

Esha Kaushal, Sachin Singhal, Sakshee Agarwal, Daniel Whitenack, Matthew A. Lanham
 Purdue University, Krannert School of Management

ekausha@purdue.edu; singha22@purdue.edu; agarw230@purdue.edu; dan_whitenack@sil.org; lanhamm@purdue.edu

ABSTRACT

There is a critical need for translation checking to localize important content across all the industries. This poster presents question-answering techniques to check the comprehensibility of a text translation. It will save several manual checking hours and will help localize text content faster and more accurately.

INTRODUCTION

- ❖ Translation checking is a significant limiting factor on the pace of localizing important content.
- ❖ There are very few available methods to test the quality of translation, which results in local language speakers lacking timely, important information in the languages they value most.
- ❖ This project will demonstrate that **question answering techniques can be used to automatically check the comprehensibility of a text translation**. In doing this, the project addresses the translation checking bottleneck.

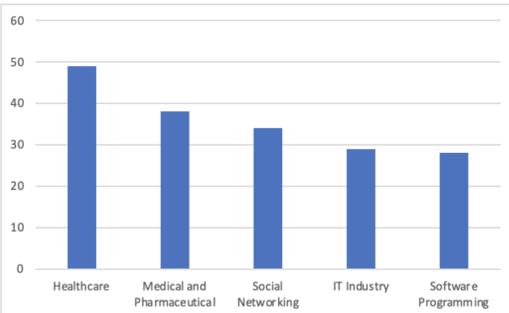
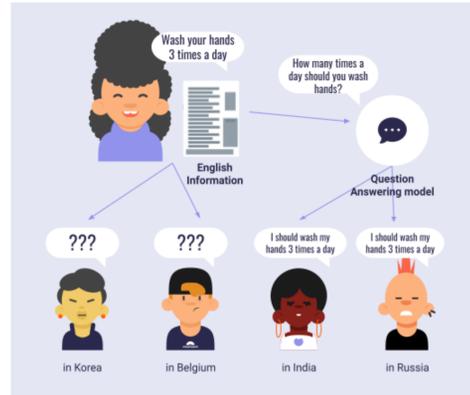


Fig. 1. Demand for translation by domain



RESEARCH QUESTIONS

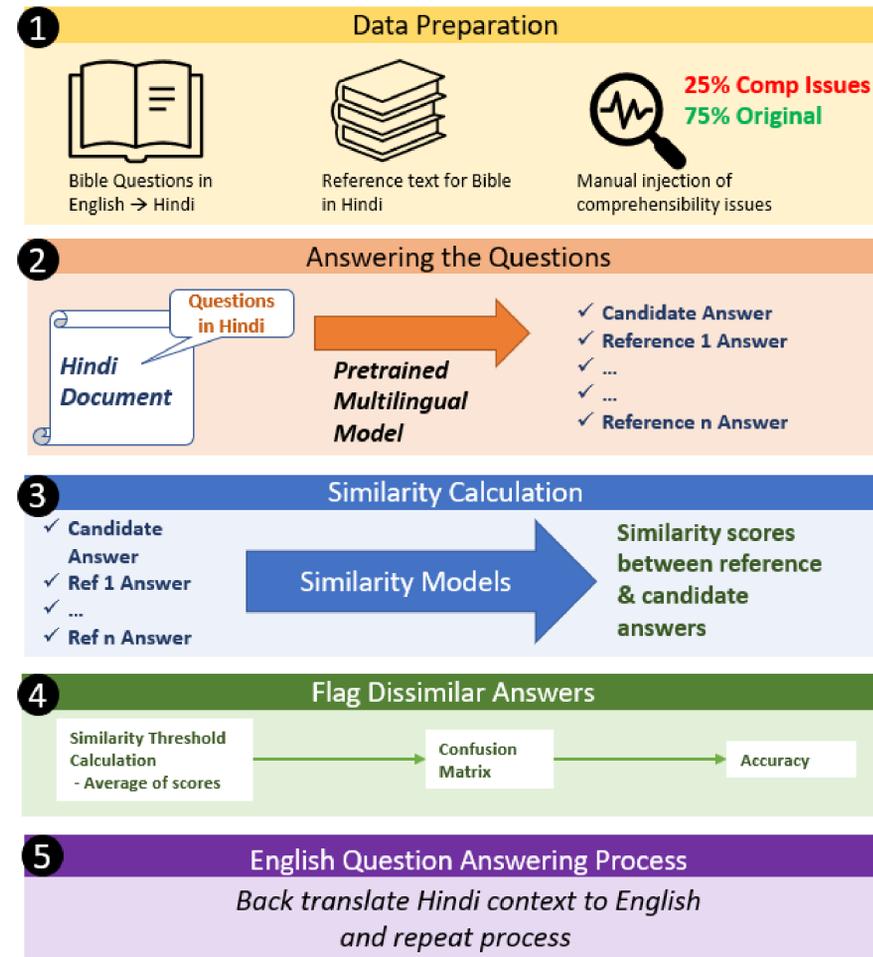
Our research focuses on answering the following questions:

