actual duration by 24.7%. The average predicted translation speed for the project, which was calculated during the execution of Activity 12, was 556.6 words per hour.

During the additional stage of the experimental verification, the effectiveness of using the ITSC obtained during the main stage in another video game localization project was tested.

The game "Sid Meier's Civilization® VI" was selected for localization and the ITSC obtained when using the developed method in the IT project for the localization of the game "Don't Starve Together" was taken. Games belonging to different genres were chosen for comparison in order to minimize the overlap of glossaries.

Unlike the video game "Don't Starve Together", the localization files of the game "Sid Meier's Civilization® VI" contain only a text string and a key, which is used to perform text substitution in the application.

When translating a localization file, the text in the original language is replaced with the text in the target language.

The localization files considered in the additional stage of experimental verification were obtained directly from the game in its original form. They were not changed during data preprocessing because the file organization proposed by the developer meets the requirements of the developed method. The number of words in the localization files was 6287.

The characteristics of the localization files are given in **Table 4.7**.

As in the main stage, the complexity of the selected localization files was assessed. As a result of the assessment, the basic predicted translation speed and duration were determined.

Table 4.7 Characteristics of the localization files of the video game "Sid Meier's Civilization® VI"

File name	Number of translation units	Number of words	Average size of translation units
Buildings_Text	59	1429	24.2
Combat_Text	75	703	9.4
CityStatePicker_FrontEndText	93	1646	17.7
DiplomacyPanel_Text	75	425	5.7
DiplomacyNotifications_Text	66	313	4.7
DiplomacyModifiers_Text	120	884	7.4
DiplomacyDeals_Text	136	887	6.5

ITSC $I_{(x,y)}$ was set to 1.5. This value was obtained at the main stage of experimental verification. Since the average TU lengths in the considered localization files are long, the corresponding ITSC value was selected.

Based on the obtained indicators of the basic predicted translation duration and ITSC, individual predicted translation speeds and durations were calculated.

The results of comparing the predicted values of translation speeds and durations with the real values for these localization files are given in **Table 4.8**.

Table 4.8 Results of comparing the predicted and real values of translation speeds and durations (additional stage)

File name	Basic predicted speed of translation, words/min	Basic predicted translation duration, min	Individual predicted speed of translation, words/min	Individual predicted translation duration, min	Real translation speed, words/min	Real translation duration, min
Buildings_Text	376	228	564	152.02	489.94	175
Combat_Text	341	123	511.5	82.46	602.57	70
CityStatePicker_Front-EndText	361	273	541.5	182.38	637.16	155
DiplomacyPanel_Text	338	75	507	50.30	520.41	49
DiplomacyNotifications_Text	415	45	622.5	30.17	722.31	26
DiplomacyModifiers_Text	408	130	612	86.67	964.36	55
DiplomacyDeals_Text	409	128	613.5	86.75	782.65	68

The predicted translation duration was 11 hours 11 minutes. The actual translation duration was 9 hours 57 minutes.

The predicted translation duration for the video game "Sid Meier's Civilization® VI", calculated using the ITSC obtained during the localization of the video game "Don't Starve Together", differs from the actual one by 12.1%. This deviation is less than that obtained for the game "Don't Starve Together" (24.7%).

Transferring ITSC from one IT localization project to another showed higher efficiency than in the original IT localization project for which the coefficient was calculated. This may be due to the following reasons:

- the peculiarities of working with the text of a particular game have become apparent;
- the sample on which the study was conducted is not large enough.

When loading localization files into the CAT tool and further processing them, it was found that the Crowdin CAT tool splits TUs stored in the ".xml" format (which contains the localization files for the game "Sid Meier's Civilization® VI") by sentences. If a translation unit consists of two sentences, then it is represented by the CAT tool as two separate TUs.

