Data-driven Crowdsourcing in Public Emergency

Hui Zhao [1,2], Ali Selamat[1]

524850884@qq.com, aselamat@utm.my

[1.]Malaysia-Japan International Institute of Technology, Universiti Technologi Malaysia, Kuala Lumpur, Malaysia

[2.] Foreign Language School, Guangdong Polytechnic of Science and Trade, Guangzhou, China

Abstract— As a new translation model in the Internet era, crowdsourcing translation has attracted the attention of worldwide scholars, this study analyzes crowdsourcing translation's application mechanism and influence in public crisis events. This research applies UTAUT model to study the translators' acceptance of the crowdsourcing platform and how the other actors involved. The model assumptions are verified by using the calculation data of UTAUT configuration. This can provide reasonable recommendations for upgrading public crowdsourcing translation services in the world. A feasible reference can also serve the whole language service industry in the public crisis management system.

Keywords: crowdsourcing translation, public emergency, translators' acceptance

1 INTRODUCTION

Since the outbreak of new coronavirus, cross-border medical assistance and medical cooperation have been launched rapidly, and the demand for multilingual translation has been increasing sharply, such as the translation of medical materials, the language barrier of communication between doctors and patients, the customs clearance logistics level, the international medical technology exchange, and the bilingual tracking of international public opinion. The translation tasks in such public health emergencies are often time tight, trivial and fragmented, and the requirements of translation audiences are different. Volunteer crowdsourcing translation projects can convene a large team of community translators in a short time to provide language solutions to urgent needs. The author has the honor to participate in the emergency Chinese epidemic prevention volunteer translation project, undertake or witness the efforts of the crowdsourced translation project in dialect service, material translation and screening, bilingual analysis of public opinion, and international medical translation, and deeply feel the urgency of cultivating crowdsourced translation talents in public crisis.

2 MAIN DEFINITIONS

2.1 Crowdsourcing

The term "crowdsourcing" was proposed by Jeff Howe of the United States in 2006 in his Wired Magazine article 'The Rise of Crowdsourcing' [1], while in his blog the same month he defined it as follows: 'Crowdsourcing is the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call'. Crowdsourcing is referred to as, how enterprises connect local network users and gather group wisdom through the Internet. Howe applied the expression to several different activities; There are examples of the non-profit/public sector National Health Museum opting to use stock photos available from iStock photo rather than commissioning photographs from a single photographer for educational kiosks on world health topics [2].

Many not-for-profit organizations contribute to humanitarian work through crowdsourcing and bridge the digital divide. The not-for-profit organization, The Rosetta Foundation, combats the lack of access to essential information on health and education and removes language barriers by involving communities on the ground. In January 2010, a group of people exchanged GPS coordinates with other people at seven critical locations across Port-au Prince, where many Haitians were known to be trapped under rubble [3]. To sum up, crowdsourcing is a successful productivity model facilitated through its data-driven features where the community's wisdom is used to achieve an organization's task.

2.2 Crowdsourcing Translation

Perrino [4] provides the term 'User-generated Translation (UGT)' for harnessing Web 2.0 services and tools to make audio or video online content accessible in various languages. According to Perrino, UGT includes professional translation networks, translation wikis, user-generated dictionaries, online subtitling practices, and volunteer website localization. Crowdsourcing translation refers specifically to the translation activities of the online community in the era of web2.0 [2]. It represents a new style of translation by outsourcing the tasks to volunteer translators other than those previously entrusted to full-time translators on the network [5]. Emerging crowdsourcing translation projects originated from the creativity and practice of enterprise crowdsourcing, and were generated by the urgent needs of the increasingly diversified translation market, thanks to the continuous development of Internet technology support [6].

With crowdsourcing translation, we mean the existence of a platform and a volunteer community/crowd willing to translate with low payment (if any) content that is submitted online. We mainly focus on the translation of text-only and not of text existent in multimedia or translation networks and wikis.

