International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING

ISSN:2147-6799 www.ijisae.org Original Research Paper

Text Simplification Improves Text Translation from Gujarati Regional Language to English: An Experimental Study

Dhawal Khem*[1], Shailesh Panchal[2], Chetan Bhatt[3]

Submitted: 24/10/2022 **Revised:** 23/12/2022 **Accepted:** 22/01/2023

Abstract

Text translation plays an important role in increasing the reach of information technology to the large portion of the population. The text translation helps to overcome the language barrier. An adequately translated and simplified text helps to improve the quality of communication. In recent times many researchers proposed research work on text translation. However, the grammar and complex formation of regional languages are bottlenecks in effective text translation. Many researchers proposed text simplification before text translation. The text simplification improves the readability and understandability of the text. However, regional language simplification and translation is still a challenging task for the researchers. Gujarati is an Indian regional language. In this paper, an experimental setup is proposed for improved text translation from Gujarati language to English language. Results show that the text simplification improves the quality of translation. We also experimented with text translation of the Indian national language - Hindi showed an improvement in translation results.

Keywords: *Text Simplification, Text Translation, Text Readability, Natural Language Processing, Indian Language*

1. Introduction

Communication is an essential part of human life. It is observed that lack of foreign language knowledge becomes a bottleneck in interaction and information exploration. Information technology helps to overcome the language barrier through text translation and text simplification. Text translation converts the text from one language to another language and text simplification improves the text readability and understandability. Historically humans have served as the intermediate person to do language translation and simplification. But in the current globalization time when people are traveling to distant locations, interacting with each other on social media, and exploring information about different regions, religions, and matters, automated text translation and text simplification become helpful. These text translations and text simplification remained essential services to improve communication [1]–[3].

To improve text translation various methods of machine translation [4], [5] and text simplification [6] have been proposed. The translation quality hampers when the translation takes place between non-sibling and resource-constrained languages. According to Wikipedia, Gujarati [7] and Hindi languages [8] are Indo-Aryan languages, meaning they are non-sibling languages to the English language. These languages are morphologically very rich languages, they use agglutination in their words, and are highly inflected languages. Compared to the English language both of these languages are resource-constrained languages. Thus facing the limitation of training NLP-based Machine Translation and Text Simplification systems. As well as they show bad translation results for local language contexts.

To improve the local language translation, text simplification as a preprocessing step is useful. Chandrasekar et al., 1994 [9] suggested improving text translation using text simplification as a preprocessing step. Panchal et al., 2015 [10] noted that text simplification helps to improve the quality of text translators and information retrieval systems. Building a text simplification engine f or a resource-constrained language is a difficult and time-consuming task. Here in this paper, we propose the text simplification engine for Indian regional language called Gujarati and an experimental study to validate the research hypothesis.

2. Literature Review

Text simplification was originally proposed as a pre-processing step for machine translation [9]. For many language pairs (e.g. English-French, English-Spanish, English-Hindi), attempts were made at rewriting input sentences using paraphrasing or textual entailment to improve the performance of MT systems [11]–[15]. Saggion et al. [16] presented an approach SIMPLEX of text simplification for Spanish. Mirkin et al. [15] manifested a way to enhance the source text before translation using SORT as a web application with Model View Controller (MVC). Paetzold and

**1Ph.D Scholar, Computer Engineering, Gujarat Technology University, GTU, Ahmedabad(Gujarat), India. ORCID ID : 0000-0002-8064-5954*
2Professor, PG-Cyber security, Graduate School of Engineering & Technology (GSET), Ahmedabad(Gujarat), India
3Professor, MCA Department, K. K. Shashtri College, Ahmedabad(Gujarat), India.
** Corresponding Author Email: khemdhawal@gmail.com*

Specia [17] have analyzed both syntactic and lexical simplification by learning tree transduction rules using Tree Transduction Toolkit (T3). Ameta et al. [18] developed a rule-based Gujarati stemmer for text simplification and improving the quality of the Gujarati-Hindi machine translation system. Patel et al.[19] presented a reordering approach. They have evaluated the quality of text in terms of BLEU, NIST, mWER, and mPER. They have concluded that adding more rules for reordering improves the translation quality.

The translation quality hampers when the translation takes place between non-sibling languages and languages which are resource-constrained. Text simplification engine to such languages helps to improve the local language sentence compatibility with the translator and may help to improve the quality of the overall language translation.

3. Experimental Setup

Here in this paper, we discussed an experimental setup for verifying the hypothesis, using text simplification as a preprocessing step improves the quality of translation, for low-resource languages like Gujarati and Hindi. We have experimented with the pairs of complex-simple sentences in Gujarati, Hindi, and English.