Splitting TUs by sentences speeds up the translation process of localization files. In the case when the level of similarity of TUs is low, the similarity of the sentences included in them may be greater. Such similarity expands the possibilities of using translation memory. However, such a division of TUs makes it difficult for the translator to understand the context, as he/she has to check the agreement of sentence parts not in a single TU containing several sentences, but in several TUs, each of which contains only one sentence.

An alternative reason for obtaining such results may be coincidence caused by an insufficient amount of test data.

In the experiments considered, the ITSC value is 1.5 for localization files with long lines.

In both experiments, this phenomenon requires further research with an increase in the number of games on which the study is conducted and the amount of text data.

4.6 Discussion of the results of the development of a method for assessing the complexity of an IT project for video game localization

The existing methodology for assessing the duration of an IT project for video game localization, based on the number of words, did not take into account the influence of a significant number of factors on this project. Therefore, a study was

conducted to identify and classify the main factors that affect the duration of such IT projects. The result of this study is a proposed classification system for the main factors that affect the IT project for video game localization, and the definition of the main groups of factors that make up the selected classes of factors.

Taking into account the proposed classification system for the main factors, a method for assessing the complexity of an IT project for video game localization was developed. The developed method consists of four stages, the implementation of which will allow calculating the predicted duration of an IT project for video game localization.

Unlike the existing methodology, the developed method takes into account the factors that affect the complexity of video game localization work, namely:

- "Localization file structure";
- "Method of organizing variables";
- "Language pair";
- "Translation quality requirements";
- "Volume of data for localization";
- "Text complexity";
- "Translators' familiarity with the subject area";
- "Translators' experience".

During the experimental testing, the developed method was compared with the existing method in the following options:

a) in the existing method, the total number of words was selected as the number of words, taking into account repeated TUs and variables, and as the translation speed, the speed recommended by sources [17–19] (300 words per hour);

b) in the existing method, the total number of words was selected as the number of words, taking into account repeated TUs and variables, and as the translation speed, the average translation speed obtained within the same project.

Comparison according to option (a) allows to verify or refute the statement that the developed method is more effective than the existing method.

Comparison according to option (b) allows to investigate the possibility of using the calculated average translation speed obtained as a result of using the method in further calculations of forecasting the duration of operations. In this case, the developed method can be used at the beginning of an IT project, when there is insufficient data to calculate the duration of translation. Then, after translating a certain number of localization files, the average translation speed obtained within the framework of a video game localization project can be used to predict the duration of translation of localization files. Further forecasting the duration of operations in an IT localization project based on the average translation speed will simplify the calculation of

the predicted duration of translation and will eliminate the need to use resources to assess the complexity of the TU translation.

For comparison, it is proposed to use the relative average error of the predicted translation duration [26]. Localization files have different sizes, so a relative, not an absolute value, was chosen for comparison.

For graphical visualization of the comparison results, it is proposed to use the relative error of the translation duration prediction and the square of the relative error of the translation duration prediction.

The absolute error of the translation duration of the localization file Δt is calculated by the formula

$$\Delta t = t_{est} - t_{real} \quad (4.15)$$

where t_{est} – the predicted file translation duration; t_{real} – the actual file translation duration.

The relative error of the localization file translation duration δt is calculated by the formula

$$\delta t = \frac{\Delta t}{t_{real}}. \quad (4.16)$$

The relative average error of the translation duration $\bar{\mu}$ is calculated by the formula

$$\bar{\mu} = \frac{1}{q} \sum_{i=1}^q \delta t_i, \quad (4.17)$$

where δt_i – the relative error of the translation duration of the i -th localization file; q – the number of localization files.

The results of the comparison of the errors of the developed method and existing methods for the localization files mentioned in **Tables 4.4–4.6** are given in **Table 4.9**.

The results of the graphical comparison of relative errors and their squares for both options are presented in **Fig. 4.5** and **Fig. 4.6**, respectively. These figures indicate: orange – data from a real IT project; blue – data calculated using the developed method; cyan – data calculated using the existing methodology according to option (a); green – data calculated using the existing methodology according to option (b).