3 LITERATURE REVIEWS

The literature analysis adopts the keyword clustering method based on co-occurrence relationship to mine the hot research topics of crowdsourcing translation. It is generally believed that the more lexical pairs appear in the same literature, the closer the relationship between the two keywords. In all keyword networks, the keywords with strong connection are close to each other and gathered together to form a keyword cluster with relatively independent concepts. After obtaining the main research topic clusters, use the visual Vosviewer to obtain the density view of clustering. From the density view, the research topics in this field concentrated can be seen.

3.1 Data Capture from Literature Reviews

The data is gathered from the WOS website from 2014 to 2022 with "crowdsourcing & translation" or as the keyword to search the literature. The records of the searched literature are selected from the documents that have been cited more than once (including once), and 10 fields such as title, author, abstract, title, research institution, publication time, cited frequency, downloaded frequency, classification and number are extracted from the literature data, import into MySQL database, literature cited less than once discarded. Excludes the options of conference procedure, book series and book, a total of 181 English documents from 2014 to 2022 are obtained.

Table 1 Top 20 high frequency keywords
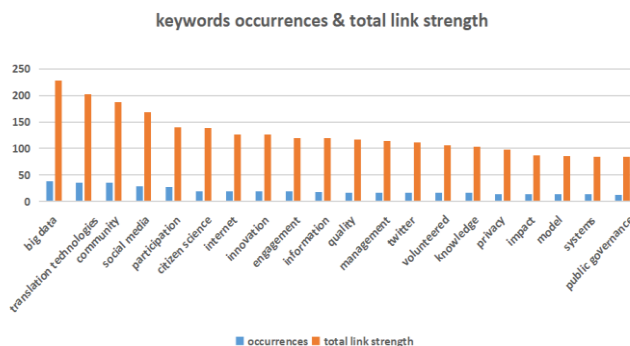
Through high-frequency keywords, we can understand the topics and hotspots in the research field. In this paper, VOSVIEWER software is used for keyword co-occurrence analysis, and the minimum number of keywords is set to 3 to get the list of the top 20 high-frequency keywords (see Table 1).

For data clearance, keywords with high coincidence rate are discarded, the highest frequency keyword is machine translation, translation technology, platform and big data, which are proved to be the technical support of crowdsourcing translation. Quality is the second largest concern in previous studies. Community translation, volunteer translation, human service, community interpreting, public governance are in the high frequency keywords list, which shows the crowdsourcing translation study is highly connected with public and community

service and governance. Facebook and Wikipedia, the commercial crowdsourcing pattern, are frequently cited, too.

The following picture is the density visualization for the above keywords. From the clusters, translation technology, quality, and community translation are research hotpots to cross and penetrate.
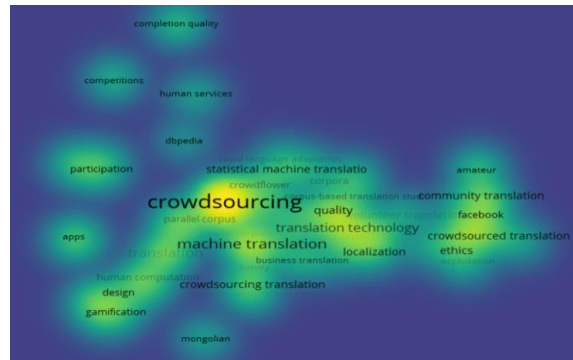
Pic 1 density visualization for keywords

To further analyze the keyword co-occurrence with time frame, the keyword co-occurrence frequency threshold are set to 3, deletes the too broad keywords with high frequency, such as crowdsourcing and translation, and the remaining 165 keywords are divided into 5 clusters. Circle nodes with different colors represent different clusters to which the research belongs. The size of the circle represents the popularity of different research topics, which is determined by the number of nodes cited and the strength of connections.
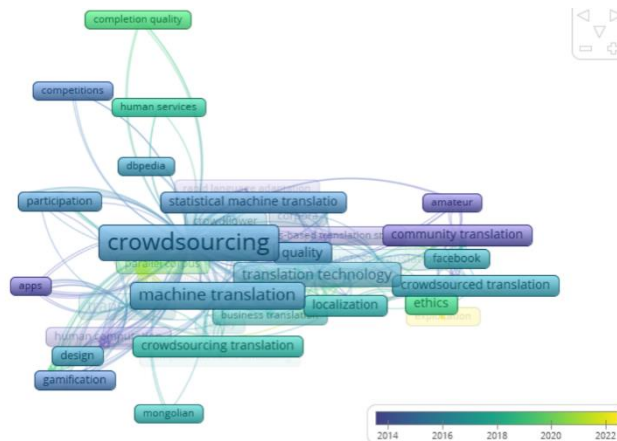
Pic. 2 Keywords co-occurrence time chart (2014-2022)