3.1. Dataset

We used two types of datasets: 1) Translated sentence pairs, and 2) Complex Simple sentences pairs.

These are collected from online resources, as well as prepared manually. For English language and translated pairs, dataset collected from online resources. For the Gujarati and Hindi language complex-simple sentence pairs, we have prepared manual text sets.

1) Translated sentences pairs

We have prepared 20 sentences text-sets over Gujarati, Hindi, and English languages, for all possible language pairs for "text - reference texts" (i.e. Gujarati-Hindi, Hindi-Gujarati, Gujarati-English, English-Gujarati, Hindi-English, and English-Hindi). A sample text and reference texts are shown in the table.

Table 1: Gujarati-English Language Text-Set

text (transliteration)	જનનીની જોડ સખી નહિ જડે રેલોલ (Jananini jod sakhi nahi jade re lol)	Gujarati Language
reference text	Friend, no one else can be found equal to mother	English Language
	Friend, no one else can be found like mother	
	Friend, no one else can be found equivalent to mother	

2) Complex Simple sentences pairs

We used an online available TURK dataset [25]. TURK dataset is a multi-reference dataset for the evaluation of sentence simplification in English. The dataset consists of 2,359 sentences from the Parallel Wikipedia Simplification (PWKP) corpus. Each sentence is associated with 8 crowdsourced simplifications that focus on only lexical paraphrasing (no sentence splitting or deletion).

In the case of Gujarati language, we are required to prepare complex simple sentence pairs with manual efforts. Many of the sentence pairs we derived from an online available book Gujarati Rudhiprayog Ane Kahevat Sangrah [26], and some of these we prepared by referring to online Gujarati literature. These texts are taken from openly available online resources.

Table 2: Complex Sentence (CS) and Simple Sentences (SS). The sentences are in the Gujarati language. Their corresponding transliterations are given below the sentences.

Complex Sentence (CS) (transliteration)	:	જનનીની જોડ સખી નહિ જડે રે લોલ (janani ni jod sakhi nahi jade re lol)
Simple Sentence (SS-1)	:	મિત્ર મા સમાન અન્ય કોઈ નહિ મળે (mitra ma saman anya koi nahi made)
Simple Sentence (SS-2)	:	મિત્ર માતા સમાન અન્ય કોઈ નહિ મળે (mitra mata saman sanya koi nahi made)
Simple Sentence (SS-3)	:	મિત્ર, માતા સમાન અન્ય કોઈ નહિ મળે (mitra, mata saman anya koi nahi made)

Table 2 shows an example of a complex sentence (i.e. a complex sentence (CS)) and simple sentence (SS) versions. These sentences have the same meaning but with different word usage. Below are the steps we followed to perform the execution.

3.2. Model

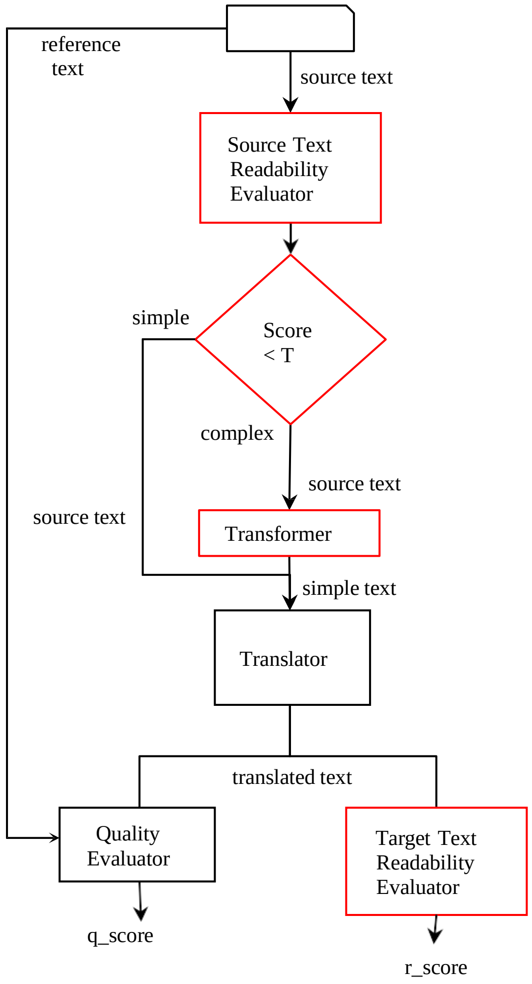


Fig 1: Execution Model

Fig 1 shows our execution model.