The obtained data shows that the error in forecasting time costs calculated using the developed method is the smallest. In some cases, the results obtained using the developed method have a relative error of less than 10% (localization files actions_1, boarlord, character_aboutme, etc.). The accuracy of the developed

method is especially evident in the conditions of using localization files that contain short TUs.

Table 4.9 Results of the comparison of the errors of the developed method and existing methods for examples of localization files

File name	Word count method (basic)		
	Δt	δt	$\delta^2 t$
actions_1	-3.80	-0.1407	0.0198
boarlord	35.4	0.8850	0.7832
carnival_crowkid	64.00	1.3617	1.8542
carnival_host	34.20	0.9500	0.9025
character_aboutme	26.40	0.7135	0.5091
character_bios	293.80	1.2293	1.5112
character_descriptions	27.00	0.5510	0.3036

File name	Developed method		
	Δt	δt	$\delta^2 t$
1	5	6	7
actions_1	0.02	0.0007	0.0000
boarlord	1.09	0.0272	0.0007
carnival_crowkid	15.01	0.3194	0.1020
carnival_host	4.93	0.1370	0.0188
character_aboutme	2.75	0.0743	0.0055
character_bios	69.87	0.2923	0.0855
character_descriptions	1.67	0.0340	0.0012

File name	Methodology based on word count (average translation speed)		
	Δt	δt	$\delta^2 t$
actions_1	-13.8182	-0.5118	0.2619
boarlord	2.8409	0.0710	0.0050
carnival_crowkid	16.0682	0.3419	0.1169
carnival_host	3.8864	0.1080	0.0117
character_aboutme	-0.9773	-0.0264	0.0007
character_bios	63.7273	0.2666	0.0711
character_descriptions	-5.8182	-0.1187	0.0141

