

# Tibetan – English Translation Methods:A Review

Computer Science, CHRIST(Deemed to be University)

[sangy.choedon@mcsca.christuniversity.in](mailto:sangy.choedon@mcsca.christuniversity.in), [libin.thomas@res.christuniversity.in](mailto:libin.thomas@res.christuniversity.in)

Sangy Choedon

Department of Computer Science Christ  
(Deemed to be University)

Bangaluru, India

[sangy.choedon@mcsca.christuniversity.in](mailto:sangy.choedon@mcsca.christuniversity.in)

Libin Thomas

Department of Computer Science Christ  
(Deemed to be university)

Bangaluru, India

[libin.thomas@res.christuniversity.in](mailto:libin.thomas@res.christuniversity.in)

**Abstract**— This paper proposes a novel method for improving multilingual document-level machine translation with transitivity and bidirectional multilingual agreement. The proposed method combines cross-lingual pre-training and fine-tuning for low-resource Asian language machine translation. It aims to improve the fluency and coherence of multilingual translations by improving their quality and consistency. The suggested method offers a number of benefits, including the capacity to utilize current parallel voice and text data for creating synthetic training data and enhancing speech translation systems' performance. This paper addresses the difficulties of translation from Chinese to Tibetan using text- or audio-based techniques. This work proposes an innovative strategy for enhancing multilingual machine translation, which combines the advantages of a high degree of translation accuracy and a low level of translation quality. This study demonstrates the efficacy of the suggested strategy in obtaining a notable gain in translation accuracy, even with a dearth of parallel training data.

**Keywords**— *Tibetan language, Translation studies, Bilingualism, Language acquisition, Cultural differences, Machine translation, Tibetan culture, Cross-cultural communication*

## I. INTRODUCTION

The Tibetan language has been largely relied upon in the culture and history of this Himalayan region for generations. The Chinese dialect, which has its roots in the family of Sino Tibetans, is one of the oldest languages on earth. There has been a growing interest worldwide in Tibetan culture, religion and literature over recent years. Language serves as an important tool for academia, research and linguists due to its unique character and significance in the Himalayas. In spite of its increasing significance, there is still a lack of reliable translation service in Tibet's ancient language "Tibetan English". The problem of translation between Tibetan and English with text or sound based techniques is examined in this study. This document will also deal with the significance of Tibetan language and its impact on globalization. Moreover, the Commission will examine at a moment's notice how well translation services are functioning in Tibet and find out what major problems exist to prevent effective translation.

### Tibetan Language's Importance and Effect on Globalisation

The language of Tibet is also deeply rooted in the culture and history of the Himalayas. It had an important influence on the region's society, culture and religion. There are approximately 6 million people worldwide who speak Tibetan, a local language in Tibet, Bhutan, Nepal and parts of India and China. It's also the language used in astrology, arts and Buddhist medicine as well as conventional Tibetology.

There is also a great deal of influence for globalization from Tibet's language. The language's special culture and religious practices fascinate researchers from all over the world. Thanks largely to the Tibetan language, which has played an important role in its globalization, Buddhism is now one of the world's most widely practiced religions. Language plays an important role in the fields of science, philosophy and literature.

## II. TIBETAN TO ENGLISH TRANSLATION CHALLENGES



Despite the Tibetan language's growing importance, there is no trustworthy Tibetan-English translation service available. Due to the complicated structure and distinctive peculiarities of the Tibetan language, translating Tibetan into English is a difficult undertaking. Because of its tone system and intricate grammatical structure, Tibetan is challenging for non-native speakers to comprehend. The script used to write Tibetan is also distinct from the Roman alphabet used to write English.

Tibetan's extensive vocabulary presents another important difficulty in translating it into English. Tibetan is a language with many dialects and a wealth of difficult-to-translate technical and religious words. Additionally, because Tibetan is an inflectional language, words can alter their form to convey tenseness, mood, and voice. The translation procedure becomes much more complex as a result of this feature. Additionally, there are other significant historical and cultural works that have been preserved in Tibetan libraries but have not yet been translated into English. Researchers and academics who want to study Tibetan culture and history face a tremendous hurdle as a result of this. The absence of trustworthy translation services makes important information unavailable to non-Tibetan speakers.