Take keyword co-occurrence time chart (Pic. 2) with VOSVIEWER software as the index board, from 2014 to 2016, the research focus is on community translation and crowdsourcing is a kind of amateur in scholars' research. To 2016, machine translation is used in crowdsourcing, participation and quality are gaining weight in crowdsourcing study. Around

2018, more translation technologies are applied in crowdsourcing, localization also falls into study scope. In 2020, ethics and completion quality are the main study topics. To 2022, parallel corpus become the direction of exploration.

3.2 More Analysis of Former Study

Taking authors and representative studies into a closer view, we can divide the previous studies into four focusing categories: the primary research, technical research, and application research, and ethical research, as shown in Table 2.

Table 2: Previous studies on crowdsourcing

Category	Research Focus	Cases	Representatives	Impact of the study	Issues still not resolved
Conceptual research	The study of the concept and characteristics, the feasibility and effectiveness, etc. of crowdsourcing translation	Profit-making and Non-profit organizations	O'Hagan (Dublin City University), F. J. Sobol (University of Arizona) [7], Michael Cronin [8] (Dublin University)	Clearly stated the open crowdsourcing platform and there is organic interaction and cooperation between these "temporary translators	What is the construction of the platform or how to make the platform a better one?
Technical research	Crowdsourcing translation software to serve language volunteers.	Gumpa Localization	MiguELA. Jiménez-Crespo [9], Rudnick (Indiana University), Chris Exton (University of Limerick) [10]	Attach importance on the development of technology in crowdsourcing	How to optimize the application of technology while remaining roles of translators

Application research	Model and application case analysis of crowdsourcing translation	Volunteer translator for 4636 Haiti in 2010; Arab Spring Egypt and Libya	SOBOL.F.J (Univerisity of Arizona.) Gwyneth Sutherlin [11] (Univerisity of Bradford), Lu Yan (Shanghai University), Cao Yixin (Sun Yesen University) [12]	Arouse the awareness of incorporating the crowdsourcing translation platform by the government and capital, and make it to embark on the road of profitability and legalization.	How to utilize crowdsourcing translation in solving emergent social issues or disasters?
Ethical research	the ethical aspects of crowdsourcing translation	Intellectual property, authorship and privacy	Joss Moorkens [13], Hao Junjie (Guangdong University of Foreign Studies) [14]	Make sense the ethical part or copyright of crowdsourcing.	How to integrate the sharing features and the ethical rules of crowdsourcing in operation?

3.3 Current Research Gap

Crowdsourcing translation is the symbiosis of translation technology and Internet technology. From its birth to its application, crowdsourcing translation is still in a tentative application stage just a few years. Most of its application research is to study its business model in profit-making organizations. The application cases focus on language service enterprise applications such as Chuanshen Company[1].and Google[2]. As an essential carrier of translation technology and applied translation research, crowdsourcing translation has achieved many international research results. The research trends cover the conceptual, ethical, technical, and application case studies of crowdsourcing translation, but the research focus is relatively single and isolated. The research starting points are primarily from the perspective of pure linguistics or software engineering.

---

[1] China Chuanshen Language Service Company who integrates language technology for quick translation
[2] Google Company who aims to integrate global information for public use and benefit everyone

4 FOCUS OF THIS RESEARCH

4.1 Existing Problems

The status-quo problems for crowdsourcing translation are, post-crisis volunteer translation team retention, internal governance mechanism of the volunteer team, adherence of volunteer loyalty, translation project management in crisis, crisis early warning mechanism of crowdsourcing translation volunteer team, language assets issues, financial problems, social positioning of the volunteer team, the lack of skillful and sophisticated crowdsourcing translator organization, etc.