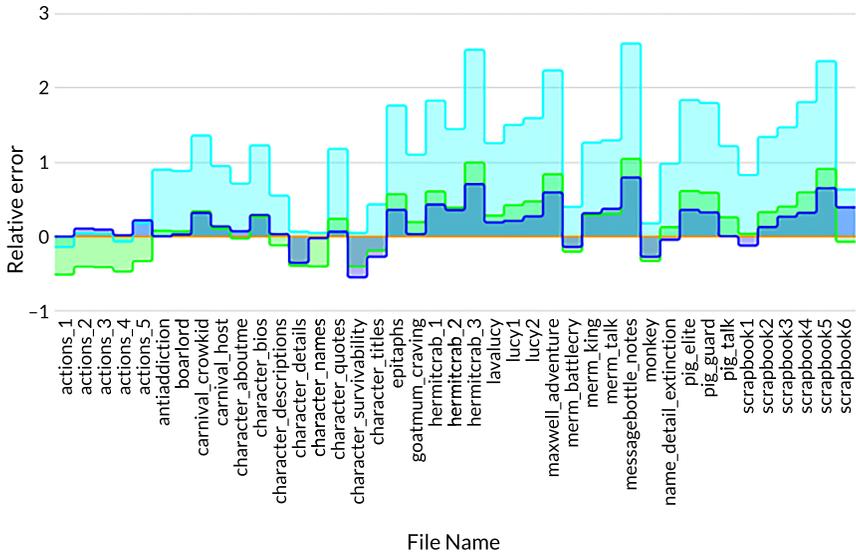


Fig. 4.5 Results of comparison of relative errors

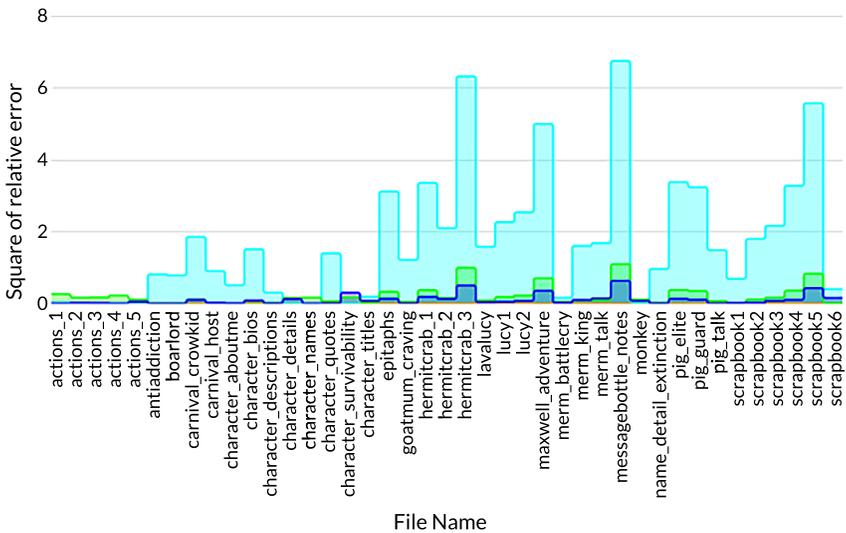


Fig. 4.6 Results of comparison of squares of relative errors

The results also showed that it is recommended to calculate the average translation speed for localization files with long and short TUs separately. This is due to the fact that the translation speed of short TUs is lower than the translation speed of long ones, and in video games, most TUs describe the game's scenario elements.

The total predicted translation duration was:

- using the developed method – 32 hours 57 minutes;
- using the existing methodology according to option (a) – 61 hours 9 minutes;
- using the existing methodology according to option (b) – 34 hours 44 minutes.

The calculated total relative average error was:

- using the developed method – 0.1639;
- using the existing methodology according to option (a) – 1.0661;
- using the existing methodology according to option (b) – 0.1739.

The actual translation duration was 26 hours 26 minutes.

The results of the predicted durations of the IT project for video game localization were compared with the actual duration of the project. The relative error of the forecast duration of the IT project for video game localization μ is calculated by the formula

$$\mu = \frac{t_{total_{est}} - t_{total_{real}}}{t_{total_{real}}}, \tag{4.18}$$

where $t_{total_{est}}$ – the predicted duration of the localization file translation; $t_{total_{real}}$ – the actual duration of the localization file translation.

The results of comparing the accuracy of the forecasts of the duration of the IT project for video game localization, obtained by both comparison options, with the actual results of the project duration are given in **Table 4.10**.

Table 4.10 Results of comparing the accuracy of the forecasts of the duration of the IT project for video game localization and the actual duration of this IT project

Calculation method	Accuracy metrics for duration forecasting for an IT video game localization project		
	Absolute duration	Relative error of duration forecast	Relative average error of localization duration
Real translation duration	26 h 26 min	-	-
Developed method	32 h 57 min	0.24	0.1639
Existing methodology (option a))	61 h 9 min	1.31	1.0661
Existing methodology (option b))	34 h 44 min	0.31	0.1739

Comparison of the obtained results showed that, in comparison with existing methods, the developed method for assessing the complexity of an IT project for video game localization showed the highest accuracy.

The main limitation of the developed method is the impossibility of its application in the case when the developer has not performed the distribution of TUs by localization files and the TUs themselves do not have meaningful names, which makes it impossible to automate the distribution of TUs by the localization team. In this case, manual implementation of such distribution is possible, but such an approach is not advisable for the following reasons:

- complication of input data analysis;
- the need to involve an excessively large number of resources;
- potential ambiguity of TUs;
- impossibility of searching for TUs in the application.

The complication of input data analysis lies in the fact that without a clear structure of localization files or meaningful names, the process of determining categories and subcategories by which TUs should be distributed becomes difficult.

The need to involve an excessively large number of resources is expressed in the fact that each TU must be read by an expert and assigned to one of the specified categories or subcategories. This operation is impractical because it consumes too much time that could be used to perform other processes by a team member.

The potential ambiguity of a TU is expressed in the fact that a video game localization expert cannot always determine which subcategory a TU belongs to based solely on its content. In this case, it must assign the TU to a subcategory based on some assumptions.