- **Text Set:** The text-set contains pairs of text and reference texts. These text-sets are divided into two categories based on the readability score, a) complex sentences, and b) simple sentences. Complex sentences are those which are difficult to read and understand, and simple sentences are those which are comparatively easy to read and understand. We have taken test cases in Gujarati, Hindi, and English languages.
- **Readability Evaluator:** It contains a set of scripts, which generate the text readability score. The available scripts are Flesch Reading Ease score [20], Flesch-Kincaid score [21], Gunning Fog Score, and others [22]. The existing scripts are working for English language sentences, so we developed our scripts for Gujarati and Hindi languages to calculate the text readability.
- **Transformer:** This is the heart of the execution model. It simplifies the input text and gives more readable simple output text. The transformer is being implemented for Gujarati language based on a knowledge based approach, using a simplified lesk algorithm [28]. The readability score of the output text is better than the input text. Here the simplified lex algorithm is used for the text simplification.
- **Translator:** It translates the input text from source language to output target language. Here the Google Translator is used for translation.
- **Quality Evaluation:** The quality of translated output text is measured using BLEU [23] and NIST [24] scores.

In the model shown in Figure 1, The text-set contains the pairs of text and reference texts. The text readability score is being evaluated to decide whether text simplification is required or not. If the text readability score is low then such complex text is given as input to the transformer, where the text simplification happens. The output simplified text is given to the Translator, which translates the input text into translated text. The reference texts along with the translated text are passed to the Quality Evaluator and Readability Evaluator. The output scores are used for the comparative analysis and plot generation.

3.3. Execution

The above model is used for executing different text-sets. The text-set contains pairs of text and reference texts. These text-sets are divided into two categories based on the readability score, a) complex sentences, and b) simple sentences. Complex sentences are those which are difficult to read and understand, and simple sentences are those which are comparatively easy to read and understand. We have used the readability measures such as the Flesch Reading Ease score, Flesch-Kincaid score, and Gunning Fog Score, to decide the readability of the sentences. The translated text quality is measured through the standard BLEU [23] and NIST [24].

Execution Steps:

1. Preparing of Text-Set pairs, i.e. Complex-Simple sentence pairs.
2. Deciding complexity of Text based on readability scores.
3. Collected translated output of Complex sentences.
4. Collected translated output of Simple sentences.
5. Measure the readability scores of translated outputs collected in step-3 and step 4.
6. Measure the BLEU and NIST scores of translated output collected in step-3 and step 4.
7. repeat step 2 for all sentences
8. and plot the comparison graph over
    a. input readability vs output readability
    b. input readability vs output quality.

Based on the language pair in the text-set, the text-sets are divided into two execution groups: 1) Non-sibling Languages, and 2) Sibling Languages.

3.3.1. Non-Sibling Languages

The text-sets having Gujarati-English, English-Gujarati, Hindi-English, and English-Hindi pairs fall under the Non-Sibling Languages group. Table 3 contains the Gujarati-English text-sets arranged in decreasing order of the input sentence complexity. The Label indicates the complexity level, Complex Sentence (CS), and Simplified Sentence (SS). The translated output scores, BLEU and NIST, show the quality of the translation. Here the translated outputs are taken from Google translate.

Table 3: Results of Gujarati to English Translation

Language Pair	Label	Input	Translated Output	BLEU	NIST
Guj_Eng	CS	જનનીની જોડ સખી નહિ જડે રે લોલ *(Jananini jod sakhi nahi jade re lol)*	The genital mutilation will not tighten	0.28	0.1
	SS1	હિત્ર િ ાઁ સિ ન અન્ય કોઈ નહિ િળે *(mitra maa samaan anya koi nahi made)*	I will not find anyone else like my friend	0.3	0.95
	SS2	હિત્ર િ ત સિ ન અન્ય કોઈ નહિ િળે *(mitra mata saman anya koi nahi made)*	No one else can be the same as a friend mother	0.37	3.73
	SS3	હિત્ર, િ ત સિ ન અન્ય કોઈ નહિ િળે *(mitra, mata saman anya koi nahi made)*	Friend, no one else can be found equal to mother.	0.41	4.09
Eng_Guj	CS	Friend, no one else can be found like mother	દોસ્ત, િ ત જવ બ ીજ કોઈ નથી િળ ી શકત ું *(dost, mata jevu biju koi nathi madi saktu)*	0.25	0.88
	SS1	Friend, no one elsecan be found equalto mother			