The training of translators is waiting to be carried out according to the content and form of crowdsourced translation and combined with the translation technology platform to ensure that the crowdsourced translation technical team is supplemented with high quality, the platform is carried out and accumulated orderly, and constantly optimized.

4.2 Research Objective

The Chinese epidemic prevention translation volunteer group is taken as an application case. Volunteer crowdsourcing translation projects in pandemics can call up a large team of translation volunteers in a short time to better solve the urgent need. However, after the public event, the crowdsourcing translators and organizations are dismissed, the valuable translation material and experience are not accumulated. Therefore, this proposal analyzes the application mechanism and influence of crowdsourcing translation in public crisis events and provides prospects and suggestions for the sustainable development of crowdsourcing organizations so that a robust crowdsourcing translation system can be established to deal with translation demands in large-scale worldwide emergencies. Hopefully, a platform that integrates translation project management, human-machine translation, technology communication, and translator training and development will be built.

4.3 Research Contents

1)Description of the crowdsourcing translation project for 2020 Chinese epidemic prevention volunteers.

2)Analysis of the advantages and pitfalls of crowdsourcing translation in solving language crisis

In public crisis events, the main body of volunteers is accredited, certified, and experienced translators, but they are not often available. The training of translators is waiting to be carried out according to the content and form of crowdsourced translation and combined with the translation technology platform to ensure that the crowdsourced translation technical team is supplemented with high quality, the platform is carried out and accumulated orderly, and constantly optimized. Through the cooperation of schools, enterprises, associations, and governments, it is expected to ensure the quality of translator training and make the crowdsourced translation platform obtain high-quality social resources.

3)Prospect of sustainable development of crowdsourcing translation in a public crisis.

(1) Technology platform development

It is proposed to design a bilingual sentence pair production platform to integrate user management, team management, corpus management, machine-aided translation, and term-assisted translation.

(2) Translator training and management mechanism of voluntary translators for public crisis

It is proposed to set up professional and stable management, training, and coordination organization for crowdsourcing translation volunteers to form strict recruitment, training, management, supervision, and evaluation system.

(3) Cultivation of core competence of differentiated language service

It is proposed to cultivate the differentiated core competence of the volunteer team according to the category of public crisis events and the language ability of the translator. The team can be gradually divided into anti-epidemic translation teams, geological disaster translation teams, biochemical crisis translation teams, and military crisis translation teams, and each group can be further subdivided.

(4) Enhance social influence

It is proposed to incorporate the crowdsourcing translation volunteer organization into the process of public crisis management. Expand the credibility and influence of popular translation organizations from the aspects of government financial allocation, provision of authoritative certification, legal security and social supervision.

4.4 Research Steps

The research steps can follow the timeline or be adjusted according to the need of the public crisis. In this study, due to the lasting and intermittent pandemic, the workflow can be in the following steps but also allows slight changes or overlapping links.
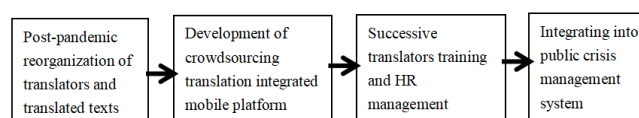
Figure. 1 Research steps conducted

The first step is the post-pandemic reorganization of translators and translated texts, to sort out all the crowdsourcing translated materials, texts, videos, pictures, or other multimodal translation, cataloging the material according to their fields, making improvements, and assembling it for publication under the condition of obtaining commercial license and copyright;

Translators shall be classified and filed according to their mastered language, translating content, translation ability, and region of their dwelling, and recorded as honorary volunteers with a precise file with time of stay. In the second step, develop a crowdsourcing translation integrated mobile platform for better and more efficient trial and Optimization of the current platform; In the third step, write the training plan of successive crowdsourcing translators, prepare corresponding training materials and supporting information-based teaching resources;

to confirm sound participation of translators in the next crisis, a well-designed human resource management system is to be built for the membership loyalty.