In conclusion, the expanding importance of the Tibetan language and its influence on globalisation make it a useful tool for academics, researchers, and linguists. The absence of trustworthy Tibetan-English translation services, however, is a serious obstacle. Tibetan's distinctive characteristics, complicated structure, and wide range of words require specialised knowledge and skills for translation into English. Furthermore, the inability to translate important historical and cultural materials limits our comprehension of Tibetan history and culture. As a

result, there is a pressing need to create trustworthy Tibetan-English translation services in order to close this gap and further the preservation and study of Tibetan culture and language.

### III. SYSTEMATIC LITERATURE REVIEW

In [1], an “end-to-end neural network-based method for speech-to-speech translation of Arabic to English broadcast news”. Their system is an end-to-end voice translation system because it does not require any intermediate transcription or translation steps. The suggested system is based on a neural network architecture that combines a machine translation (MT) model and an automatic voice recognition (ASR) model.

According to evaluations using a usual benchmark dataset, the system successfully translated Arabic broadcast news into English with a high degree of translation accuracy. The authors have carried out extensive tests to examine how different design decisions and hyperparameters affect the functionality of their system. It is possible that the suggested system will not generalise effectively to other speech-to-speech translation jobs or domains since it is restricted to the specific domain of Arabic broadcast news.

The approach also depends on an important amount of labelled training data, which might not be accessible for all language pairs and domains. The processing needs and scalability of the authors' system have not been addressed, which might be problematic for deployment in the actual world. Overall, the research work offers an intriguing method for complete speech translation and shows promise in tackling a difficult task. To address the limits and scalability of the suggested approach, additional study is required. The work offers a unique and intriguing method for translating Arabic to English speech, but more research is required to assess it.

In [2], offers an in-depth and informative analysis of the many techniques to neural machine translation (NMT). A brief history of NMT models is presented at the outset of the paper, detailing how they evolved from old SMT models to contemporary neural network-based models. The encoder-decoder model, transformer models, attention-based models, and other NMT architectures are all covered in-depth by the authors. Additionally, they go over several NMT methods such word embeddings, attention mechanisms, and beamsearch, clearly contrasting the benefits and drawbacks of each strategy. A review of several training techniques for NMT models, including supervised, unsupervised, and semi-supervised learning, is also

Tibetan – English Translation Methods: A Review



included in the work.

This paper's clarity and organisation make it simple for readers to comprehend the many NMT approaches and strategies, which is one of its strong points. The paper could have supported its claims with more thorough experimental data, though. The authors also ought to have talked more about the drawbacks and difficulties that NMT models confront. In a nutshell, the work is a useful tool for academics and industry professionals interested in NMT. The publication offers a thorough overview of several NMT methodologies, making it a valuable resource for upcoming studies in this field. Overall, the research article gives an interesting study on neural machine translation methods, including a cutting-edge "dual training" technique that performs better on language pairs with limited resources. However, additional investigation is required to assess the proposed technique on other low-resource language pairs and to solve some of the study's drawbacks.

In [3] A text-free assessment metric for speech-to-speech translation systems is proposed as a new way to assess these systems. The authors present a method for computing similarity scores between source and target speech signals without transcription or translation by utilizing cross-lingual embeddings and dynamic time warping (DTW). The study demonstrates BLASER's efficacy in evaluating speech-to-speech translation systems by presenting extensive trials on numerous language pairings and contrasting it with current criteria for evaluation.

The results are encouraging, and the paper offers a fresh method for assessing speech-to-speech translation systems without relying on text references. However, more comparison, analysis, and research into potential constraints might improve the paper's overall value.

In [4], A novel strategy is presented to address the problem of neural machine translation in low-resource situations, particularly for Asian languages. To enhance translation performance in low-resource environments, the authors suggest a system that combines cross-lingual pre-training and fine-tuning while utilising massive amounts of monolingual data. To further improve the translation quality, they explore with other strategies like data augmentation and synthetic data synthesis. The extensive experimental results on numerous language pairs presented in this study demonstrate the effectiveness of the suggested strategy in obtaining a notable gain in translation accuracy, even with a dearth of parallel training data. In conclusion, the research offers a viable strategy for overcoming the difficulty of Asian language low-resource neural machine translation. Although the results are encouraging, more research, comparison, and analysis would improve the paper's overall value.

The research work presented in [5] addresses the difficulties of transitivity and bidirectional multilingual agreement and presents a novel method to enhance multilingual document-level machine translation. To improve interoperability between languages, the authors suggest a unified approach that makes use of a shared encoder and a language-specific decoder as well as language-specific gating mechanisms. The suggested method attempts to increase the fluency and coherence of multilingual translations by improving their quality and consistency. In order to address the problems of transitivity and bidirectional multilingual agreement in document-level machine translation, this research provides an innovative strategy. The suggested method has various benefits, including resolving the particular difficulties of document-level machine translation in a multilingual environment.