દોસ્તો, િ ત સિ ન બ ીજ કોઈ નથી િળી શકત ું	0.34	1.24

			(dosto, mata saman biju koi nathi madi saktu)		
	SS2	Friend, no one else can be found equivalent mother	હિત્ર, બ ીજ કોઈ સિકક્ષ િ ત નથી િળ ી શકત ું *(mitra, biju koi samkaksha mata nathi madi saktu)*	0.38	1.48
	SS3	Friend, no one can be found equal to mother	દોસ્તો, િ ત સિ ન કોઈ નથી િળ ી શકત ું *(dosto, mata saman koi nathi madi saktu)*	0.43	1.46
Eng_Hin	CS	Friend, no one else can be found like mother	द ोस्त, कि सी और ि ो म ि ी तरह नह ीं मिल सकि त ह ै *(dost, kisi or ko maa ki tarah nahi mil sakta he)*	0.45	1.77
	SS1	Friend, no one else can be found equal to mother	द ोस्त, ि ोई और म ि े बर बर नह ीं प यज सकि त ह ै *(dost, koi or maa ke barabar nahi paya ja sakta he)*	0.64	2.42
	SS2	Friend, no one else can be found equivalent mother	द ोस्त, कि सी और ि ो समि क्ष म नह ीं मिल सकि ती ह ै	*(dosto, kisi or ko samkaksh maa nahimil sakti he)*	

0.5	1.94

	SS3	Friend, no one can be found equal to mother	दोस्त, िोई भी म ा िे बर बर नह ीं प य ज सित है *(dost, koi bhi maa ke barabar nahi paya ja sakta he)*	0.55	2.14
	CS	दोस्त, म ाँ िे बर बर और िोई नह ीं ममल सित । *(dost, maa ke barabar or koi nahi mil sakta)*	Friends, nobody can get equal to mother	0.43	1.29
Hin_Eng	SS1	म त िे सम नऔर िोई नह ीं ममलेग , ममत्र। *(mata ke saman or koi nahi milga, mitra)*	Friends, no one else can get it.	0.43	0.8
	SS2	ममत्र, म त िे सम न और िोई नह ीं ममलेग । *(mitra, mata ke saman or koi nahi milega)*	Mother and no one will get, friend.	0.21	0.4
	SS3	*दोस्त, म ाँ िे बर बर और िोई नह ीं ममल सित ।* *(dost, maa ke barabar or koi nahi mil sakta)*	Friends, no one else and no one will get.	0.33	1.04

We observe that the output of the complex sentence is wrong and irrelevant. The output of simplified using our proposed regional language simplification engine (i.e. sentence SS1) is comparatively better than the complex sentence CS. The output of the simplified sentence SS2 is better than the SS1 output. And similarly, the output of the simplified sentence SS3 is better than the SS2. The corresponding BLEU and NIST scores show the improvements.

3.3.2. Sibling Languages

Here we tried the text-sets having Gujarati-Hindi and Hindi-Gujarati sibling languages pairs. The pairs of sentences used for Sibling, and Non-Sibling execution groups are the same, meaning the sentence translations are used. This is to observe the behavior of the sentence translation among the languages. The corresponding results are shown in Table 4.

Table 4: Results of Gujarati to Hindi Translation for OS

Language Pair	Label	Input	Translated Output	BLEU	NIST
Guj_Hin	CS	જનનીની જોડ સખી નહિ જડે રે લોલ *(Jananini jod sakhi nahi jade re lol)*	जनन ◌ीं◌ग◌ो ि◌◌ो खोजन म◌ुश◌्कि◌◌ ल नह◌ीं◌ होग । *(Jananaango ko khojana muskil nahi hoga)*	0.44	0.41
	SS1	હિત્ર િ◌ ◌ાઁ◌સિ ન અન્ય કોઈ નહિ િ◌ળે *(mitra maa samaan anya koi nahi made)*	मुझे अपने द◌ोस्त ि◌◌ी तरह ि◌◌ोई और नह ◌ीं◌ बर बर		
	SS2	હિત્ર િ◌ ત સિ ન અન્ય કોઈ નહિ િ◌ળે *(mitra mata saman anya koi nahi made)*			
	SS3	હિત્ર, િ◌ ત સિ ન અન્ય કોઈ નહિ િ◌ળે *(mitra, mata saman anya koi nahi made)* दोस्त, म◌ाँ◌ ि◌◌◌े			