In the fourth step, to enhance the influence of the existing crowdsourcing translation organizations, obtain experience from other volunteer organizations, enhance effect through cross-border cooperation and cross-organizational publicity, promote public welfare cooperation alliances, and integrate the crowdsourcing translation organization into the public crisis management system.

4.5 Research Methodology

The China Anti-Pandemic Crowdsourcing Translation Team is taken as an application case to make clear the actors involved in the network concerned.

This research applies UTAUT to study the translators' acceptance of the crowdsourcing platform and how the other actors involved. The UTAUT model is an evolved version of Davis's Technology Acceptance Model (TAM). Despite its revision, such as UTAUT2, it still keeps its comprehensiveness and high explanatory power. Venkatesh, Morris, Davis, and Davis offered the UTAUT after benchmarking eight different models related to information technology (IT) acceptance. The eight models are the following: the theory of rational action (TRA), the TAM, the motivational model, the theory of planned behavior (TPB), the combined TAM and TPB model, the model of PC utilization, the innovation diffusion theory, and the social cognitive theory.

The critical part of this research then is how to design an effective crowdsourcing platform, a platform that integrates all technologies needed and is well-accepted by all crowdsourcing undertakers. The willingness and behavior to use the platform is the utmost intention we should and could survey.

UTAUT model aims to explain users' acceptance of new technologies and speculate on subsequent user behavior. The model calculates the impact of five core dimensions, namely performance expectancy (PE), effort expectancy (EE), social influence (SI), facilitating conditions (FC), and flexibility (FL) on users' behavior intention (BI) using crowdsourcing technology. This research involves language learners, language service providers, digital technology, human-computer interactive platform, and related rules and regulations.
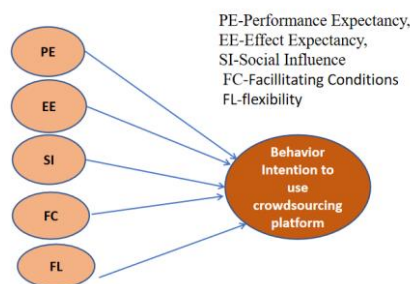

Figure. 2 Theoretical hypothesis model of research

PE is directed to digital technology providers, EE refers to language service providers, SI organizations and associations or government policies, FC language service market and FL universities and colleges.

According to the UTAUT model, thus we form our hypothesis:

A well-designed crowdsourcing platform provides effective and immediate language service in public emergency.

And we assume the following sub-hypothesis.

H1: performance expectation PE positively impacts the willingness of language service providers to use the platform.

H2: Effort Expectancy positively affects language service providers' willingness to use the crowdsourcing platform.

H3: Social Influence positively affects language service providers' willingness to use the crowdsourcing platform.

H4: Facilitating Conditions positively affects the willingness of language service providers to use crowdsourcing platform.

H5: Flexibility positively impacts language service providers' willingness to use the platform.

Questionnaires and research models are designed on the basis of the main hypothesis and five sub-hypotheses.

Because the crowdsourcing translators are formed in a great party by university students of language studies, and they are the army of future translators. Therefore, our questionnaire survey is done among the students and graduates of 10 colleges and universities in China. About 256 questionnaires were received from the participants, 50 invalid questionnaires were eliminated, and 205 valid questionnaires were obtained.

The demographic characteristics of the sample are shown in the table below, in which the male sample accounts for 9.75% and the female sample accounts for 90.24%. The age is concentrated between 18-30 years old, mainly with college education.

Table 3 demographic characteristics of the respondents

Column	Options	Sample	Ratio
Gendre	male	20	9.75%
	female	185	90.24%
Age	18-22	192	93.65%
	22-30	9	4.39%
	30-40	3	1.46%
	40-50	1	0.4%
Education Background	college	193	94.14%
	undergraduate	8	3.90%
	postgraduate	4	1.95%

The proportion of female in the sample is in the majority. The biggest reason is that most of the language service providers are females (most language learners are girls too). From the

perspective of sample age distribution and education level, the majority of the sample collection groups are college students. Because only in the crowdsourcing project of public emergency can the college degree language users enter without too much requirements and limitations, which also forms the big concern for translation quality in crowd-sourcing projects.