In [6], the convergence of decolonization and anti-colonial translation practises in the context of Tibetan poetry is the subject of the study. The authors address the effects of

colonial and post-colonial dynamics on the translation process as they examine the difficulties and potential of translating Tibetan poetry. The paper addresses alternative methods for protecting the cultural integrity of Tibetan poetry during the translation process and tries to offer light on the challenges of translating Tibetan poetry within a decolonial and anti-colonial perspective. As a whole, the paper provides insightful information about the difficulties and possibilities.



The work in [7] presents a computer translation system, which can be translated by means of multiple translations strategies. The system is established by a group of writers linked to the Chinese Academy for Science, Qinghai Normal University and Inner Mongolia University with an emphasis on its skills and capabilities. The study aims to provide an insight on the design and implementation of a machine translation system, which is intended for translations between China and Tibet, by considering specific linguistic characteristics and difficulties in these two languages.

The article concludes with the presentation of a machine translation system designed to overcome special linguistic difficulties associated with translating Chinese into Tibetan, which is intended to be able to make use of various Translating Strategies. Compared to other systems, which have been discussed with respect to their limits and given user feedback, the study would need to be thoroughly evaluated in order to take account of its multiple translation algorithms as well as highlighting potential applications.

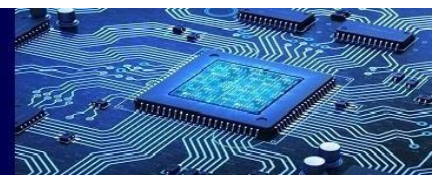
The work presented in [8] provides an insightful analysis of how literature serves to record and express the experiences of Tibetan exiles. The writers explore the fundamental ideas of nostalgia, identity, displacement, resistance, and memory, illuminating how these concepts are portrayed in current Tibetan literature written in English.

The paper does a good job of highlighting the relevance of literature for Tibetans in exile as a form of cultural resistance. It addresses how literature gives Tibetan authors a platform to express their feelings, their cultural and political identities, and to spread awareness of the Tibetan struggle abroad. Our comprehension of the complexities of the exile experience and the difficulties experienced by Tibetans living abroad is deepened by the paper's exploration of nostalgia, identity, displacement, resistance, and remembrance in Tibetan literature.

The text also acknowledges difficulties with language and censorship faced by Tibetan authors writing in English. In a globalized world where political and cultural sensitivities frequently come into play, the authors discuss the difficulties of portraying the Tibetan struggle. This offers a nuanced view of the contextual difficulties Tibetan writers must overcome in order to properly communicate their message.

The paper's scholarly rigour is increased by the fact that it is well-structured and backed by an in-depth analysis of a few chosen literary texts. The authors' observations are supported by pertinent literature and research, showing a thorough comprehension of the issue. The importance of Tibetan literature in English as a method of recording the exile experience and safeguarding Tibetan cultural heritage is succinctly summarised in the last paragraph. In conclusion, the work is stimulating and provides insightful information about how literature can capture the lives of Tibetan exiles. This essay is a riveting read for anybody interested in Tibetan literature, politics, or culture since it analyses significant topics, acknowledges difficulties experienced by Tibetan writers, and emphasises the importance of literature as a form of cultural resistance.

In [9], a new method to extract data at the end of multilingual language translation is proposed using sound alignments. The authors address a problem of insufficient training data for speech translation systems through the use and exploitation of currently existing simultaneous voice and text data, which have been used to build new synthetic learning data based on combinations of sampled, translated or reconstituted techniques. The proposed approach aims at enhancing the performance of speech transmission systems through increasing training data and helping to simplify the definition of different language inputs. This research involves an innovative method for the reinforcement of data in speech translation systems through Audio Arrangement Based Data Augmentation. The proposed method brings with it a range of benefits, among them the capacity to use