मिलेग । (muze apne dost ki tarah koi or nah	0.43	1.92		i milega) मि म ि म के स म न	और कोई नहीं हो सकत (mit	ra maa ke samaan or koi nahi ho sakta)		
						दोस्त, माँ के बराबर और कोईनहीं मिल सकत। (dost, maa ke barabar or koi nahi mil sakta)	0.56	4.69
						િત થી બર બર કોઈ 0.47 5.13		
			Hin_Guj	CS	और कोई नहीं मिल सकत ।	હિત્રનેછી શકતી નથી.	0.39	1.18

	(dost, maa ke barabar or koi nahi mil sakta)	*(matathi barabar koi mitrane madi sakto nathi)*		
SS1	*म त ि◌◌े सम न और ि◌◌ोई नह ◌ी◌ं िमलेग , िमत्र।* *(mata ke saman or koi nahi milga, mitra)*	*િ◌ ત , િમજવે કઈન◌ી◌ુ◌ંિ◌ળ .* *(mata, mitra jeva koi nahi made)*	*0.28*	*1.06*
SS2	*िमत्र, म त ि◌◌े समन और ि◌◌ोई नह ◌ी◌ं िमलेग ।* *(mitra, mata ke saman or koi nahi milega)*	*િ◌ ત જવે િમત્ર કઈ નિ◌ી◌ુ◌ંિ◌ળ* *(mata jeva mitra koine nahi made)*	*0.28*	*1.15*
SS3	*दोस्त, म ◌ाँ ि◌◌े बर बर और ि◌◌ोई नह ◌ी◌ं िमल सि◌त ।* *(dost, maa ke barabar or koi nahi mil sakta)*	*દોસ્ત કઈપણ િ◌ ત ની બરબર ન િ◌◌ોઈ શકે.* *(dost koi pan matani barabar na koi shake)*	0.13	0.4

The output of complex sentence CS is wrong and irrelevant. While the output of the simplified sentences SS1, SS2, and SS3 shows improvement in their corresponding BLEU and NIST scores.

4. Results

Here the test-sets summaries and their corresponding plots are shown. The example summaries contain all the possible one-to- one translations among the three languages: Gujarati, Hindi, and English. Table 5 and Figure 2 contain the BLEU scores of the test-sets, and Table 6 and Figure 3 contain the NIST scores of the test-sets.

Table 5: BLEU Evaluation Scores of Translated Text. CS: ComplexSentence, SS: Simple Sentence

	Language Pair	**CS**	**SS1**	**SS2**	**SS3**
Non-Sibling	Guj_Eng	0.28	0.3	0.37	0.41

Language	Eng_Guj	0.25	0.34	0.38	0.43
	Eng_Hin	0.45	0.64	0.5	0.55
	Hin_Eng	0.43	0.43	0.21	0.33
Sibling Language	Guj_Hin	0.44	0.43	0.47	0.56
	Hin_Guj	0.39	0.28	0.28	0.13

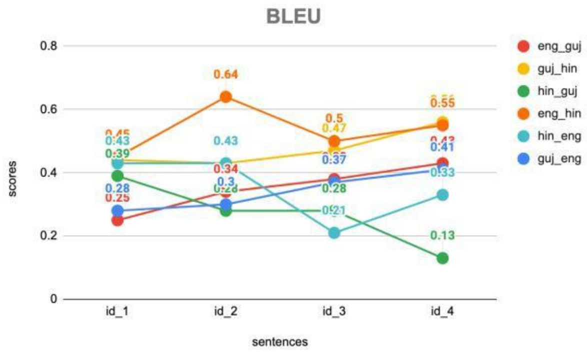


Fig 2: BLEU Score Improved Graph

Table 6: NIST Evaluation Scores of Translations. CS: Complex Sentence, SS: Simple Sentence

	Language Pair	CS	SS1	SS2	SS3
Non-Sibling Language	Guj_Eng	0.1	0.95	3.73	4.09
	Eng_Guj	0.88	1.24	1.48	1.46
	Eng_Hin	1.77	2.42	1.94	2.14
	Hin_Eng	1.29	0.8	0.4	1.04
Sibling Language	Guj_Hin	0.41	1.92	5.13	4.69
	Hin_Guj	1.18	1.06	1.15	0.4

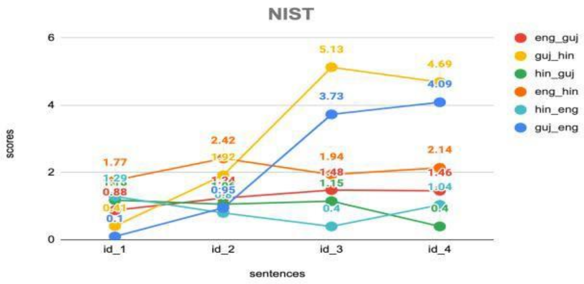


Fig 3: NIST Score Improved Graph

We can observe that in most of the cases the Simplified Statement SS3 shows higher BLEU and NIST scores compared to the other input statements.