4.6 Reliability and Validity Analysis

In order to ensure the reliability and validity of the questionnaire, smart PLS partial least squares method is used to test the reliability and validity of the questionnaire. The results are shown in the table below.

In order to ensure the reliability and validity of the questionnaire, smart PLS partial least squares method is used to test the reliability and validity of the questionnaire. The results are shown in the table below.

When Cronbach's alpha value of Cronbach's coefficient > 0.7, composite reliability value > 0.7 and average variance extracted value > 0.7, it shows that the measurement model has good reliability. It can be seen from the following table that the Cronbach's alpha values of PE, EE, SI, FC, FL and BI are more significant than 0.9, and the combined reliability and Ave values of each variable are more critical than 0.9, indicating that the variables explain the model very well, and the model construction has good reliability and validity.

Table 4 Reliability and validity of the questionnaire

Construct Reliability and Validity

	Cronbach's Alpha	rho_A	Composite Reliability	Average Variance Extracted (AVE)
BI	0.961	0.961	0.975	0.928
EE	0.916	0.941	0.946	0.854
FC	0.886	0.894	0.930	0.815
FL	0.941	0.945	0.962	0.894
PE	0.948	0.957	0.967	0.906
SI	0.906	0.909	0.941	0.843

The measurement model has good reliability and validity and can be further analyzed by a correlation matrix. The square root of the Ave value of each dimension is significantly greater than the correlation coefficient of other dimensions, which meets the evaluation standard of discriminant validity, indicating that the measurement model has good discriminant validity. The correlation matrix is shown in the table below.

Table 5 Matrix Table (Note: the last data in each line is the square root of Ave)

	BI	EE	FC	FL	PE	SI
BI						
EE	0.531					
FC	0.842	0.690				
FL	0.847	0.746	0.893			
PE	0.669	0.761	0.801	0.807		
SI	0.798	0.554	0.835	0.802	0.674	

BI- users' behavior intention
EE- effect efficiency
FC- facilitating conditions

FL- flexibility
PE- performance efficiency
SI- social influence

The model assumptions are verified by using the calculation data of UTAUT configuration. The calculation diagram of UTAUT configuration is as follows:
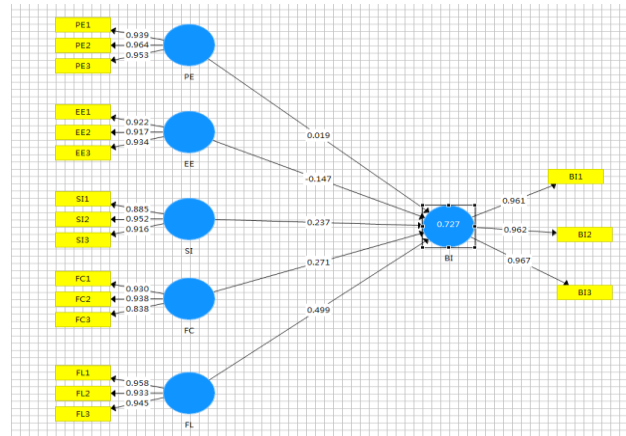
Pic. 3 The calculation diagram of UTAUT configuration

Combined with the calculation data, the verification results of the model assumptions are shown in the table below. If the path coefficient is more significant than 0.01, it reaches the significance level, which proves that the positive impact is tenable.

Table 6 Verification result table

hypothesis	path	Path coefficient and its significance	Conclusion (whether the hypothesis is supported)
H1	PE-BI	0.019	YES
H2	EE-BI	-0.147	NO
H3	SI-BI	0.237	YES
H4	FC-BI	0.271	YES
H5	FL-BI	0.499	YES

From the result, values prove to verify the significance of using the crowdsourcing platform, except that the expected value of EE is less than 0.01 and does not reach the significant level, maybe because the value of this variable is related to the low accessibility of language service providers using the platform.