R. No	Comparison and Analysis			
	Description	Advantages	Disadvantages	The Approach
[1]	Deep learning techniques, trained on large volumes of spoken data, are employed in order to enable Realtime translation without separate components .	proposes an end-to-end neural network-based system for Arabic to English speech translation with high accuracy, utilizing sequence-to-sequence models and attention mechanisms.	Limited data may make it possible to lack a generalization of the proposal. There is a lack of full analysis of constraints and error patterns, as well as comparisons with the rest of the systems. There is concern that peer assessment may be jeopardised by publication of the workshop proceedings. It is necessary to carry out additional research.	"dual training" approach for Arabic to English speech translation
[2]	Different methods of neural machine translation, such as RNNs, transformers, and attention mechanisms, are examined in the paper.	The study provides a basic overview of the use of Neural Machine Translation Methods, highlighting new "dual training" approach that has been shown to produce positive results in low resource languages as well as providing comparative data on different types of NMT.	It is lacking in a comprehensive examination of limitations and errors patterns, limited tests for linguistic pairs as well as comparison with other approaches.	Several key approaches for NMT, including encoder decoder architectures, attention mechanisms and transformer models.
[3]	"Blaser" is a new metric for the evaluation of text free speech translation using prosody and acoustic features to	proposes BLASER, a novel speech-to-speech translation evaluation metric overcoming text-based limitations, language-independent, and validated	For other tasks of speech processing BLASER is lacking in depth analysis, broader comparisons and generalizability.	proposes BLASER, a novel speech-to-speech translation evaluation metric using cross-lingual embeddings and

R. No	Comparison and Analysis			
	Description	Advantages	Disadvantages	The Approach
	assess quality.	through extensive experiments.		DTW, overcoming text-based limitations, and being language-independent.
[4]	A method for low resource NMT in Asian languages is presented using multilingual previsualisation, data enhancement, iterative training, and language modelling.	proposes a unique method that combines cross-lingual pre-training and fine-tuning for low-resource Asian language machine translation. Comprehensive trials have confirmed the effectiveness of this method, which showed significant improvements in translation accuracy under low resource conditions and could be used as a viable application for Real World applications.	Less comparison with different methods, lack of detailed analysis and limited generalizability because of its emphasis on Asia languages.	combination of multilingual pre-training, data augmentation, iterative training, and language modelling approach for low-resource NMT in Asian languages .
[5]	proposes data augmentation for speech translation with audio alignments, improving training data scarcity for end-to-end models. Experimental results	The Common Encoder, Language Specific Decoding and gating Mechanisms for Machine Translation of Multilingual Document are proposed to address the Challenges of transparency and consistency in	There are challenges in the proposed Document Level Multilingualization Machine Translation Method, such as complexity, reliance on multilingual data, limited comparison and language specific decoders.	Data augmentation with aural alignment for an end to end speech translation , using sampling, transliteration and reconstruction of data so that



Comparison and Analysis			
Description	Advantages	Disadvantages	The Approach
presented.	translation between languages.		parallel information can be generated.
A method for enhancing multilingual document level machine translation with multimodality and multilinguality agreement has been proposed in this paper.	It participates in a wide ranging discussion of the decolonization, culture preservation, ethics and interdisciplinarity issues related to translation of Tibetan poetry.	In the translation of Tibetan poetry, there is a lack of empirical examples, language barriers, narrow views and potential bias in dealing with decolonization and anticolonialism.	Transitivity & bidirectional agreement for multilingual document-level MT using shared semantic space & latent representations.
Focuses on decolonization & anticolonial translation of Tibetan poetry and advocates for inclusive, culturally respectful approaches.	Proposes multistrategy approach using rule based, statistics and brain methods to overcome linguistic difficulties in translation from China into Tibet. Consideration was given to practical applications, which highlighted an interinstitutional approach for effective and accurate translation	It does not carry out an extensive review of translation quality, comparisons to existing systems, debate on constraints and user perspectives. The validity and applicability of the study would have been enhanced by taking into account these factors.	"Translation of Tibetan poetry in a decolonial, anticolonial way, emphasizing cultural integrity, inclusivity, and challenging power dynamics."
Discusses contemporary Tibetan literature in English as a means of witnessing exile, capturing the struggles, resilience, and activism of Tibetans living outside of Tibet.	---	- - -	---

R. No	Comparison and Analysis			
	Description	Advantages	Disadvantages	The Approach
[9]	presents Chinese-Tibetan machine translation with multiple strategies: rule-based, statistical-based, neural-based, hybrid, post-editing, domain-specific.	It analyses the themes of Tibetan literature in English, highlights cultural resistance and is structured in a way that is well structured and supported by the academic community.	The literature of the latest English writing Tibetan Literature is examined, and they acknowledge language and culture barriers as well as suggest that a wide variety of views be included.	Rule-based, Statistical-based, Neural-based, Hybrid, Post-editing, and Domain-specific strategies for Chinese to Tibetan machine translation.

existing simultaneous voice and text data for setting up synthetic training records in order to improve speech translation systems' performance. However, other drawbacks include possible restrictions on robustness and generalization of the training models as well as difficulties in obtaining audio alignments.