Other Non-Sibling Text Sets Results:

BLEU **NIST**

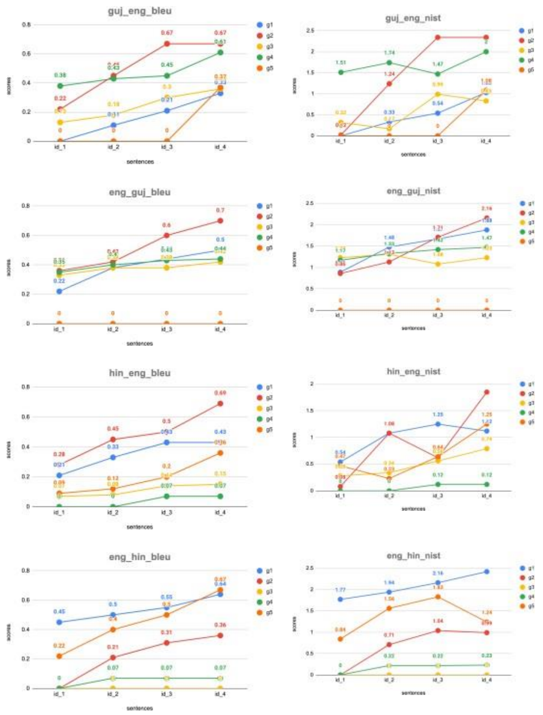


Fig 4: BLEU and NIST Scores for Other Non-Sibling Text Sets Results

Other Sibling Text Sets Results:

BLEU **NIST**

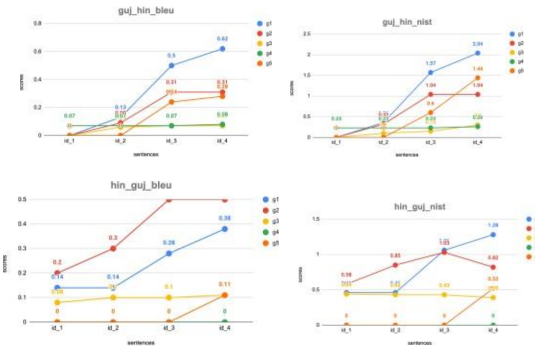
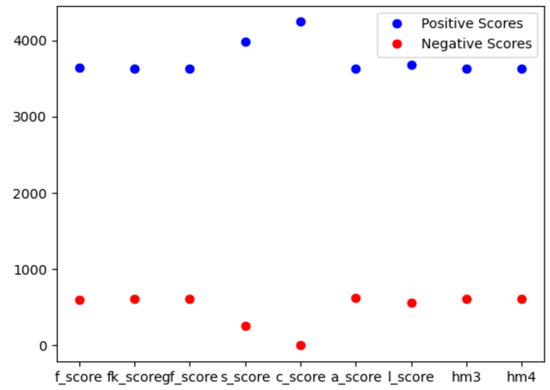


Figure 5: BLEU and NIST Scores for Other Sibling Text Sets Results

Table 7: Gujarati Complex vs Simple Readability Score. Human Prepared 4244 complex and simple Gujarati sentence pairs. The positive count indicates the corresponding readability score of the simple sentences is higher than the complex sentences. The negative count indicates the readability score of the complex sentences is higher than the simple sentences.

Gujarati	**Positive**	**Negative**	**%**
flesch_score	3639	605	85.74
flesch_kincaid_score	3635	609	85.65
gunning_fog_grade_score	3629	615	85.51
smog_readability_score	3981	263	93.8
coleman_liau_score	4244	0	100
automated_readabilityv	3626	618	85.44
linsear_write_score_diff	3683	561	86.78
hm3	3636	608	85.67
hm4	3636	608	85.67



Figs 6: Gujarati Complex vs Simple Readability Score

Table 8: English Complex vs Simple Readability Score. Human Prepared 4244 complex and simple English sentence pairs. The positive count indicates the corresponding readability score of the simple sentences is higher than the complex sentences. The negative count indicates the readability score of the complex sentences is higher than the simple sentences.