5 CONCLUSIONS

The questionnaire samples are mainly obtained from language service providers in the China Epidemic Prevention Translation Volunteer Group. The assumptions are established through

UTAUT acceptance model and verified by smart PLS minimum binary algorithm. The Cronbach's alpha value of user behavior intention Bi is as high as 0.962, indicating that crowdsourcing professionals and non-professionals have a strong willingness to use a highly-integrated crowdsourcing platform. In contrast, PE (performance efficiency), SI(social influence), FC(facilitating conditions), and FL(flexibility) positively impact translators' willingness and behavior intention. A better crowdsourcing platform can be established through all these actors' improved models.

As a new translation model in the Internet era, crowdsourcing translation has attracted the attention of worldwide scholars. Public crisis governance is also a global problem. More and more countries have attached great importance to the research of social organizations participating in public crisis governance, especially the language service in a worldwide emergency. China's crowdsourcing translation service for the Covid-19 pandemic plays an essential role in the anti-pandemic campaign and thus becomes a vital issue for public crisis management. This research can provide reasonable recommendations for upgrading public crowding translation services globally by elevating the crowdsourcing platform to improve its availability in the language service of public crisis. A feasible reference is also carried out for the whole language service industry in the public crisis management system.

ACKNOWLEDGMENT

REFERENCES

[1] O'HAGAN M. From Fan Translation to Crowdsourcing; Consequences of Web 2.0 User Empowerment in Audiovisual Translation [J]. Approaches to Translation Studies, 2012b, (36):25-41

[2] Minako O'Hagan, The Routledge Handbook of Translation and Technology, 2019, Part 2.

[3] RUDNICK A, SKIDMORE T, SAMANIEGO A, et al. Guampa: A Toolkit for Collaborative Translation [EB/OL]. [2015-10-01]. http://wwww.lrecconforg/proceedings/lrec2014/gdf/151_Paper.pdf, 2014

[4] Perrino S. User-generated translation: the future of translation in a Web 2.0 environment. The Journal of Specialised Translation 2009; 12: 55–78, www.jostrans.org/issue12/art_perrino.pdf (accessed June 2011).

[5] Lu Yan, research on crowdsourcing translation model [J] Shanghai translation, 2012 (3): 74-78

[6] Zhu Lin, the rise of translation crowdsourcing and its impact on the development of translation industry [J]. Shanghai translation, 2016 (2): 71-78

[7] SOBOL F J. Translation in Translation: The Feasibility and Effectiveness of Crowdsourced Translation Versus Professional Translation [D]. Tucson: Univeristy of Arizona,2012

[8] Cronin M. The translation crowd. Revista traduma`tica 2010; 08 (available online)

[9] DESILETS A.VAN DER MEER J. Co-Creating a Repository of Best-Practices for Collaborative Translation [J]. Linguistica Antverpiensia New Series, 2011, 10(2): 27-45

[10] Miguel A. Jiménez-Crespo, crowdsourcing and online collaborative translations, 2017, Chapter 2, P39

[11] Chris Exton, Micro Crowdsourcing, a NewWay of Software Localization, 2009, The International Journal of Localization.

[12] Gwyneth Sutherlin, A voice in the crowd: Broader implications for crowdsourcing translation during crisis, Journal of information science, 2013

[13] Cao Yixin, network crowdsourcing translation model and the evolution of translator's subjectivity [J] Oriental translation, 2015a (1): 31

[14] Joss Moorkens, Post-editing evaluations: Trade-offs between novice and professional participants, 2015,

[15] Hao Junjie, a review of crowdsourcing translation research at home and abroad [J] Oriental translation, 2019:25-29

[16] B. Latour, Reassembling the social: An introduction to actor-network-theory, Oxford University Press, 2007:43

[17] Silvia Vermicelli, et al. How can crowdsourcing help tackle the COVID-19 pandemic? An explorative overview of innovative collaborative practices, 2021:183-195

[18] Narongdej Phanthaphoommee, Crowdsourcing Translation for Community-Based Tourism Empowerment in Thailand,2022:156

[19] LI Qian, Crowdsourcing Translation: A New Approach of Chinese Culture Translation, 2021:64-67.