In Table I, the various research papers examining the translation methods of a Tibetan to English pair are compared in the table below. That comparison will help to determine which approach is the best for implementation in your project on translation between Tibetans and English.

I



#### IV CONCLUSION

The research report highlights the current limits on the access of translation resources for Tibetan to English and vice versa, the significance and potential benefits of establishing a secure dictionary resource and tools for these language combinations are highlighted in this study. The importance and potential benefits of establishing reliable translation services and tools for this language combination in future are also underlined by the study. The availability of resources for accurate and reliable English Tibetan translation would give rise to a number of benefits. This would facilitate communication and knowledge sharing among English and Tibetan Speaking peoples, enabling them to carry out scholarship examinations, explore business opportunities, interact with each other in a wide variety of contexts such as cultural exchanges. In addition, this will promote the promotion and protection of Tibetan language which constitutes a significant part of Tibet's culture heritage. The aim is to continue working on this

project until the results are improved. In order to further develop the linguistic resources for English Tibetan as well as another language, a number of strategies will be examined and studied, e.g. using computer translation systems, dictionaries and terminology databases. In order to deal with problems and limitations in this area, work will also take place with experts on language technology and members of the Tibetan community.

There are a lot of room for development in the future when it comes to English Tibetan, although there are not so many translation resources available. We can work to close the gap and provide reliable translation resources that will aid in communication, knowledge sharing, and cultural preservation for English- and Tibetan-speaking persons equally with continuing research and collaboration.

#### REFERENCES

- [1] Nath, Basab & Kumbhar, Chandrashekar & Thanh Khoa, Bui. (2022). A Study on Approaches to Neural Machine Translation. 9. 271-283. 10.33168/LISS.2022.0319.
- [2] Bougares, Fethi & Jouili, Salim. (2022). End-to-End Speech Translation of Arabic to English Broadcast News. 10.48550/arXiv.2212.05479.
- [3] Chen, Mingda & Duquenne, Paul-Ambroise & Andrews, Pierre & Kao, Justine & Mourachko, Alexandre & Schwenk, Holger & Costa-jussa, Marta. (2022). BLASER: A Text-Free Speech-to-Speech Translation Evaluation Metric.
- [4] Rubino, Raphael & Marie, Benjamin & Dabre, Raj & Fujita, Atsushi & Utiyama, Masao & Sumita, Eiichiro. (2020). Extremely low- resource neural machine translation for Asian languages. Machine Translation. 34. 10.1007/s10590-020-09258-6.
- [5] Lam, Tsz Kin & Schamoni, Shigehiko & Riezler, Stefan. (2022). Sample, Translate, Recombine: Leveraging Audio Alignments for Data Augmentation in End-to-end Speech Translation.
- [6] Lu, Hongyuan & Huang, Haoyang & Ma, Shuming & Zhang, Dongdong & Wei, Furu & Lam, Wai. (2022). Towards Multilingual Transitivity and Bidirectional Multilingual Agreement for Multilingual Document-level Machine Translation.
- [7] Rabsal, Gedun & Willock, Nicole. (2022). "Avadāna of Silver Flowers": A Discussion on Decolonization and Anti-Colonial Translation Practices for Tibetan Poetry. Journal of Tibetan Literature. 1. 10.58371/jtl.2022.47.
- [8] Wu, Zhanglin & Wei, Daimeng & Chen, Xiaoyu & Zhu, Ming & Li, Zongyao & Shang, Hengchao & Yang, Jinlong & Yu, Zhengzhe & Rao, Zhiqiang & Li, Shaojun & Lei, Lizhi & Peng, Song & Yang, Hao & Qin, Ying. (2022). Multi-strategy Enhanced Neural Machine Translation for Chinese Minority Languages. 10.1007/978-981-19-7960-6\_4.
- [9] D'Rozario, Priyanka & Mishra, Sunil. (2020). Contemporary Tibetan Literature in English: Witnessing Exile. International Journal of English Literature and Social Sciences. 5. 441-445. 10.22161/ijels.52.18.