English	**Positive**	**Negative**	**%**
flesch_score	3637	607	85.7
flesch_kincaid_score	3647	597	85.93
gunning_fog_grade_score	3664	580	86.33
smog_readability_score	4244	0	100
coleman_liau_score	3634	610	85.63
automated_readabilityv	3630	614	85.53
linsear_write_score_diff	3687	557	86.88
hm3	0	0	0
hm4	0	0	0

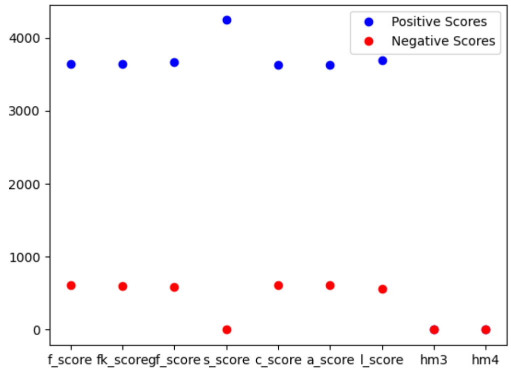


Fig 7: English Complex vs Simple Readability Score

5. Observations

From the above results and graphs, we found that,

- Simplifying input sentences by reducing their complexity improves the quality of the output statement.
- Gujarati to other language translations shows better improvement compared to other pairs.
- Gujarati and English show higher NIST scores compared to other pairs.
- Sibling languages show better translation quality than non-sibling languages.

These improvements are seen in the manually prepared text-sets. When simplified texts are generated automatically, we observe inconsistencies in the BLEU graph and the NIST graph. This could be because of the limitation of the auto text simplifier to generate correct and meaningful outputs compared to human simplification; as well as the limitation of readability scores to evaluate the correctness of the sentence compared to human evaluation.

6. Conclusions

Here we talked about improvement in text translation using text simplification as a pre-processing step. We showed experimental results of Gujarati, Hindi, and English language translations. We identify that a simplified input statement improves the translation quality of the output statement. We measured the sentence readability using readability scores (i.e., Flesch Reading Ease score, Flesch-Kincaid score, Gunning Fog Score, etc.). We measured the translation quality improvement using the BLEU and NIST scores. We observed an improvement in translation quality in the manually simplified sentences. The auto text simplifier outputs are not accurate compared to human simplification. The readability scores are not accurate compared to human evaluations. As a result, inconsistencies are found in the BLEU graph and the NIST graph, when results are generated automatically. In the manually prepared sentences, we observed that the translation quality is higher in the sibling languages compared to the non-sibling languages.

References

B. B. CK Bhensdadia Pushpak Bhattacharyya, “Introduction to Gujarati wordnet,” *Third Natl. Workshop Indowordnet Proc.*, vol. 494, 2002.

C. Boitet, “The French National MT-Project: Technical organization and translation results of CALLIOPE-AERO,” *Comput. Transl.*, vol. 1, no. 4, pp. 239–267, 1986, doi: 10.1007/BF00936424.

L. Feng, “Text simplification: A survey,” *City Univ. N. Y. Tech Rep*, pp. 7–23, 2008. Hautli-Janisz, “Pushpak Bhattacharyya: Machine translation,” *Mach. Transl.*, vol. 29, no. 3–4, pp. 285–289, Dec. 2015, doi: 10.1007/s10590-015-9170-7.

G. V. Garje and G. K. Kharate, “Survey of Machine Translation Systems in India,” *Int. J. Nat. Lang. Comput.*, vol. 2, no. 5, pp. 47–67, Oct. 2013, doi: 10.5121/ijnlc.2013.2504.

L. Feng, “Text Simplification: A Survey,” p. 35.

W. Contributors, “Gujarati Language,” *Definitions*, 2020. https://en.wikipedia.org/w/index.php?title=Gujarati_language&oldid=962021892 (accessed Jun. 08, 2020).

Wikipedia contributors, “Hindi Language,” in *Definitions*, Qeios, 2020. doi: 10.32388/W2U5JG.

R. Chandrasekar, C. Doran, and B. Srinivas, “Motivations and methods for text simplification,” in *Proceedings of the 16th conference on Computational linguistics* -, Morristown, NJ, USA, 1996, vol. 2, p. 1041. doi: 10.3115/993268.993361.

S. SPanchal, P. P Shukla, P. R Panchal, J. S Kolte, and B. H N, “Gujarati WordNet A Lexical Database,” *Int. J. Comput. Appl.*, vol. 116, no. 20, pp. 6–8, 2015, doi: 10.5120/20450-2803.