The impossibility of finding a TU in the application is that if there are named TU keys, localization team members can quickly and accurately determine the location of this TU in the application, which is necessary for verification and testing.

In the case where there are no named keys and the structuring is done by the developer, localization team members can only make assumptions about the location of the TU in the application based on its content.

Another limitation of this method is that it is intended for third-level localization. The amount of content localized at the first and second levels of localization is limited to several thousand words, so the use of the developed method is redundant. The fourth level of localization includes both text localization and the work of the voice-over team, the assessment of which is not provided for in the method.

The main disadvantage of the considered method is the need to assess the complexity of each TU. Such an assessment requires additional time and human resources, but allows to increase the accuracy of forecasts of the duration of operations in an IT project for video game localization.

The use of machine learning and natural language processing methods for assessing the complexity of TU translation is considered as the main vector for further development of the method.

The use of such tools will allow to get rid of the main disadvantage of the developed method – the time spent by experts on assessing the complexity of TU translation. In addition, it is expected to reduce ITSC fluctuations for different pairs of "expert-translator", since the assessment will be carried out by the same expert.

4.7 Conclusions

A new method for assessing the complexity of video game localization IT projects has been developed, which takes into account a larger number of complexity factors than the existing method based on the number of words. The developed method takes into account not only the number of words, but also other factors that affect the duration of localization. It is effective even in the absence of retrospective data on the speed of translation, which is an important advantage for the initial phase of projects.

An experimental verification of the obtained result has been carried out. The course and results of the calculations confirm the possibility of using the developed method to assess the characteristics of video game localization IT projects. It has been proven that using the developed method allows to obtain estimates of the time spent on the implementation of an IT project from the third level of video game localization, which are more accurate than the estimates obtained using the existing method. The total predicted duration of the translation was:

- using the developed method – 32 hours 57 minutes;
- using the existing method according to option (a) – 61 hours 9 minutes;
- using the existing methodology according to option (b) – 34 hours 44 minutes.

The calculated total relative average error was:

- using the developed method – 0.1639;
- using the existing methodology according to option (a) – 1.0661;
- using the existing methodology according to option (b) – 0.1739.

The actual translation duration was 26 hours 26 minutes.

Conflict of interest statement

The authors declare that there is no conflict of interest in relation to this paper, as well as the published research results, including the financial aspects of conducting

the research, obtaining and using its results, as well as any non-financial personal relationships.

Use of artificial intelligence statement

The authors declare that they did not use artificial intelligence tools in preparing this manuscript.

References

1. Wolf, M. J. P. (Ed.) (2008). *The video game explosion: A history from PONG to Playstation and beyond*. Bloomsbury Academic, 380.
2. Pirrone, M., D'Ulizia, A. (2024). The Localization of Software and Video Games: Current State and Future Perspectives. *Information*, 15 (10), 648. <https://doi.org/10.3390/info15100648>
3. Bernal-Merino, M. Á. (2011). Breve historia de la localización de videojuegos. *Trans. Revista de Traductología*, 15, 11–17. <https://doi.org/10.24310/trans.2011.v0i15.3191>
4. Esselink, B.; de Vries, A.-S., O'Brien, S. (Eds.) (2000). *A Practical Guide to Localization*. John Benjamins Publishing Company. <https://doi.org/10.1075/liwd.4>
5. Fry, D., Lommel, A. (2003). *The Localization Industry Primer*. Localization Industry Standards Association (LISA), 52. Available at: <https://www.scribd.com/document/581372706/The-Localization-Industry-Primer>
6. Toftedahl, M. (2020). Localization and regional aspects of game production – a research overview. *Proceedings of the 14th IADIS International Conference on Interfaces and Human Computer Interaction 2020 and 13th IADIS International Conference on Game and Entertainment Technologies 2020*, 135–143. https://doi.org/10.33965/ihci_get2020_2020101017
7. O'Hagan, M., Mangiron, C. (2013). *Game localization: translating for the global digital entertainment industry*. John Benjamins Publishing. <https://doi.org/10.1075/btl.106>
8. Ressin, M., Abdelnour-Nocera, J., Smith, A. (2011). Lost in Agility? Approaching Software Localization in Agile Software Development. *Agile Processes in Software Engineering and Extreme Programming*. Springer, 320–321. https://doi.org/10.1007/978-3-642-20677-1_25