- ❖ Can we create a Hindi data set for benchmarking the flagging of comprehensibility issues?
- ❖ Can we evaluate the comprehensibility of Hindi text using a pre-trained Hindi question answering model?
- ❖ Can we evaluate the comprehensibility of Hindi text using a pre-trained English question answering model and a Hindi-to-English machine generated back translation?

LITERATURE REVIEW

Study	Question-Answering	Machine translation	Similarity Calculation
Amazon Alexa Models	✓	✓	
A Qualitative Comparison of CoQA, SQuAD 2.0 and QuAC	✓	✓	
FQuAD: French Question Answering Dataset	✓		✓

METHODOLOGY



Pretrained Multilingual Question Answering Models Used –

- ❖ Bert-multi-cased-finetuned-xquadv1
- ❖ Monsoon-nlp/hindi-bert
- ❖ Monsoon-nlp/hindi-tpu-electra

Similarity Calculation Models Used –

- ❖ Language-Agnostic BERT Sentence Embedding (LaBSE)
- ❖ Bilingual Evaluation Understudy (BLEU)
- ❖ Language –Agnostic Sentence Representation (LASER)

Process of calculating Similarity Threshold–

- ❖ Take average of all similarity scores
- ❖ Determine quartiles and choose boundary
- ❖ Flag context whose score falls below threshold
- ❖ Create confusion matrix to judge accuracy

STATISTICAL RESULTS

Text_Similarity_Method	TN	FP	FN	TP	Precision	Recall	Accuracy
BLEU (AND)	169	59	31	43	42%	58%	70%
BLEU (Avg)	147	81	16	58	42%	78%	68%
LaBSE (OR)	166	62	15	59	49%	80%	75%
LaBSE (Avg)	206	22	17	57	72%	77%	87%
Eng_LaBSE (OR)	113	115	4	70	38%	95%	61%
Eng_LaBSE (Avg)	185	43	12	62	59%	84%	82%

Result Summary:

- ❖ **Hindi Question answering model** for Hindi text: BERT (base-multilingual-cased) fine-tuned for multilingual Q&A
- ❖ **English question answering model** for Back-translated English text: DistilBERT base uncased distilled SQuAD
- ❖ **Similarity Measure:** LaBSE word embedding
- ❖ **Flagging Method:** Averaging of similarity scores
- ❖ **Highest Recall rate:** using DistilBERT base uncased distilled SQuAD model on Back translated English text using average LaBSE similarity

BUSINESS IMPACTS

- ❖ Going forward, potential savings to be realized by SIL will be **\$3000 per project (50+ man hours)**.
- ❖ **A set of power tools for translation consultants**, which will allow them to do their checking work more thoroughly and more consistently
- ❖ **An early warning system for translators**, which will alert them to obvious problems so they can address them earlier in the translation process
- ❖ **An equalizer for administrators and strategists**, which will allow them to compare and evaluate methodologies and products (e.g., machine translation systems) on an equal footing

CONCLUSION

- ❖ A Hindi question answering model is a valid method to determine the comprehensibility of a translation with 85% accuracy
- ❖ A question-answering model in English on back translated data from any target language can be used to evaluate the comprehensibility of a text with 90% accuracy.
- ❖ This method can now be used by several companies in order to save several manual checking hours and will help localize text content faster and more accurately.

ACKNOWLEDGEMENTS

We thank our industry partner Daniel Whitenack for his trust, support and encouragement while approaching this problem. We also thank Professor Matthew Lanham for constant guidance on this project.