C. Callison-Burch, P. Koehn, and M. Osborne, “Improved Statistical Machine Translation Using Paraphrases,” in *Proceedings of the Human Language Technology Conference of the NAACL, Main Conference*, New York City, USA, Jun. 2006, pp. 17–24. Accessed: Aug. 21, 2022. [Online]. Available: https://aclanthology.org/N06-1003

S. Mirkin, “Confidence-driven Rewriting for Improved Translation,” Sep. 2013, Accessed: Aug. 21, 2022. [Online]. Available: https://www.academia.edu/4090244/Confidence_driven_Rewriting_for_Improved_Translation

W. Aziz, M. Dymetman, L. Specia, and S. Mirkin, “Learning an Expert from Human Annotations in Statistical Machine Translation: the Case of Out-of-Vocabulary Words,” Saint Raphaël, France, May 2010. Accessed: Aug. 21, 2022. [Online]. Available: https://aclanthology.org/2010.eamt-1.31

S. Tyagi, D. Chopra, I. Mathur, and N. Joshi, “Classifier based text simplification for improved machine translation,” in *2015 International Conference on Advances in Computer Engineering and Applications*, Mar. 2015, pp. 46–50. doi: 10.1109/ICACEA.2015.7164711.

S. Mirkin, S. Venkatapathy, M. Dymetman, and I. Calapodescu, “SORT: An Interactive Source-Rewriting Tool for Improved Translation,” in *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, Sofia, Bulgaria, Aug. 2013, pp. 85–90. Accessed: Aug. 21, 2022. [Online]. Available: https://aclanthology.org/P13-4015

H. Saggion, E. Gómez-Martínez, E. Etayo, A. Anula, and L. Bourg, “Text Simplification in Simplext. Making Text More Accessible,” vol. 47, Sep. 2011.

G. H. Paetzold and L. Specia, “Text Simplification as Tree Transduction,” 2013. Accessed: Aug. 21, 2022. [Online]. Available: https://aclanthology.org/W13-4813

J. Ameta, N. Joshi, and I. Mathur, “Improving the quality of Gujarati-Hindi Machine Translation through part-of-speech tagging and stemmer-assisted transliteration,” *Int. J. Nat. Lang. Comput.*, vol. 2, no. 3, pp. 49–54, Jun. 2013, doi: 10.5121/ijnlc.2013.2305.

P. Pimpale and R. Patel, “Reordering rules for English-Hindi SMT,” Apr. 2013, Accessed: Aug. 21, 2022. [Online]. Available: https://www.academia.edu/7421948/Reordering_rules_for_English_Hindi_SMT

J. N. Farr, J. J. Jenkins, and D. G. Paterson, “Simplification of Flesch Reading Ease Formula,” *J. Appl. Psychol.*, vol. 35, no. 5, pp. 333–337, 1951, doi: 10.1037/h0062427.

M. Solnyshkina, R. Zamaletdinov, L. A. Gorodetskaya, and A. I. Gabitov, “Evaluating Text Complexity and Flesch-Kincaid Grade Level,” *J. Soc. Stud. Educ. Res.*, vol. 8, pp. 238–248, Nov. 2017.

“THE GUNNING FOG READABILITY FORMULA.” https://readabilityformulas.com/gunning-fog-readability-formula.php (accessed Aug. 21, 2022).

K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, “BLEU,” in *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics - ACL ’02*, Morristown, NJ, USA, 2001, p. 311. doi: 10.3115/1073083.1073135.

Y. Zhang, S. Vogel, and A. Waibel, “Interpreting Bleu/NIST scores: How much improvement do we need to have a better system,” 2004.

Xu, W., Napoles, C., Pavlick, E., Chen, Q., & Callison-Burch, C. (2016). “Optimizing Statistical Machine Translation for Text Simplification”. Transactions of the Association for Computational Linguistics, 4, 401–415. Retrieved from https://cocoxu.github.io/publications/tacl2016-smt-simplification.pdf

Gujarati Rudhiprayog Ane Kahevat Sangrah. (n.d.). Retrieved from https://drive.google.com/uc?id=1gH7v1XoJ3f5ajsUg0Rz286LmNkKYfg2h&export=download

Ramesh, G., Doddapaneni, S., Bheemaraj, A., Jobanputra, M., Ak, R., Sharma, A., … Khapra, M. S. (2021). Samanantar: The Largest Publicly Available Parallel Corpora Collection for 11 Indic Languages. ArXiv [Cs.CL]. Retrieved from http://arxiv.org/abs/2104.05596

Bataa, B., & Altangerel, K. (2012). Word Sense Disambiguation in Gujarati Language. Proceedings - 2012 7th International Forum on Strategic Technology, IFOST 2012, (1), 44–47.

doi:10.1109/IFOST.2012.6357625