9. Zorrakin-Goikoetxea, I. (2023). Survey on Localization From the Development Perspective. *Games and Culture*, 19 (3), 391–415. <https://doi.org/10.1177/15554120231168201>
10. Toftedahl, M. (2025). Towards a deeper understanding of localization as a game development process: Game studio studies in Sweden and China. *Entertainment Computing*, 54, 100944. <https://doi.org/10.1016/j.entcom.2025.100944>
11. Skadinš, R., Pinnis, M., Vasiljevs, A., Skadina, I., Hudik, T. (2014). Application of machine translation in localization into low-resourced languages. *Proceedings of the 17th Annual Conference of the European Association for Machine Translation, EAMT 2014*. <https://doi.org/10.13140/2.1.5047.6168>
12. Wang, X., Chen, C., Xing, Z. (2019). Domain-specific machine translation with recurrent neural network for software localization. *Empirical Software Engineering*, 24 (6), 3514–3545. <https://doi.org/10.1007/s10664-019-09702-z>
13. Moreno García, L. D., Mangiron, C. (2024). Exploring the potential of GPT-4 as an interactive transcreation assistant in game localisation: A case study on the translation of Pokémon names. *Perspectives*, 1–18. <https://doi.org/10.1080/0907676x.2024.2378346>
14. Bilous, O. M. (2013). *Teoriia i tekhnolohiia perekladu*. Kirovohrad: RVV KDPU im. V. Vynnychenka, 200.
15. ISO/IEC/IEEE 12207:2017. *Systems and software engineering – Software life cycle processes (2017)*. IEEE. <https://doi.org/10.1109/ieeestd.2017.8100771>
16. O'Donnell J. (2022). How long will it take to localize my game? – Localization scheduling. IGDA. Available at: <https://igda.org/news-archive/how-long-will-it-take-to-localize-my-game-localization-scheduling/>
17. Tarasiewicz, J. (2024). How much does the software localization cost? ATL. Available at: <https://www.atltranslate.com/articles/how-much-does-software-localization-cost>
18. Realistic translation times using human translators. PACTRANZ. Available at: <https://www.pactranz.com/translation-times/#speed>
19. Giulia, C. (2017). How do I calculate a translation time frame. Trusted Translations. Available at: <https://www.trustedtranslations.com/blog/calculate-translations-time-frame>
20. Crowdin. Available at: <https://crowdin.com>
21. Phrase Strings. Phrase. Available at: <https://phrase.com/platform/strings/>
22. Vasylytova, N., Osypchuk, D. (2024). Analiz pokaznykiv, shcho vplyvaiut na skladnist robot z lokalizatsii informatsiinykh system. *Radioelektronika ta Molod u XXI Stolitti*. Vol. 6: Konferentsiia "Informatsiini Intelektualni Systemy". *KhNURE*, 231–233. <https://doi.org/10.30837/iyf.iis.2024.231>

23. Chandler, H. M., Deming, S. O. M. (2011). Game Localization Handbook. Jones & Bartlett Publishers.
24. Nastanova do zvodu znan z upravlinnia proiektamy. Nastanova PMBOK (2021). Project Management Institute, Inc. Available at: https://learn.ztu.edu.ua/pluginfile.php/274061/mod_resource/content/1/PMBOK7_Ukr_ForPersonal-UseOnly.pdf
25. Don't Starve Together on Steam (2016). STEAM. https://store.steampowered.com/app/322330/Dont_Starve_Together/
26. Taylor, J. R. (1997). An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements. University Science Books, 327.