

Mitigating Translationese in Low-resource Languages: The Storyboard Approach

What is “Translationese”?

- Translationese refers to the linguistic phenomenon that occurs when translations exhibit characteristics that are not typical of the target language (Gellerstam, 1986)
- Translationese also refers to the use of more explicit and standardised constructions compared to original text (Baker et al., 1993)

Why mitigate “Translationese”?

- Detrimental impact on Downstream Tasks like Machine Translation
- Diminishes fluency and naturalness of the text

Storyboard

- Field Linguists use “storyboards” to collect “authentic” language data
- We want minimal influence from another language
- Rather than asking someone to translate, we ask them to narrate a story

Storyboard Sample (Lakka and the Ghosts)

Lakka is always seeing ghosts



Storyboard Sample (Lakka and the Ghosts)



"A ghost is hiding behind a tree!" she shouts.

Language of Focus

Our study focuses on four low-resource African languages: Swahili, Yorùbá, Hausa, and Ibibio.

Control Group: Text Translation

English Sentence: Lakka is always seeing ghosts



Control Group Sample



Yoruba: Gbogbo ìgbà ni Lakka máa ń rí
iwin

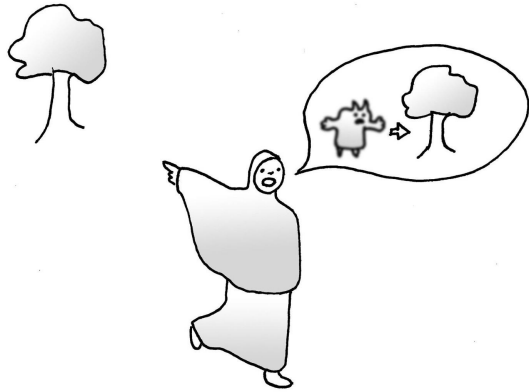
*(Google Translate:
Lakka always sees ghosts)*

Storyboard-Based Translation

English Sentence: Lakka is always seeing ghosts



Storyboard-Based Translation



English Sentence: "A ghost is hiding behind a tree!" she shouts.

Data Collection Process

Wait 1 hour



Storyboard-Based Translation



Given the image above, in one or more sentences (in the African language you selected earlier) write what the image is portraying

Storyboard-Based Translation Sample

Yoruba: Ọmọdẹ yìí sáà mǎa ñ rí òkú



*(Google Translate:
This child often sees dead bodies)*

Annotation Result

Language	Text Translation	Storyboard-Based
Hausa	1154	968
Ibibio	887	883
Swahili	1334	1211
Yorùbá	1448	1033

Accuracy Evaluation

- Human Evaluation
- LASER Embedding Cosine Similarity

Accuracy Human Eval Result

Language	Text Translation	Storyboard-Based	Both
Hausa	78%	21%	0%
Ibibio	79%	18%	2%
Swahili	64%	14%	22%
Yorùbá	68%	7%	23%

Embedding Similarity Result

Language	Text Translation	Storyboard-Based
Hausa	0.74	0.64
Swahili	0.69	0.58
Yorùbá	0.74	0.74

Fluency

- Human Evaluation
- Lexical Diversity (MTLD)
- Syntactic Complexity (POS perplexity)

Fluency Human Evaluation Result

Language	Text Translation	Storyboard-Based	Both
Hausa	39%	60%	0%
Ibibio	26%	36%	38%
Swahili	41%	48%	11%
Yorùbá	18%	34%	47%

Lexical Diversity (MTLD) Result

Language	Text Translation	Storyboard-Based
Hausa	11.4	7.8
Ibibio	7.2	8.0
Swahili	5.0	6.8
Yorùbá	14.2	16.1

Syntactic Complexity (POS Perplexity) Result

Language	Text Translation	Storyboard-Based
Hausa	49.4	6.7
Swahili	14.6	5.6
Yorùbá	$7.9 * 10^7$	$6.5 * 10^2$

